exercices de mathématiques

oraux x-ens

Serge Francinou
Hervé Gianella
Serge Nicolas

algèbre 1 analyse 1

algèbre 2 analyse 2

algèbre 3 analyse 3

Deuxième édition revue et augmentée

CASSINI

Powered by Jyloup™

exercices de mathématiques

oraux x-ens

algèbre 1

Serge Francinou
Hervé Gianella
Serge Nicolas

CASSINI

EXERCICES DE MATHÉMATIQUES
ORAUX X-ENS

Enseignement des mathématiques

1. J.-Y. Ouvrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouvrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*

SERGE FRANCINOU
HERVÉ GIANELLA
SERGE NICOLAS

Exercices de mathématiques des oraux de l'École polytechnique et des Écoles normales supérieures

Algèbre. Tome I

Deuxième édition, revue et augmentée

CASSINI

SERGE FRANCINOU, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Henri IV.

HERVÉ GIANELLA, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Saint-Louis.

SERGE NICOLAS, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Henri IV.

ISBN 978-2-84225-132-1
(1ere édition, 2001, ISBN-10 : 2-84225-030-3, ISBN-13 : 978-2-84225-030-0)

© Cassini, Paris, 2007.

Table des matières

Introduction 1

Chapitre 1. Combinatoire 5

1.1. Nombres de Fibonacci 6
1.2. Nombre de dérangements (1) 7
1.3. Nombre de dérangements (2) 9
1.4. Nombre de dérangements (3) 10
1.5. Nombres de Catalan 12
1.6. Nombres de Bell 14
1.7. Cardinal d'une relation d'équivalence 16
1.8. Cardinal de $\mathrm{GL}_n(\mathrm{K})$ et $\mathrm{SL}_n(\mathrm{K})$ 17
1.9. Cardinal de $\mathrm{SO}_2(\mathbb{Z}/p\mathbb{Z})$ 17
1.10. Nombre d'involutions 19
1.11. Partitions d'un entier 20
1.12. Un problème de théorie extrémale des ensembles 24
1.13. Ensembles définis par récurrence 25
1.14. Théorème de Howie 27
1.15. Distribution du premier chiffre des puissances de 2 29
1.16. Un théorème de Gauss 31
1.17. Théorème de Beatty (1926) 33
1.18. Un exercice d'Olympiades 34

Chapitre 2. Théorie des groupes 37

2.1. Existence d'un idempotent 37
2.2. Groupes dont l'ensemble des sous-groupes est fini 38
2.3. Morphismes de $\mathbb{Q}$ dans $\mathbb{Z}$ 38
2.4. Équivalence $\mathrm{Ker}\, f = \mathrm{Ker}\, f^2 \iff \mathrm{Im}\, f = \mathrm{Im}\, f^2$ 39
2.5. Sous-groupes finis de $\mathbb{Q}/\mathbb{Z}$ 40
2.6. Groupes abéliens de cardinal pq 41
2.7. Un cas particulier du lemme de Cauchy 41
2.8. Exposant d'un groupe abélien fini 43
2.9. Puissances dans un groupe abélien d'exposant fini 44
2.10. Lemme de Cauchy 46
2.11. Centre d'un p-groupe 47
2.12. Nombre de classes de conjugaison 49
2.13. Sous-groupe d'indice infini 50
2.14. Un théorème de Frobenius (1895) 52
2.15. Classes de conjugaison 53

2.16. Sous-groupes finis de $SO_3(\mathbb{R})$ 54
2.17. Groupe dérivé 57
2.18. Groupes quasi-cycliques de Prüfer 59
2.19. Le groupe modulaire 60
2.20. Unicité dans le théorème de structure des groupes abéliens finis 64
2.21. Génération du groupe symétrique 67
2.22. Plongement de $\mathcal{S}_n$ dans $\mathcal{A}_{n+2}$ 68
2.23. Morphismes d'un sous-groupe de $\mathcal{S}_p$ dans $\mathbb{Z}/p\mathbb{Z}$ 69
2.24. Morphismes de $\mathcal{S}_4$ dans $\mathcal{S}_3$ 70
2.25. Automorphismes de $\mathcal{S}_n$ 74
2.26. Structure de groupe ordonné sur $\mathbb{Z}^2$ 78

Chapitre 3. Anneaux et corps **81**

3.1. Inversibilité à droite 82
3.2. Calcul d'inverse 83
3.3. Anneaux tels que $x^3 = x$ 84
3.4. Commutativité ou anti-commutativité 84
3.5. Anneaux réguliers 85
3.6. Idéaux principaux 87
3.7. Idéaux maximaux 88
3.8. Anneau des nombres décimaux 90
3.9. Anneau $\mathbb{Z}[X]$ 90
3.10. Anneaux factoriels 91
3.11. Anneaux euclidiens 94
3.12. Anneau des entiers de Gauss (1) 98
3.13. Anneau des entiers de Gauss (2) 100
3.14. Une extension quadratique de $\mathbb{C}[X]$ 104
3.15. Anneau sans idéal non premier 106
3.16. Automorphismes de $\mathbb{Q}(\sqrt{2})$ 106
3.17. Nombres algébriques 107
3.18. Valuations sur $\mathbb{Q}$ 110
3.19. Valeurs absolues non-archimédiennes sur $\mathbb{C}(X)$ 112
3.20. Indépendance des valeurs absolues sur $\mathbb{Q}$ 114

Chapitre 4. Arithmétique **117**

4.1. Étude de l'irréductibilité d'une fraction 118
4.2. Équation $a^b = b^a$ dans $\mathbb{N}$ 119
4.3. Divisibilité 120
4.4. Points du réseau $\mathbb{Z}^n$ visibles de l'origine 122
4.5. Produits d'entiers consécutifs 123
4.6. Parties de $\mathbb{N}$ additivement stables 124
4.7. Un exercice pour les années impaires 125

4.8. Équation du second degré dans $\mathbb{Z}/p\mathbb{Z}$ 126
4.9. Un problème de congruence 127
4.10. Un multiple de 1996 qui ne s'écrit qu'avec des 4 127
4.11. Somme des puissances k-ièmes dans $\mathbb{Z}/p\mathbb{Z}$ 127
4.12. Théorème de Wilson (1759) 128
4.13. Cyclicité du groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$ 129
4.14. Critères de primalité 130
4.15. Diviseurs premiers communs aux termes d'une suite arithmétique 131
4.16. Nombres de Fermat 132
4.17. Infinité des nombres premiers congrus à 3 modulo 4 134
4.18. Version faible du théorème de la progression arithmétique de Dirichlet (1837) 135
4.19. Plus petit nombre premier ne divisant pas n 137
4.20. Théorème de Kurschak (1918) 138
4.21. Théorème de Legendre (1808) 139
4.22. Un produit de trois entiers consécutifs n'est jamais une puissance k-ième 139
4.23. Théorème de Palfy-Erdös 140
4.24. Valuation p-adique de $C_{p^n}^k$ 143
4.25. Congruences de Lucas (1878) 144
4.26. Un problème de congruence 145
4.27. Le problème de Ducci 147
4.28. Expression de $\sum_{k=1}^{n} \tau(k)$ 150
4.29. Une majoration de σ 151
4.30. Équation faisant intervenir σ 151
4.31. Sur la fonction σ 152
4.32. Un théorème d'Erdös (1946) 153
4.33. Probabilité pour que deux entiers soient premiers entre eux 156
4.34. Écriture d'un nombre premier comme somme de deux carrés 158
4.35. Théorème des deux carrés, preuve combinatoire 160
4.36. Théorème des quatre carrés de Lagrange (1770) 162
4.37. Lemme de Davenport-Cassels 164
4.38. Une équation diophantienne 165
4.39. Théorème de Sophie Germain (1823) 167

Chapitre 5. Polynômes **171**

5.1. Égalité polynomiale 172
5.2. Une sous-algèbre de $\mathbb{R}[X]$ 173
5.3. Condition de divisibilité (1) 174
5.4. Condition de divisibilité (2) 174
5.5. Condition pour que $(P')^p$ divise P^q 175

5.6. Une équation fonctionnelle (1) 176
5.7. Une équation fonctionnelle (2) 177
5.8. Un théorème de Liouville (1879) 179
5.9. Théorème de Mason (1984) 180
5.10. Résultant de deux polynômes 182
5.11. Caractérisation d'un polynôme par les antécédents de deux points distincts 184
5.12. Image réciproque d'une partie finie par un polynôme 185
5.13. Sur le nombre d'antécédents d'un point par un polynôme . 186
5.14. Polynôme rationnel inséparable de degré 5 187
5.15. Un polynôme irréductible de $\mathbb{Z}[X]$ 188
5.16. Critère d'Eisenstein 188
5.17. Irréductibilité de Φ_p dans $\mathbb{Q}[X]$ 190
5.18. Décomposition de $1 + X + X^2 + \cdots + X^{n-1}$ 192
5.19. Polynômes complexes d'image réelle 193
5.20. Sommes de deux carrés dans $\mathbb{R}[X]$ 194
5.21. Polynômes positifs sur $[-1, 1]$ 195
5.22. Polynôme positif 197
5.23. Diviseurs d'un polynôme de $\mathbb{Z}[X]$ 197
5.24. Polynômes de Hilbert 198
5.25. Interpolation de Lagrange 199
5.26. Polynômes complexes envoyant surjectivement $\mathbb{Q}$ sur $\mathbb{Q}$ . . 200
5.27. Caractérisation des polygones réguliers 201
5.28. Polynômes entiers et fonctions polynomiales induites sur $\mathbb{Z}/p^i\mathbb{Z}$ 202
5.29. Images des racines d'un polynôme 206
5.30. Un calcul de $\zeta(2)$ 207
5.31. Formules de Newton (1707) 209
5.32. Polynômes réels scindés 211
5.33. Un théorème de Kronecker 213
5.34. Racines réelles de $nX^n - X^{n-1} - \cdots - X - 1$ 214
5.35. Un polynôme scindé sur $\mathbb{R}$ 215
5.36. Dénombrement de racines réelles 216
5.37. Dérivation et polynômes réels scindés 216
5.38. Un théorème de Laguerre 217
5.39. Plans vectoriels de polynômes scindés 219
5.40. L'ouvert des polynômes scindés à racines simples sur $\mathbb{R}$ dans l'ensemble des polynômes unitaires de degré n 221
5.41. Polynômes de Tchebychev 222
5.42. Inégalités de Bernstein et de Markov 225
5.43. Théorème de Gauss-Lucas 229
5.44. Polynôme à racines réelles 231
5.45. Théorème d'Eneström-Kakeya 233

5.46. Construction d'un polynôme satisfaisant des conditions sur le module de ses valeurs 234
5.47. Inégalité de Landau 234
5.48. Critère de Routh-Hurwitz pour le degré 3 235
5.49. Règle de Descartes 237
5.50. Théorème de Sturm 238
5.51. Décomposition en éléments simples 240
5.52. Inégalité de Bernstein 241
5.53. Inversion de la matrice de Hilbert 242
5.54. Automorphismes de K(X) 244
5.55. Approximation d'un irrationnel algébrique par des rationnels 247
5.56. Transcendance de e 249
5.57. Polynômes à plusieurs variables à valeurs entières 252
5.58. Applications de $\mathbb{R}^2$ dans $\mathbb{R}$ polynomiales par rapport à chaque variable 254
5.59. Un théorème de Bezout 255

Chapitre 6. Espaces vectoriels. Algèbres **259**

6.1. Intersection de sous-espaces 259
6.2. Supplémentaire commun 261
6.3. Drapeaux 263
6.4. Lemmes de factorisation 266
6.5. Condition pour que $\operatorname{rg} g \leqslant \operatorname{rg} f$ 267
6.6. Endomorphismes stabilisant les sous-espaces de dimension k 269
6.7. Exemple d'utilisation des espaces quotients 269
6.8. Majoration de l'indice de nilpotence 270
6.9. Produit commutatif d'endomorphismes nilpotents 271
6.10. Inégalité de Sylvester 272
6.11. Pseudo-inverse 274
6.12. Endomorphismes u tels que $\operatorname{Ker} u = \operatorname{Im} u$ 275
6.13. Endomorphismes u tels que $\operatorname{Ker} u \oplus \operatorname{Im} u = \mathrm{E}$ 276
6.14. Décomposition de Fitting 277
6.15. Endomorphismes tels que $\mathrm{E} = \operatorname{Ker} u \oplus \operatorname{Im} u$ 279
6.16. Endomorphisme annulé par un polynôme de degré 2 à racines simples 280
6.17. Équation linéaire dans $\mathcal{L}(\mathrm{E})$ 282
6.18. Projecteurs 284
6.19. Une somme de projecteurs 285
6.20. Endomorphismes de $\mathbb{C}[X]$ 286
6.21. Formule de Burnside 289
6.22. Théorème de Maschke 292
6.23. Automorphismes de la K-algèbre $\mathcal{L}(\mathrm{E})$ 293
6.24. Simplicité de $\mathcal{L}(\mathrm{E})$ 295

6.25. Idéaux à gauche de $\mathcal{L}(E)$ 295
6.26. Idéaux à droite de $\mathcal{L}(E)$ 297
6.27. Orthogonalité duale en dimension quelconque 299
6.28. Familles libres d'applications 300
6.29. Familles positivement génératrices 302
6.30. Familles positivement génératrices de E^* 304
6.31. Caractérisation de $\mathbb{C}$ 307
6.32. Théorème de Frobenius 308
6.33. Sous-algèbres de dimension finie de $\mathcal{C}^0(\mathbb{R},\mathbb{R})$ 310
6.34. Racine carrée de la dérivation 311
6.35. Φ-dérivation (1) 311
6.36. Φ-dérivation (2) 312
6.37. Φ-dérivation (3) 314
6.38. Étude d'une algèbre 315

Chapitre 7. Matrices **319**

7.1. Condition pour que $\operatorname{rg} g \leqslant \operatorname{rg} f$ 320
7.2. Factorisation bilatère 321
7.3. Fonctions multiplicatives et inversibilité 322
7.4. Degré et valuation du polynôme $\det(XA+B)$ 322
7.5. Endomorphismes de $\mathcal{M}_n(\mathbb{C})$ stabilisant le groupe linéaire . . 323
7.6. Endomorphismes de $\mathcal{M}_n(\mathbb{C})$ conservant le rang 325
7.7. Équation matricielle $X + {}^tX = (\operatorname{Tr} X)A$ 328
7.8. Dual de $\mathcal{M}_n(K)$ 329
7.9. Tout hyperplan de $\mathcal{M}_n(K)$ coupe $\mathrm{GL}_n(K)$ 331
7.10. Dimension maximale de sous-espaces de matrices de rang $\leqslant p$ 331
7.11. Orthogonalité duale 335
7.12. Crochets de Lie de $\mathcal{M}_n(K)$ 336
7.13. Diagonales possibles dans une classe de similitude 337
7.14. Traces modulo p 339
7.15. Matrices monotones 341
7.16. Puissances d'une matrice strictement stochastique 343
7.17. Théorème de Frobenius-König (1912-1916) 345
7.18. Décomposition de Bruhat et drapeaux 347
7.19. Bases d'un groupe abélien 350
7.20. Première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$ 352
7.21. Matrices équivalentes dans $\mathcal{M}_n(\mathbb{Z})$ 354
7.22. Structure des groupes abéliens finis 358
7.23. Lemme de Siegel 360

Index **369**

Introduction

Cet ouvrage est le premier des six tomes d'un recueil d'exercices de mathématiques destiné à la préparation des oraux des concours d'entrée aux Écoles normales supérieures et à l'École polytechnique. Trois volumes seront consacrés à l'algèbre et trois à l'analyse.

La vocation première des Écoles normales est de former des chercheurs ou des enseignants-chercheurs. Le concours d'entrée vise donc à détecter les qualités scientifiques du candidat, son aptitude à la recherche. À l'oral, on jugera avant tout la capacité de prendre des initiatives, d'utiliser une indication, de mener à bien une démarche. On ne sera pas surpris que les exercices posés aient un contenu mathématique riche, qu'ils soient très éloignés du simple exercice technique, d'application du cours, qu'ils soient souvent difficiles. Ils visent la plupart du temps à la démonstration d'un résultat mathématique significatif. Ils pourraient apparaître excessivement difficiles, si on perdait de vue le déroulement concret de l'épreuve. L'oral des ENS est un long dialogue (l'épreuve dure environ cinquante minutes, comme d'ailleurs à l'École polytechnique) entre le candidat et l'examinateur, qui tout au long de l'épreuve fournit des indications, quand c'est nécessaire, pour relancer la réflexion du candidat et tester ses réactions. Il est d'ailleurs impossible de rendre pleinement compte dans un recueil d'exercices du caractère oral de l'épreuve.

L'École polytechnique, quant à elle, est plus généraliste. Les exercices posés au concours sont de facture plus classique et, en règle générale, l'examinateur intervient moins. C'est au candidat de montrer sa maîtrise du programme dans la résolution d'un exercice dont la difficulté est cependant très variable. Certains sont proches des exercices d'ENS. Les énoncés circulent d'ailleurs d'un concours à l'autre, ou peuvent même être repris d'exercices d'Olympiades.

Les énoncés qui figurent dans ce recueil, ont été donnés entre 1994 et 2006. Ils sont extraits pour l'essentiel des listes publiées chaque année par la RMS (*Revue des mathématiques de l'enseignement supérieur* aux éditions Vuibert jusqu'en 2003 et désormais *Revue de la filière Mathématiques* aux éditions e.net) dont nous remercions les auteurs pour l'aide précieuse qu'ils apportent ainsi aux élèves et aux professeurs des classes préparatoires. Il s'agit de versions communiquées par les étudiants, reflétant la compréhension que ceux-ci ont eue de l'exercice et le déroulement conjoncturel de leur oral, comme le montrent les variations d'une année à l'autre pour un même exercice. Nous n'avons pas hésité à les modifier, pour rectifier des erreurs, compléter un énoncé

quand manifestement l'exercice s'est arrêté avant que le résultat que l'examinateur avait en vue ne soit atteint, ou ajouter des indications.

Nous avons choisi de laisser quelques énoncés «bruts», ceux pour lesquels nous estimons qu'une démarche naturelle (qui peut être longue et ardue) permet de conduire à la solution. Pour d'autres exercices, nous avons pris la liberté de rajouter des questions intermédiaires, qui auraient pu être celles posées par l'examinateur. Quitte à perdre en concision, nous avons tenu à rédiger les solutions les plus pédagogiques possible, essayant d'exposer clairement les idées et démarches de raisonnement, sans escamoter les détails ou calculs qui peuvent paraître évidents. On évite autant que possible l'introduction d'une astuce ou d'un objet *ad hoc* permettant d'atteindre rapidement la solution. S'il n'y a pas moyen d'expliquer l'origine de cette astuce, c'est que l'exercice est peu intéressant et que l'étudiant en tirera peu de profit.

À l'intérieur de chaque chapitre, les exercices ont été regroupés thématiquement et à l'intérieur de chaque thème, souvent par ordre de difficulté croissante. Ainsi regroupés, ils apparaîtront plus accessibles, car plongés dans leur contexte mathématique, éclairés par d'autres exercices voisins. Les introductions historiques qui ouvrent chaque chapitre, outre leur intérêt propre, visent au même but. Enfin, nous avons agrémenté les énoncés de quelques remarques préliminaires. Sans faire de rappels de cours systématiques, nous avons énoncé, voire redémontré, certains résultats : lemmes classiques, intervenant dans la résolution d'un grand nombre d'exercices, ou résultats au contraire à la lisière du programme, mais utiles, pour lesquels des éclaircissements étaient nécessaires. On trouvera aussi des remarques de synthèse ou des généralisations qui, nous l'espérons, pourront amener le candidat curieux à approfondir ses connaissances. Les quelques indications bibliographiques ont le même objectif.

Le lecteur ne tirera profit de ce livre que s'il cherche des solutions personnelles des exercices avant d'en étudier les corrigés. Une bonne connaissance du cours est indispensable. En effet, les théorèmes du programme fournissent bon nombre de schémas de démonstration. Rappelons aussi quelques démarches générales qui peuvent faciliter l'appréhension des exercices difficiles :

▷ l'étude de cas particuliers permet souvent d'entrevoir les idées qu'il faudra mettre en œuvre dans la résolution du cas général. Ce peut être l'étude du problème pour les petites dimensions, dans un exercice d'algèbre linéaire, l'étude du problème dans $\mathbb{Z}$ pour un exercice où il est question d'un anneau quelconque... ;

▷ le renforcement des hypothèses peut aboutir à un problème plus simple : cas où le corps est algébriquement clos, s'il est question d'un corps K ou d'un K-espace vectoriel ; cas où les matrices sont diagonali-

sables dans un exercice d'algèbre linéaire ; étude du cas où la fonction f est supposée dérivable au lieu d'être seulement continue en analyse... ;

▷ parfois, l'étude de certains cas particuliers permet d'atteindre le cas général : problème linéaire qu'il suffit de traiter sur une base ; passage au cas général par des arguments de densité...

Au-delà des étudiants en classe préparatoire, ces ouvrages intéresseront aussi les candidats au CAPES et à l'Agrégation, qui y trouveront matière à réviser les principales notions du programme, ainsi que des exemples pour nourrir un développement pour leur oral.

Ce premier tome est consacré à l'algèbre générale (combinatoire, groupe, anneaux, corps, arithmétique, polynômes) et aux éléments de base de l'algèbre linéaire (espaces vectoriels, matrices). Il couvre de manière assez complète le programme d'algèbre des classes préparatoires scientifiques et pourra déjà être utilisé par des élèves de première année. La matière des différents chapitres est assez classique, sauf peut-être le premier qui fait appel à des techniques assez diverses. Le second tome d'algèbre contiendra la réduction, l'algèbre bilinéaire et multilinéaire et la géométrie. Enfin le troisième et dernier tome d'algèbre portera sur les espaces euclidiens, les espaces hermitiens, les formes quadratiques et la géométrie.

Nous remercions André Bellaïche, René Cori et Rached Mneimné, ainsi que nos élèves, pour leur relecture approfondie de l'ouvrage et leurs nombreuses suggestions, tant sur le fond que sur la forme.

Outre les notations habituelles, indiquons que $\mathrm{E}(x)$ désigne la partie entière du réel x et que le cardinal d'un ensemble fini X est noté indifféremment $\mathrm{Card}\,\mathrm{X}$ ou $|\mathrm{X}|$. Nous avons choisi de conserver la notation de Pascal C_n^k pour désigner les coefficients binomiaux plutôt que la notation anglo-saxonne $\binom{n}{k}$.

Enfin, si vous souhaitez nous contacter pour nous faire part de vos remarques, vous pouvez envoyer un courriel à l'adresse *fgn.cassini@free.fr*.

Chapitre 1

Combinatoire

L'Analyse combinatoire (ou plus brièvement la combinatoire) est la partie des Mathématiques qui s'occupe des configurations finies. Par nature, elle intervient dans de nombreuses branches des Mathématiques : arithmétique, théorie des graphes, théorie des groupes... Les techniques utilisées sont très variées : séries génératrices, analyse asymptotique...

Les exercices qui suivent traitent essentiellement du dénombrement, qui consiste à déterminer le nombre de configurations possibles d'un certain type : nombre de bijections sans point fixe d'un ensemble de cardinal n, nombre de matrices inversibles de taille n sur un corps fini... À la base du dénombrement se trouve l'étude de structures élémentaires : nombre d'éléments d'un produit cartésien, d'une réunion, nombre d'arrangements, de combinaisons – qui ont donné leur nom à la combinatoire.

La méthode la plus naturelle pour dénombrer un ensemble de configurations consiste à établir une bijection entre cet ensemble et un ensemble dont le nombre d'éléments est déjà connu. Mais ce n'est pas toujours la plus facile. Donnons un exemple où il s'agit non pas d'effectuer un dénombrement, mais d'établir une identité. Pour montrer que dans un ensemble E, *il y a autant de sous-ensembles de cardinal pair que de sous-ensembles de cardinal impair, on peut utiliser la formule du binôme de Newton :*

$$0 = (1-1)^n = \sum_{k=0}^{n} (-1)^k \mathrm{C}_n^k = \sum_{k \text{ pair}} \mathrm{C}_n^k - \sum_{k \text{ impair}} \mathrm{C}_n^k.$$

On peut aussi établir une bijection entre l'ensemble des sous-ensembles de E *de cardinal pair et celui des sous-ensembles de cardinal impair, par exemple comme ceci : on fixe $a \in \mathrm{E}$; à* A *on associe $\mathrm{A} \setminus \{a\}$ si* A *contient a, $\mathrm{A} \cup \{a\}$ sinon. De telles démonstrations où les égalités d'entiers traduisent des bijections entre ensembles finis sont appelées «démonstrations combinatoires».*

Dans les cas où un dénombrement exact est impossible, on s'intéressera à une évaluation asymptotique du nombre de structures.

Les exercices proposés dans ce chapitre concernent des sujets très variés et illustrent plusieurs techniques combinatoires importantes. Ils sont en général assez difficiles.

Le premier exercice voit apparaître la célèbre suite de Fibonacci.

Léonard de Pise, plus connu sous le nom de Leonardo Fibonacci est un des rares mathématiciens du Moyen Âge qui soit passé à la postérité. C'est dans le Liber Abacci. *écrit en 1202, qu'apparaît la suite qui porte son nom. Cet ouvrage remarquable, qui a introduit l'usage des chiffres arabes en Europe, réunissait presque toutes les connaissances en arithmétique et en algèbre de l'époque. La suite de Fibonacci a révélé par la suite de multiples propriétés extraordinaires. On la rencontre aussi bien en botanique en comptant les pétales des fleurs ou les écailles des ananas (c'est la phyllotaxie), que sur les marchés financiers où les analystes graphiques essaient de s'en servir pour prévoir l'amplitude des mouvements de hausse ou de baisse d'une action (ratio de Fibonacci). Elle apparaît dans l'exercice suivant, qui fournit un premier exemple de situation combinatoire où l'on utilise un raisonnement par récurrence. De telles situations se rencontrent chaque fois qu'il est possible de construire les configurations de rang n à partir de configurations de rang inférieur, ce qui est fréquent en combinatoire. Historiquement, c'est d'ailleurs pour donner une preuve solide de la formule des combinaisons (donnant la valeur de C_n^k) que Pascal a formulé pour la première fois de façon claire le principe de la démonstration par récurrence.*

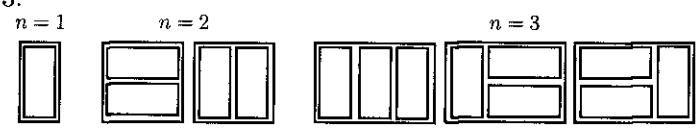
1.1. Nombres de Fibonacci

Déterminer le nombre a_n de manières de recouvrir un damier de dimension $2 \times n$ avec des pièces de dimension 1×2. Montrer que si n est assez grand, a_n est la partie entière de $\frac{1}{2} + \frac{1}{\sqrt{5}} \left(\frac{1+\sqrt{5}}{2} \right)^{n+1}$.

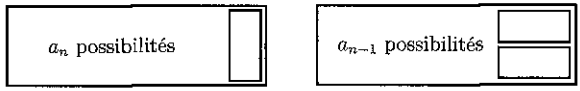
(École polytechnique)

▷ **Solution.**

On va effectuer ce dénombrement en établissant une relation de récurrence. Les configurations ci-dessous montrent que $a_1 = 1$, $a_2 = 2$ et $a_3 = 3$.



On partitionne l'ensemble des dispositions dans un échiquier de taille $n+1$ selon que la case $(1, n+1)$ est recouverte par un domino vertical ou horizontal. Dans le premier cas il reste à recouvrir un échiquier de taille $2 \times n$ ce qu'on peut faire de a_n manières. Dans le second cas, le domino qui recouvre la case $(2, n+1)$ est lui aussi horizontal et on a a_{n-1} possibilités de recouvrir l'échiquier $2 \times (n-1)$ restant.



La suite (a_n) vérifie donc la relation de récurrence $a_{n+1} = a_n + a_{n-1}$. On peut poser $a_0 = 1$ et on reconnaît alors, à un décalage d'indice près, la suite de Fibonacci. L'équation caractéristique associée à cette suite récurrente linéaire est $x^2 - x - 1 = 0$ et ses racines sont $\Phi = \dfrac{1+\sqrt{5}}{2}$, le nombre d'or[1] et $-\dfrac{1}{\Phi}$. On obtient alors, en tenant compte des conditions initiales, la formule[2]

$$\begin{aligned} a_n &= \frac{1}{\sqrt{5}}\left(\Phi^{n+1} - \left(-\frac{1}{\Phi}\right)^{n+1}\right) \\ &= \frac{1}{\sqrt{5}}\left(\left(\frac{1+\sqrt{5}}{2}\right)^{n+1} - \left(\frac{1-\sqrt{5}}{2}\right)^{n+1}\right). \end{aligned}$$

Comme $u_n = \dfrac{1}{2} + \dfrac{1}{\sqrt{5}}\left(\dfrac{1+\sqrt{5}}{2}\right)^{n+1} - a_n$ tend vers $\dfrac{1}{2}$, la partie entière de u_n est nulle pour n assez grand. Or celle-ci vérifie l'égalité

$$\mathrm{E}(u_n) = \mathrm{E}\left(\frac{1}{2} + \frac{1}{\sqrt{5}}\left(\frac{1+\sqrt{5}}{2}\right)^{n+1}\right) - a_n.$$

D'où le résultat de la seconde question. ◁

Nous abordons maintenant le classique problème des dérangements posé pour la première fois en 1708 par Pierre Rémond de Montmort dans son livre Essai d'analyse sur les jeux de hasard *: si n personnes quittant une réunion, prennent au vestiaire un parapluie au hasard, quelle est la probabilité pour que personne n'ait son propre parapluie ? Nous allons aborder ce problème de plusieurs manières différentes. Le premier énoncé en propose une résolution élémentaire.*

1.2. Nombre de dérangements (1)

1. Calculer $\displaystyle\sum_{k=0}^{p} (-1)^k \mathrm{C}_n^k \mathrm{C}_{n-k}^{p-k}$ pour $0 \leqslant p \leqslant n$.

1. Ainsi noté en l'honneur du grec Phidias (V^e siècle avant notre ère) qui décora notamment le Parthénon.

2. Dite formule de Binet (1843) mais injustement dénommée puisque Euler l'avait déjà découverte en 1765.

2. Soit D_n le nombre de permutations de $\mathcal{S}_n$ n'ayant pas de point fixe. Montrer que $\sum_{k=0}^{n} C_n^k D_{n-k} = n!$ (on pose $D_0 = 1$).

3. Montrer que $D_n = n!\left(\sum_{k=0}^{n} \frac{(-1)^k}{k!}\right)$.

(**École polytechnique**)

▷ **Solution.**

1. On a

$$\begin{aligned}\sum_{k=0}^{p}(-1)^k C_n^k C_{n-k}^{p-k} &= \sum_{k=0}^{p}(-1)^k \frac{n!(n-k)!}{k!(n-k)!(p-k)!(n-p)!}\\ &= \sum_{k=0}^{p}(-1)^k C_n^p C_p^k\\ &= C_n^p(1-1)^p = \begin{cases} 0 \text{ si } p \geqslant 1\\ 1 \text{ si } p = 0\end{cases}\end{aligned}$$

2. Soit $n \geqslant 1$. Notons P_k l'ensemble des permutations de $\mathcal{S}_n$ ayant exactement k points fixes. Il est clair que $\{P_0, P_1, \ldots, P_n\}$ est une partition de $\mathcal{S}_n$. En particulier $n! = \sum_{k=0}^{n} \mathrm{Card}(P_k)$. Nous allons montrer que $\mathrm{Card}(P_k) = C_n^k D_{n-k}$ pour tout k. On a par définition $\mathrm{Card}(P_0) = D_n$ et $\mathrm{Card}(P_n) = 1 = C_n^n D_0$, car seule l'identité possède n points fixes. Pour $k \geqslant 1$, un élément de P_k est parfaitement déterminé par le choix de ses points fixes (C_n^k possibilités) et par le choix de la permutation induite sur les $n-k$ éléments restants (D_{n-k} possibilités puisque cette permutation est un dérangement d'un ensemble à $n-k$ éléments). Remarquons en passant que $D_1 = 0$ et que effectivement P_{n-1} est vide : une permutation ne peut avoir exactement $n-1$ points fixes ! On a donc

$$\boxed{n! = \sum_{k=0}^{n} C_n^k D_{n-k} = \sum_{k=0}^{n} C_n^k D_k}.$$

3. On part du membre de droite de l'égalité demandée. On a

$$n!\left(\sum_{k=0}^{n}\frac{(-1)^k}{k!}\right) = \sum_{k=0}^{n}(-1)^k C_n^k (n-k)! = \sum_{k=0}^{n}(-1)^k C_n^k \sum_{p=0}^{n-k} C_{n-k}^p D_p$$

en remplaçant le terme $(n-k)!$ par la formule obtenue dans la question 2. En échangeant l'ordre de sommation, on obtient grâce à la question 1,

$$n!\left(\sum_{k=0}^{n}\frac{(-1)^k}{k!}\right)=\sum_{p=0}^{n}\left(\sum_{k=0}^{n-p}(-1)^k C_n^k C_{n-k}^p\right)D_p=D_n. \lhd$$

Une seconde possibilité, pour aboutir directement à cette dernière relation, consiste à utiliser la formule du crible (celle-ci est rappelée dans la solution de l'exercice 4.33). En effet, si on note U_i l'ensemble des permutations de $\mathcal{S}_n$ fixant l'entier i, pour $1 \leqslant i \leqslant n$, on obtient $D_n = n! - \mathrm{Card}\left(\bigcup_{i=1}^{n} U_i\right)$. Le cardinal de la réunion s'évalue alors facilement à l'aide de la formule du crible. Nous invitons le lecteur à achever ce calcul.

Dans l'exercice suivant, l'utilisation des séries entières fournit une troisième voie pour inverser le système linéaire donné par les relations de la question 2 ci-dessus et obtenir la formule de la question 3.

1.3. Nombre de dérangements (2)

On note toujours D_n le nombre de permutations d'un ensemble à n éléments n'ayant pas de point fixe.

1. Calculer $\sum_{k=0}^{n} C_n^k D_k$.

2. On considère la série entière $\sum_{k=0}^{+\infty}\frac{D_k z^k}{k!}$, (on l'appelle série *génératrice exponentielle* de la suite $(D_n)_{n\in\mathbb{N}}$). On note D sa somme. Minorer son rayon de convergence R. Calculer $D(z)$ pour $|z| < R$.

3. En déduire que D_k est la partie entière de $\frac{k!}{e}+\frac{1}{2}$.

(École polytechnique)

▷ **Solution.**

1. Voir la solution dans l'exercice précédent : $\sum_{k=0}^{n} C_n^k D_k = n!$.

2. On a évidemment $0 \leqslant \frac{D_k}{k!} \leqslant 1$, donc le rayon de la série entière définissant $D(z)$ est supérieur ou égal à 1.

Pour calculer $D(z)$ on utilise la relation précédente. Elle peut s'écrire, pour tout $n \in \mathbb{N}$, $n! = \sum_{k=0}^{n}\frac{n!}{k!(n-k)!}D_k$, ou encore $1 = \sum_{k=0}^{n}\frac{D_k}{k!}\frac{1}{(n-k)!}$. On reconnaît le coefficient d'ordre n du produit de Cauchy des séries entières $\sum\frac{D_k}{k!}z^k$ et $\sum\frac{1}{k!}z^k$. On en déduit que pour $|z| < 1$,

$$D(z)e^z = \sum_{n=0}^{+\infty} z^n = \frac{1}{1-z}$$

d'où l'on tire, $\mathrm{D}(z) = \dfrac{e^{-z}}{1-z}$ (c'est ici que s'effectue l'inversion). Il ne reste plus qu'à développer en série entière pour obtenir D_k. On a, pour tout $z \in \mathbb{C}$, $e^{-z} = \sum_{k=0}^{+\infty} \dfrac{(-1)^k}{k!} z^k$. Il en résulte que pour tout $k \in \mathbb{N}$, $\dfrac{\mathrm{D}_k}{k!} = \sum_{p=0}^{k} \dfrac{(-1)^p}{p!}$. C'est bien le résultat obtenu dans l'exercice précédent.

3. On a, d'après ce qui précède, $\dfrac{1}{e} = \sum_{p=0}^{+\infty} \dfrac{(-1)^p}{p!} = \dfrac{\mathrm{D}_k}{k!} + \mathrm{R}_k$, où $\mathrm{R}_k = \sum_{p=k+1}^{+\infty} \dfrac{(-1)^p}{p!}$. Ainsi, $\dfrac{k!}{e} + \dfrac{1}{2} = \mathrm{D}_k + \dfrac{1}{2} + k!\mathrm{R}_k$. La série $\sum \dfrac{(-1)^p}{p!}$ vérifiant le critère des séries alternées, on sait que $|\mathrm{R}_k| \leqslant \dfrac{1}{(k+1)!}$. Pour $k \geqslant 1$, on a donc $|k!\mathrm{R}_k| < \dfrac{1}{k+1} \leqslant \dfrac{1}{2}$ et donc

$$\boxed{\mathrm{D}_k = \mathrm{E}\left(\frac{k!}{e} + \frac{1}{2}\right)}.$$

La formule est par ailleurs aussi vérifiée au rang $k = 0$. ◁

Remarquons que la proportion de dérangements $\dfrac{\mathrm{D}_n}{n!}$ *tend vers le nombre* $\dfrac{1}{e} = 0{,}368...$ *lorsque* n *tend vers l'infini. En particulier le rayon de convergence de la série entière* D *ci-dessus est* e.

Dans l'exercice suivant on établit de manière combinatoire, une autre relation de récurrence qui permet le calcul de D_n.

1.4. Nombre de dérangements (3)

Les notations sont les mêmes que dans les deux exercices précédents.

1. Établir, par une preuve combinatoire, que pour tout $n \geqslant 2$,

$$\mathrm{D}_{n+1} = n(\mathrm{D}_n + \mathrm{D}_{n-1}).$$

2. En déduire que pour tout $n \geqslant 2$, $\mathrm{D}_n = n\mathrm{D}_{n-1} + (-1)^n$.

3. Retrouver la valeur de D_n.

(École polytechnique)

▷ **Solution.**

1. Soit $\mathcal{D}_n$ l'ensemble des dérangements de $[\![1, n]\!]$. Si $\sigma \in \mathcal{D}_{n+1}$, l'image de $n+1$ par σ est dans $[\![1, n]\!]$. Pour tout $k \in [\![1, n]\!]$, on note

F_k l'ensemble des éléments σ de $\mathcal{D}_{n+1}$ tels que $\sigma(n+1) = k$. Les F_k partitionnent $\mathcal{D}_{n+1}$.

Nous allons prouver que le cardinal de F_k est égal à $\mathrm{D}_n + \mathrm{D}_{n-1}$ quel que soit l'entier k. On fixe donc k et on partitionne F_k en deux : G_k désigne l'ensemble des $\sigma \in \mathrm{F}_k$ tels que $\sigma(k) = n+1$ et H_k désigne le complémentaire de G_k dans F_k. Il est clair qu'un élément de G_k est parfaitement déterminé par sa restriction à l'ensemble $\{1, \dots, k-1, k+1, \dots, n\}$, cette restriction devant être un dérangement. Il en résulte que $\mathrm{Card}(\mathrm{G}_k) = \mathrm{D}_{n-1}$. Cherchons maintenant le cardinal de H_k. Soit $\sigma \in \mathrm{H}_k$. Alors $\sigma^{-1}(n+1) \neq k$. Considérons le dérangement $\tilde{\sigma}$ de $[\![1, n]\!]$ défini par $\tilde{\sigma}(i) = \sigma(i)$ pour $i \neq \sigma^{-1}(n+1)$, et $\tilde{\sigma}(\sigma^{-1}(n+1)) = k$ (on a simplement «court-circuité» $n+1$). Il est clair que l'application de H_k dans $\mathcal{D}_n$ qui à σ associe $\tilde{\sigma}$ est bijective de sorte que $\mathrm{Card}(\mathrm{H}_k) = \mathrm{D}_n$. On obtient donc $\mathrm{Card}(\mathrm{F}_k) = \mathrm{D}_{n-1} + \mathrm{D}_n$, pour tout k, ce qui donne la relation de récurrence $\mathrm{D}_{n+1} = n(\mathrm{D}_n + \mathrm{D}_{n-1})$.

2. On effectue une récurrence sur n, le résultat étant vrai pour $n = 2$ puisque $\mathrm{D}_2 = 1$ et $\mathrm{D}_1 = 0$. Supposons la formule vraie au rang n. On a alors

$$\begin{aligned}\mathrm{D}_{n+1} &= n(\mathrm{D}_n + \mathrm{D}_{n-1}) = n\mathrm{D}_n + n\mathrm{D}_{n-1} \\ &= n\mathrm{D}_n + \mathrm{D}_n - (-1)^n = (n+1)\mathrm{D}_n + (-1)^{n+1},\end{aligned}$$

ce qui est la formule au rang $n+1$.

3. En divisant la relation précédente par $n!$ il vient, pour tout $n \geqslant 2$, $\dfrac{\mathrm{D}_n}{n!} = \dfrac{\mathrm{D}_{n-1}}{(n-1)!} + \dfrac{(-1)^n}{n!}$. En sommant cette relation pour les indices $2, 3, \dots, n$, on retrouve la formule

$$\mathrm{D}_n = n! \left(\sum_{k=0}^{n} \frac{(-1)^k}{k!} \right). \quad \lhd$$

Notons que les relations des questions 1 et 2 s'obtiennent facilement à partir de la série entière D *de l'exercice précédent. Il suffit pour cela de dériver la relation* $(1-z)\mathrm{D}(z) = e^{-z}$. *Dans son ouvrage de combinatoire*[3], *Comtet suggère l'existence d'une preuve combinatoire directe de la relation de la question 2. Cela ne semble toutefois pas très facile...*

Dans l'exercice suivant, on rencontre les célèbres nombres de Catalan à travers le problème des parenthésages. C'est un exemple où l'emploi des séries génératrices se révèle incontournable.

3. COMTET (L.), *Analyse combinatoire*, tome 2. PUF, 1970, p. 11.

1.5. Nombres de Catalan

On considère un ensemble muni d'une loi de composition interne non associative et n éléments $a_1, \ldots, a_n$ de cet ensemble. Combien de valeurs différentes peut prendre le produit $a_1 a_2 \ldots a_n$?

(École Polytechnique)

▷ **Solution.**

Commençons par préciser l'énoncé. Pour calculer un produit de n éléments $a_1, \ldots, a_n$ on doit indiquer l'ordre dans lequel faire les opérations à l'aide d'un parenthésage, comme par exemple $((a_1(a_2a_3))a_4)$ pour $n = 4$. Bien entendu, pour des éléments donnés, il se peut que deux parenthésages distincts conduisent au même résultat : ainsi pour une loi associative, comme l'addition sur $\mathbb{R}$, tous les parenthésages donnent la même valeur ! L'énoncé demande en fait de calculer le nombre total de parenthésages du produit $a_1 a_2 \ldots a_n$. Il est en effet facile de construire une loi de composition interne pour laquelle deux parenthésages distincts conduisent toujours à deux produits distincts.

Nous noterons $\mathcal{C}_n$ ce nombre de parenthésages. On a clairement $\mathcal{C}_1 = \mathcal{C}_2 = 1$ et pour $n = 3$, on peut écrire $(a_1a_2)a_3$ ou $a_1(a_2a_3)$, donc $\mathcal{C}_3 = 2$. Pour $n \geqslant 2$, tout produit des a_i s'écrit d'une manière et d'une seule $(a_1 \ldots a_k)(a_{k+1} \ldots a_n)$ avec $1 \leqslant k \leqslant n-1$. En effet, il suffit de considérer le dernier produit effectué. Les deux termes qui forment ce dernier produit sont bien définis. Puisqu'il y a respectivement $\mathcal{C}_k$ parenthésages possibles pour le premier facteur $a_1 \ldots a_k$ et $\mathcal{C}_{n-k}$ pour le second facteur $a_{k+1} \ldots a_n$, il y a $\mathcal{C}_k\mathcal{C}_{n-k}$ parenthésages possibles où le premier facteur dans le dernier produit effectué contient k termes. On obtient donc

$$\mathcal{C}_n = \sum_{k=1}^{n-1} \mathcal{C}_k \mathcal{C}_{n-k}.$$

Cette relation fait penser à un produit de Cauchy. Il convient d'observer tout de même que la relation n'est pas valable pour $n = 1$. Considérons $f(x) = \sum_{n=1}^{+\infty} \mathcal{C}_n x^n$ la série entière génératrice des $\mathcal{C}_n$.

Analyse. Supposons que son rayon de convergence R soit différent de 0. Dans ces conditions, pour $x \in \mathbb{R}$, $|x| < \mathrm{R}$ on a

$$f(x)^2 = \sum_{n=2}^{+\infty} \left(\sum_{k=1}^{n-1} \mathcal{C}_k \mathcal{C}_{n-k} \right) x^n = \sum_{n=2}^{+\infty} \mathcal{C}_n x^n = f(x) - x.$$

Cette équation $f(x)^2 - f(x) + x = 0$ du second degré en $f(x)$ a un discriminant $\Delta = 1 - 4x$ et pour $x \leqslant \dfrac{1}{4}$, on aura donc $f(x) = \dfrac{1 + \varepsilon(x)\sqrt{1-4x}}{2}$.

avec $\varepsilon(x) = \pm 1$. Comme f est continue sur son intervalle ouvert de définition, il en est de même de ε qui est donc constante. Comme $f(0) = 0$, on a $\varepsilon(0) = -1$ et donc

$$f(x) = \frac{1 - \sqrt{1-4x}}{2}.$$

Synthèse. Considérons la fonction $f : x \longmapsto \dfrac{1-\sqrt{1-4x}}{2}$ qui est développable en série entière en 0 avec un rayon de convergence égal à $\dfrac{1}{4}$. Elle vérifie $f(x)^2 - f(x) + x = 0$. Si on écrit $f(x) = \sum_{n=1}^{+\infty} \mathrm{C}_n x^n$ son développement en série entière, la relation $f(x)^2 = f(x) - x$ donne, par unicité des coefficients, pour $n \geqslant 2$

$$\mathrm{C}_n = \sum_{k=1}^{n-1} \mathrm{C}_k \mathrm{C}_{n-k}.$$

Comme $f(x) = x + o(x)$ en 0, on en déduit que $\mathrm{C}_1 = 1 = \mathcal{C}_1$ et finalement par récurrence que $\mathrm{C}_n = \mathcal{C}_n$ pour tout $n \in \mathbb{N}^*$ et $\mathrm{R} = \dfrac{1}{4}$.

Pour l'expression de $\mathcal{C}_n$, dérivons f. Pour $|x| < \dfrac{1}{4}$, on a

$$f'(x) = \frac{1}{\sqrt{1-4x}} = (1-4x)^{-1/2}.$$

Le développement classique de $(1-u)^\alpha$ donne en l'occurrence

$$\begin{aligned} f'(x) &= \sum_{n=0}^{+\infty} \left(-\frac{1}{2}\right)\left(-\frac{3}{2}\right)\cdots\left(-\frac{2n-1}{2}\right)\frac{(-4x)^n}{n!} \\ &= \sum_{n=0}^{+\infty} \frac{1}{2^n}\frac{(2n)!}{2.4\cdots(2n-2).(2n)}\frac{4^n x^n}{n!} = \sum_{n=0}^{+\infty} \frac{(2n)!}{4^n (n!)^2} 4^n x^n \\ &= \sum_{n=0}^{+\infty} \mathrm{C}_{2n}^n x^n. \end{aligned}$$

Comme $f(0) = 0$, on en déduit que $f(x) = \sum_{n=0}^{+\infty} \dfrac{\mathrm{C}_{2n}^n}{n+1} x^{n+1}$. On aboutit à

$$\boxed{\mathcal{C}_n = \frac{\mathrm{C}_{2n-2}^{n-1}}{n}}. \quad \triangleleft$$

Les nombres de Catalan se rencontrent dans de nombreux problèmes de combinatoire. Par exemple, $\mathcal{C}_n$ est aussi le nombre de manières de

trianguler un polygone régulier à $n+1$ côtés ou encore le nombre d'arbres binaires à $n-1$ sommets.

L'exercice suivant s'intéresse au nombre de partitions d'un ensemble fini. C'est aussi le nombre de relations d'équivalence sur cet ensemble. On y fait encore usage de séries génératrices.

1.6. Nombres de Bell

Pour tout $n \in \mathbb{N}^*$, on note B_n le nombre de partitions de l'ensemble $[\![1, n]\!]$, avec par convention, $B_0 = 1$.

1. Calculer B_1, B_2, B_3.

Démontrer que, pour tout $n \in \mathbb{N}$, on a $B_{n+1} = \sum_{k=0}^{n} C_n^k B_k$.

2. On pose $f(z) = \sum_{n=0}^{+\infty} \frac{B_n}{n!} z^n$ (série génératrice exponentielle de la suite $(B_n)_{n\geqslant 0}$). Montrer que le rayon de convergence R de cette série entière n'est pas nul. Calculer $f(z)$ pour $z \in]-R, R[$.

3. Exprimer B_n comme somme d'une série.

(École polytechnique)

▷ **Solution.**

1. On obtient $B_1 = 1$, $B_2 = 2$, $B_3 = 5$. Pour $k \in [\![0, n]\!]$, considérons l'ensemble E_k des partitions de $[\![1, n+1]\!]$ pour lesquelles la partie de $[\![1, n+1[\![$ contenant $n+1$ est de cardinal $k+1$. On a $\operatorname{Card} E_k = C_n^k B_{n-k}$. En effet, pour constituer la partie contenant $n+1$, il faut choisir k élément dans $[\![1, n]\!]$, puis il faut réaliser une partition des $n-k$ éléments restants. Comme $E_0, E_1, \ldots, E_n$ forment une partition de l'ensemble des partitions de $[\![1, n+1]\!]$, on obtient

$$B_{n+1} = \sum_{k=0}^{n} C_n^k B_{n-k} = \sum_{j=0}^{n} C_n^j B_j,$$

en faisant le changement d'indice $j = n-k$ dans la seconde égalité.

2. Pour minorer le rayon de convergence de la série $\sum_{n\geqslant 0} \frac{B_n}{n!} z^n$ il faut majorer B_n. Montrons donc, par récurrence sur $n \in \mathbb{N}$, que $B_n \leqslant n!$. C'est vrai pour $n \leqslant 3$ et si la propriété est vérifiée jusqu'au rang n, alors

$$B_{n+1} \leqslant \sum_{k=0}^{n} C_n^k k! = n! \sum_{k=0}^{n} \frac{1}{(n-k)!} \leqslant (n+1)!.$$

On a donc, pour $n \in \mathbb{N}^*$, $\frac{B_n}{n!} \leqslant 1$ et le rayon de convergence R de la série entière est supérieur ou égal à 1.

Calculons $f(z)$, pour $z \in]-\mathrm{R}, \mathrm{R}[$ en utilisant la relation de récurrence démontrée dans 1. On a, pour $z \in]-\mathrm{R}, \mathrm{R}[$, $f(z) = 1 + \sum_{k=0}^{+\infty} \frac{\mathrm{B}_{n+1}}{(n+1)!} z^{n+1}$. La fonction f est dérivable sur $]-\mathrm{R}, \mathrm{R}[$ et $f'(z) = \sum_{n=0}^{+\infty} \frac{\mathrm{B}_{n+1}}{n!} z^n$. On en déduit que, pour $z \in]-\mathrm{R}, \mathrm{R}[$, on a

$$f'(z) = \sum_{n=0}^{+\infty} \frac{1}{n!} \left(\sum_{k=0}^{n} \mathrm{C}_n^k \mathrm{B}_k \right) z^n = \sum_{n=0}^{+\infty} \left(\sum_{k=0}^{n} \frac{\mathrm{B}_k}{k!} \frac{1}{(n-k)!} \right) z^n.$$

On reconnaît dans le terme de droite un produit de Cauchy, celui des séries $\sum \frac{\mathrm{B}_n}{n!} z^n$ et $\sum \frac{z^n}{n!}$. La première a pour somme $f(z)$ et la seconde e^z. Elles ont toutes deux un rayon de couvergence supérieur ou égal à R. On a donc, pour $z \in]-\mathrm{R}, \mathrm{R}[$, $f'(z) = f(z)e^z$. On en déduit qu'il existe $\mathrm{C} \in \mathbb{R}$ tel que, pour $z \in]-\mathrm{R}, \mathrm{R}[$, $f(z) = \mathrm{C}e^{e^z}$. Sachant que $f(0) = \mathrm{B}_0 = 1$, on en déduit que $\mathrm{C} = \frac{1}{e}$. Finalement, on obtient pour $z \in]-\mathrm{R}, \mathrm{R}[$,

$$\boxed{f(z) = \frac{1}{e} e^{e^z} = e^{e^z - 1}}.$$

3. La série entière définissant la fonction exponentielle ayant un rayon de convergence infini, on a, pour tout $z \in \mathbb{C}$,

$$e^{e^z} = \sum_{n=0}^{+\infty} \frac{e^{nz}}{n!} = \sum_{n=0}^{+\infty} \frac{1}{n!} \sum_{k=0}^{+\infty} \frac{(nz)^k}{k!}.$$

Considérons la série double $(u_{n,k})_{(n,k)\in\mathbb{N}^2}$ définie par $u_{n,k} = \frac{(nz)^k}{n!k!}$. On a, pour $n \in \mathbb{N}$,

$$\sum_{k=0}^{+\infty} |u_{n,k}| = \sum_{k=0}^{+\infty} \frac{|nz|^k}{n!k!} = \frac{e^{|nz|}}{n!},$$

puis

$$\sum_{n=0}^{+\infty} \frac{e^{|nz|}}{n!} = \sum_{n=0}^{+\infty} \frac{\left(e^{|z|}\right)^n}{n!} = e^{e^{|z|}}.$$

La série double est donc sommable, pour tout $z \in \mathbb{C}$. On peut donc échanger l'ordre des sommations et en déduire que, pour tout z de $]-\mathrm{R}, \mathrm{R}[$,

$$f(z) = \frac{1}{e} \sum_{n=0}^{+\infty} \left(\sum_{k=0}^{+\infty} u_{n,k} \right) = \frac{1}{e} \sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} u_{n,k} \right) = \sum_{k=0}^{+\infty} \left(\sum_{n=0}^{+\infty} \frac{n^k}{n!} \right) \frac{z^k}{k!}.$$

D'après l'unicité du développement en série entière de f, on obtient, pour tout $k \in \mathbb{N}$,

$$\boxed{\mathrm{B}_k = \frac{1}{e}\sum_{n=0}^{+\infty} \frac{n^k}{n!}}. \quad \triangleleft$$

Il résulte des calculs précédents que le rayon de convergence de la série entière définissant f est en fait infini. Cette formule peut servir de point de départ à l'étude asymptotique de la suite B_n. Le lecteur intéressé pourra par exemple étudier la première épreuve du concours Mines-Ponts de 2002 qui se propose d'obtenir un équivalent de B_n.

Voici maintenant un petit exercice facile qui concerne les relations d'équivalence.

1.7. Cardinal d'une relation d'équivalence

Soit A un ensemble de cardinal n, $\mathcal{R}$ une relation d'équivalence sur A avec k classes. Soit m le cardinal du graphe de $\mathcal{R}$. Montrer que $n^2 \leqslant km$.

(École polytechnique)

▷ **Solution.**

Notons $\mathrm{A}_1, \ldots, \mathrm{A}_k$ les différentes classes d'équivalence. On va commencer par expliciter les quantités n et m. Comme les classes d'équivalence partitionnent A on a déjà

$$n = \operatorname{Card} \mathrm{A} = \sum_{i=1}^{k} \operatorname{Card} \mathrm{A}_i.$$

D'autre part, si $\overline{x}$ désigne la classe d'équivalence d'un élément x de A, on a

$$m = \operatorname{Card}\{(x,y) \in \mathrm{A}^2,\ (x,y) \in \mathcal{R}\} = \sum_{x \in \mathrm{A}} \operatorname{Card} \overline{x}$$

$$= \sum_{i=1}^{k} \sum_{x \in \mathrm{A}_i} \operatorname{Card} \mathrm{A}_i = \sum_{i=1}^{k} (\operatorname{Card} \mathrm{A}_i)^2.$$

L'inégalité demandée résulte alors de l'inégalité de Cauchy-Schwarz. $\triangleleft$

Le thème commun aux exercices qui suivent est l'algèbre linéaire sur des corps finis.

1.8. Cardinal de $\mathrm{GL}_n(\mathrm{K})$ et $\mathrm{SL}_n(\mathrm{K})$

Soit K un corps commutatif fini de cardinal q. Déterminer le cardinal de $\mathrm{GL}_n(\mathrm{K})$ et celui de $\mathrm{SL}_n(\mathrm{K})$.

(École normale supérieure)

Solution.

Choisir une matrice inversible de taille n revient à choisir la famille $(\mathrm{C}_1, \ldots, \mathrm{C}_n)$ de ses n colonnes, famille qui constitue une base de K^n. Pour le choix de C_1, on a donc $q^n - 1$ possibilités : elle doit simplement être non nulle. Pour chacun de ces choix, on a $q^n - q$ possibilités pour la colonne C_2 (elle ne doit pas appartenir à la droite engendrée par C_1). Puis on a $q^n - q^2$ possibilités pour C_3 (elle ne doit pas être dans le plan $\mathrm{KC}_1 \oplus \mathrm{KC}_2$). Plus généralement, si $\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_{k-1}$ sont fixées, on a $q^n - q^{k-1}$ choix possibles pour C_k. En effet, il doit simplement être en dehors de l'espace $\mathrm{KC}_1 \oplus \cdots \oplus \mathrm{KC}_{k-1}$ et cet espace de dimension $k-1$ est de cardinal q^{k-1}. Finalement, on obtient

$$\boxed{|\mathrm{GL}_n(\mathrm{K})| = (q^n - 1)(q^n - q)(q^n - q^2)\ldots(q^n - q^{n-1})}.$$

L'application det : $\mathrm{GL}_n(\mathrm{K}) \to \mathrm{K}^*$ est un morphisme surjectif de groupes de noyau $\mathrm{SL}_n(\mathrm{K})$. Nous savons alors (cela est redémontré dans l'exercice 2.4) que $|\operatorname{Im}\det| \times |\operatorname{Ker}\det| = |\mathrm{GL}_n(\mathrm{K})|$, de sorte que

$$\boxed{|\mathrm{SL}_n(\mathrm{K})| = \frac{(q^n - 1)(q^n - q)(q^n - q^2)\ldots(q^n - q^{n-1})}{q-1}}.\triangleleft$$

Dans l'exercice suivant on s'intéresse à l'analogue du groupe spécial orthogonal sur le corps $\mathbb{Z}/p\mathbb{Z}$.

1.9. Cardinal de $\mathrm{SO}_2(\mathbb{Z}/p\mathbb{Z})$

Soit p un nombre premier. On note $\mathrm{SO}_2(\mathbb{Z}/p\mathbb{Z})$ l'ensemble des matrices M de $\mathcal{M}_2(\mathbb{Z}/p\mathbb{Z})$ telles que ${}^t\mathrm{MM} = \mathrm{I}_2$ et $\det \mathrm{M} = 1$. Soit u_p le cardinal de $\mathrm{SO}_2(\mathbb{Z}/p\mathbb{Z})$. Montrer que $u_2 = 2$, que $u_p = p - 1$ si $p \equiv 1\ [4]$ et enfin que $u_p = p + 1$ si $p \equiv 3\ [4]$.

(École normale supérieure)

Solution.

Soit $\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{Z}/p\mathbb{Z})$. Alors $\mathrm{M} \in \mathrm{SO}_2(\mathbb{Z}/p\mathbb{Z})$ si et

seulement si $\det M = 1$ et $M^{-1} = {}^tM$. Or, si $\det M = 1$, M^{-1} est égale à la transposée de la comatrice de M, c'est-à-dire à la matrice $M' = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Ainsi, $M' = {}^tM$ équivaut à $a = d$ et $b = -c$. Finalement, M appartient à $SO_2(\mathbb{Z}/p\mathbb{Z})$ si et seulement si (a, b, c, d) vérifient le système

$$\begin{cases} a = d \\ b = c \\ ad - bc = 1, \end{cases} \quad \text{qui équivaut au système} \quad \begin{cases} a = d \\ b = c \\ a^2 + b^2 = 1. \end{cases}$$

On a donc une bijection entre $SO_2(\mathbb{Z}/p\mathbb{Z})$ et l'ensemble (qu'on a envie d'appeler «cercle») $S = \{(a, b) \in (\mathbb{Z}/p\mathbb{Z})^2,\ a^2 + b^2 = 1\}$. Dans le cas réel, S est le cercle unité de $\mathbb{R}^2$ et en le paramétrant par $(\cos\theta, \sin\theta)$, $\theta \in \mathbb{R}$, on obtient l'écriture classique des matrices de rotation de taille $(2, 2)$. Évidemment ce paramétrage ne va pas nous servir ici. Cependant, le cercle admet un paramétrage rationnel très important qu'on retrouve à partir du précédent grâce aux formules trigonométriques donnant $\cos\theta$ et $\sin\theta$ en fonction de $t = \tan\dfrac{\theta}{2}$:

$$\cos\theta = \frac{1 - t^2}{1 + t^2} \text{ et } \sin\theta = \frac{2t}{1 + t^2}.$$

Plus précisément, lorsque t décrit $\mathbb{R}$, le point $M(t)$ de coordonnées $\left(\dfrac{1 - t^2}{1 + t^2}, \dfrac{2t}{1 + t^2}\right)$ décrit bijectivement le cercle S privé du point $(-1, 0)$ (car $\theta = 2\arctan t$ décrit bijectivement l'intervalle $]-\pi, \pi[$).

On va examiner si on a encore cette bijection dans le cas du corps $\mathbb{Z}/p\mathbb{Z}$. Un problème survient du fait que la quantité $1 + t^2$ peut s'annuler (si -1 est un carré dans $\mathbb{Z}/p\mathbb{Z}$). Laissons de côté le cas $p = 2$ qui se traite directement : on a aisément dans ce cas $S = \{(0, 1); (1, 0)\}$ et donc $u_2 = 2$. On suppose dans la suite que p est un nombre premier impair et on distingue deux cas.

• Supposons que -1 n'est pas un carré dans $\mathbb{Z}/p\mathbb{Z}$, c'est-à-dire que $p \equiv 3\ [4]$ (voir exercice 4.34). Pour tout élément t de $\mathbb{Z}/p\mathbb{Z}$ on peut considérer le point $A(t) = \left(\dfrac{1 - t^2}{1 + t^2}, \dfrac{2t}{1 + t^2}\right)$. Il est aisé de vérifier que $A(t) \in S$ et que $A(t) \neq (-1, 0)$. Montrons que A est injective. Soit t, t' tels que $A(t) = A(t')$. On a $\dfrac{1 - t^2}{1 + t^2} = \dfrac{1 - t'^2}{1 + t'^2}$, ce qui conduit à $2t^2 = 2t'^2$. Comme 2 est inversible dans $\mathbb{Z}/p\mathbb{Z}$, on a $t^2 = t'^2$. En regardant les secondes coordonnées des points $A(t)$ et $A(t')$, on obtient alors $2t = 2t'$ soit $t = t'$. Donc A est une application injective. Soit $(x, y) \in S$ un point distinct de $(-1, 0)$. On cherche t tel que $A(t) = (x, y)$. L'équation

$\frac{t^2}{1+t^2} = x$ conduit à $t^2 = \frac{1-x}{1+x}$, ce qui est possible puisque $x \neq -1$. En substituant dans l'égalité $\frac{2t}{1+t^2} = y$, on obtient $t = \frac{y}{1+x}$. On vérifie réciproquement, que pour cette valeur de t on a $A(t) = (x, y)$. On en déduit que

$$u_p = |\mathbb{Z}/p\mathbb{Z}| + 1 = p + 1.$$

• Supposons maintenant que -1 est un carré dans $\mathbb{Z}/p\mathbb{Z}$, c'est-à-dire que $p \equiv 3\ [4]$ et notons $\pm\varepsilon$ ses deux racines carrées. L'application A restreinte à $\mathbb{Z}/p\mathbb{Z} \setminus \{\pm\varepsilon\}$ reste bien entendu injective et à valeurs dans $S \setminus \{(-1, 0)\}$. Elle reste en fait surjective, car si $(x, y) \in S \setminus \{(-1, 0)\}$, le réel $t = \frac{y}{1+x}$ obtenu précédemment vérifie $t^2 = \frac{1-x}{1+x} \neq -1$ de sorte que $t \notin \{\pm\varepsilon\}$. On a donc dans ce cas,

$$u_p = |\mathbb{Z}/p\mathbb{Z} \setminus \{\pm\varepsilon\}| + 1 = p - 1.$$

D'où le résultat. ◁

L'exercice suivant montre un exemple d'utilisation des opérations de groupes dans un problème combinatoire.

1.10. Nombre d'involutions

Soit K un corps fini de caractéristique différente de 2. Déterminer le nombre de matrices $A \in GL_n(K)$ telles que $A^2 = I_n$.

(**École normale supérieure**)

Solution.

• Soit $A \in GL_n(K)$ telle que $A^2 = I_n$. Comme la caractéristique de K est différente de 2, le polynôme $X^2 - 1$ est scindé à racines simples sur K. Il en résulte que A est diagonalisable avec $K^n = \mathrm{Ker}(A - I_n) \oplus \mathrm{Ker}(A + I_n)$. Autrement dit, A est la matrice de la symétrie par rapport au sous-espace $F = \mathrm{Ker}(A - I_n)$ parallèlement au sous-espace $G = \mathrm{Ker}(A + I_n)$. Soit alors f l'application qui à cette matrice A associe le couple des sous-espaces supplémentaires (F, G). Il est clair que f est injective (A est parfaitement déterminée par la connaissance des sous-espaces F et G) et que son image est exactement l'ensemble des couples de sous-espaces supplémentaires. L'exercice revient donc à dénombrer ces couples.

• Pour $p \in [\![0, n]\!]$ on note E_p l'ensemble des couples (F, G) où $\dim F = p$ et $F \oplus G = K^n$. Le groupe $GL_n(K)$ opère naturellement sur E_p et cette opération est clairement transitive. Notons $(e_1, \ldots, e_n)$ la base canonique de K^n, $F_p = \mathrm{Vect}(e_1, \ldots, e_p)$ et $G_p = \mathrm{Vect}(e_{p+1}, \ldots, e_n)$. On a $(F_p, G_p) \in E_p$. Le cardinal de E_p est égal à l'indice du stabilisateur

de l'élément (F_p, G_p) dans $GL_n(K)$. Or, une matrice de $GL_n(K)$ stabilise (F_p, G_p) si et seulement si elle est de la forme

$$\begin{pmatrix} X & 0 \\ 0 & Y \end{pmatrix}$$

où $X \in GL_p(K)$ et où $Y \in GL_{n-p}(K)$ (évidemment, si $p = 0$ ou $p = n$, l'un des deux éléments disparaît). On a donc

$$\boxed{\mathrm{Card}\{A \in GL_n(K), A^2 = I\} = \sum_{p=0}^{n} \frac{|GL_n(K)|}{|GL_p(K)|\,|GL_{n-p}(K)|}},$$

où par convention $|GL_0(K)| = 1$. On peut conclure en remplaçant les cardinaux des groupes linéaires par la valeur déterminée dans l'exercice 1.8. ◁

L'exercice qui suit concerne un sujet qui forme un pan entier de la combinatoire : la théorie des partitions. On appelle partition d'un entier naturel n toute décomposition de n sous la forme $n = a_1 + a_2 + \cdots + a_k$ où $0 < a_1 \leqslant a_2 \leqslant \cdots \leqslant a_k$. Les a_i sont appelés les parts de la partition. L'entier 5 admet ainsi 7 partitions qui sont $5 = 1+4 = 2+3 = 1+1+3 = 1+2+2 = 1+1+1+2 = 1+1+1+1+1$. Cette notion est étudiée depuis Euler à qui l'on doit déjà de très beaux résultats, par exemple le fait qu'un entier admet autant de partitions en parts distinctes que de partitions en parts impaires. Ce résultat peut s'établir à l'aide de la série génératrice $\sum p(n)z^n$ où on note usuellement $p(n)$ le nombre de partitions de n. A l'aide de l'analyse complexe, on peut démontrer la formule de Ramanujan (1887-1920)

$$\ln p(n) \sim \pi\sqrt{\frac{2n}{3}}.$$

L'exercice qui suit est un joli problème extrémal sur les partitions.

1.11. Partitions d'un entier

1. Pour quelles partitions de n sous la forme $n = a_1 + \cdots + a_k$ (où les a_i sont dans $\mathbb{N}^*$), le produit $\prod_{i=1}^{k} a_i$ est-il maximal ?

2. Résoudre le même problème si on considère seulement les partitions constituées d'entiers 2 à 2 distincts.

(**École normale supérieure**)

Solution.

1. Voici pour les premiers entiers les partitions réalisant le produit maximal :

$2 \quad 2;3=3;4=2+2=4;5=2+3;6=3+3;7=2+2+3=3+4;$
$8=2+3+3;9=3+3+3+;\ldots$

Soit n un entier que l'on suppose supérieur à 5. L'ensemble des partitions de n étant fini, on est certain d'en trouver au moins une donnant un produit maximal : $n=a_1+a_2+\cdots+a_k$ avec $a_1\leqslant a_2\leqslant\cdots\leqslant a_k$. Étudions les propriétés que doit avoir une telle partition.

• Les a_i sont inférieurs ou égaux à 4. En effet, si $a_k\geqslant 5$ on peut le remplacer par $(a_k-3)+3$ et comme $3(a_k-3)>a_k$ (car $2a_k>9$), on a une partition de n donnant un produit strictement plus grand. C'est absurde.

• On peut remplacer les coefficients égaux à 4 par $2+2$ sans changer le produit. On peut donc supposer $a_k\leqslant 3$.

• On a $a_1\geqslant 2$. En effet, si $a_1=1$, la partition $(1+a_2)+a_3+\cdots+a_k$ donne un produit strictement plus grand.

• Il y a au plus deux a_i égaux à 2. En effet, s'il y en a trois, le terme $2+2+2$ peut être remplacé par $3+3$ qui donne encore un produit supérieur.

Conclusion.

⋆ si $n\equiv 0\ [3]$, $n=3k$ $(k\geqslant 1)$, l'unique partition de produit maximal est $n=3+3+\cdots+3$ avec k fois le terme 3.

⋆ si $n\equiv 1\ [3]$, $n=3k+1$ $(k\geqslant 1)$, les partitions de produit maximal sont $n=2+2+3+3+\cdots+3=4+3+3+\cdots+3$ avec $(k-1)$ fois le terme 3.

⋆ si $n\equiv 2\ [3]$, $n=3k+2$ $(k\geqslant 0)$, l'unique partition de produit maximal est $n=2+3+3+\cdots+3$ avec k fois le terme 3.

2. Le problème est plus difficile mais la démarche est la même. L'existence est claire pour les mêmes raisons que ci-dessus. On commence par regarder les premiers entiers. On donne toutes les partitions en entiers distincts, celle(s) de produit maximal étant en gras.

$2=\mathbf{2}$

$3=1+2=\mathbf{3}$

$4=1+3=\mathbf{4}$

$5=1+4=\mathbf{2+3}=5$

$6=1+2+3=1+5=\mathbf{2+4}=6$

$7=1+2+4=2+5=\mathbf{3+4}=7$

$8=1+2+5=1+3+4=2+6=\mathbf{3+5}=8$

$9=1+2+6=1+3+5=\mathbf{2+3+4}=3+6=4+5=9$

$10=1+2+3+4=1+2+7=1+3+6=1+4+5=\mathbf{2+3+5}=3+7=4+6=10\ldots$

Soit $n \geqslant 5$ que l'on écrit $n = a_1 + a_2 + \cdots + a_k$, avec $a_1 < a_2 < \cdots < a_k$, où le produit est maximal. On va ici encore étudier les propriétés que doit avoir cette partition. Sur les exemples ci-dessus, il semblerait qu'il faille prendre des entiers presque consécutifs. Précisons cela.

• Fait 1 : il y a au plus un entier de l'intervalle $[\![a_1, a_k]\!]$ qui n'apparaît pas dans la partition. En effet, dans le cas contraire, on peut trouver deux éléments $a < b$ de la partition tels que $a+1$ et $b-1$ ne soient pas dans la partition et $a < a+1 < b-1 < b$. Mais si on remplace alors $a+b$ par $(a+1)+(b-1)$, le produit obtenu est strictement plus grand en vertu de l'inégalité $(a+1)(b-1) > ab$ (car $b-a > 1$), ce qui est contradictoire.

La partition est donc formée d'entiers consécutifs avec au plus un «trou». Sur les exemples ci-dessus, il semblerait aussi qu'elle commence toujours par 2 ou 3.

• Fait 2 : on a $a_1 \in \{2, 3\}$ (on a supposé $n \geqslant 5$). En effet, si $a_1 \geqslant 7$ on le remplace par $3 + (a_1 - 3)$ et le produit est strictement plus grand. Si $a_1 = 6$ on le remplace par $2 + 4$ et si $a_1 = 5$ on le remplace par $2 + 3$. Supposons que $a_1 = 4$. Si $a_2 = 5$ on remplace $a_1 + a_2 = 4 + 5$ par $2 + 3 + 4$ avec $2 \times 3 \times 4 = 24 > 4 \times 5 = 20$ et si $a_2 = 6$ on remplace $a_1 + a_2 = 4 + 6$ par $2 + 3 + 5$ avec $2 \times 3 \times 5 = 30 > 4 \times 6 = 24$. Il reste enfin à voir que a_1 ne peut être égal à 1 ce qui est évident car si on le rajoute à a_2, le produit est plus grand.

Enfin, voici une dernière propriété qui n'est pas visible sur les exemples ci-dessus mais que l'on découvre par exemple en décomposant $14 = 3 + 5 + 6 = 2 + 3 + 4 + 5$. La partition maximale est la seconde : le 6 est avantageusement remplacé par $2 + 4$. Plus généralement,

• Fait 3 : si $a_1 = 3$ et si la partition admet un «trou», celui-ci est nécessairement entre a_{k-1} et a_k (c'est-à-dire en dernière position). En effet, dans le cas contraire, la partition s'écrit

$$n = 3 + 4 + \cdots + (m-1) + (m+1) + (m+2) + \cdots + a_k,$$

où m est l'entier manquant. Mais la partition

$$n = 2 + 3 + \cdots + (m-1) + m + (m+1) + (m+3) + \cdots + a_k$$

donne un produit strictement plus grand car $2m > m + 2$, ce qui est absurde.

• On peut maintenant effectuer la synthèse de ce travail. On fixe un entier $m \geqslant 3$ et on pose $S_m = 2 + 3 + \cdots + m = \dfrac{m(m+1)}{2} - 1$. On note enfin E_m l'ensemble des suites finies strictement croissante d'entiers $s = (a_1, a_2, \ldots, a_k)$ ayant les propriétés suivantes :

(*i*) les a_i sont des entiers consécutifs avec éventuellement un «trou» ;

(*ii*) $a_1 = 2$ ou $a_1 = 3$;

(*iii*) lorsque $a_1 = 3$, s'il y a un «trou», il est entre a_{k-1} et a_k ;

(*iv*) la somme $a_1 + \cdots + a_k$ est inférieure ou égale à S_m.

L'application f qui à une suite s de E_m associe la somme de ses termes, a pour image l'intervalle $[\![2, S_m]\!]$ privé de l'entier 4 : c'est ce qui résulte de l'étude précédente.

On se propose de montrer que f est bijective, c'est-à-dire que la partition réalisant le produit maximal est unique, ce qu'on a pu observer sur les exemples ci-dessus. Pour cela il suffit de montrer que le cardinal de E_m est égal au cardinal de l'image de f à savoir $S_m - 2 = \dfrac{m^2+m-6}{2}$.

Comptons les éléments de E_m.

⋆ Il y a les suites commençant par 2. Il y a $m-1$ suites sans «trou» : (2) ; $(2,3)$; $(2,3.4)$; … ; $(2,3,\ldots,m)$.

Dans chacune de ses suites (sauf les deux premières), on peut placer un «trou» dans n'importe quelle position, ce qui à partir de la suite $(2,3,\ldots,k)$, donne $k-3$ suites supplémentaires et donc $k-2$ suites en tout. On a au total $1+1+2+3+\cdots+(m-2) = 1 + \dfrac{(m-2)(m-1)}{2}$ telles suites.

⋆ Il y a les suites commençant par 3 sans «trou» : (3) ; $(3,4)$; … ; $(3,4,\ldots,m)$. Cela fait $m-2$ suites.

⋆ Il y a enfin les suites commençant par 3 avec un «trou» en dernière place : $(3,5)$; $(3,4,6)$; … ; $(3,4,5,\ldots,m-2,m)$; mais aussi (à ne pas oublier !) la suite $(3,4,5,\ldots,m-1,m+1)$ dont la somme est $S_m - 1$. On a donc $m-3$ telles suites.

Au total, on obtient

$$|E_m| = 1 + \frac{(m-2)(m-1)}{2} + (m-2) + (m-3) = \frac{m^2+m-6}{2},$$

ce qui est le résultat voulu.

Conclusion. Tout entier $n \geqslant 2$ admet une unique partition en entiers distincts de produit maximal, partition décrite ci-dessus. ◁

Remarquons qu'il est facile de trouver la partition optimale de proche en proche. Supposons celle de n connue, $n = a_1 + \cdots + a_k$, et voyons comment obtenir celle de $n+1$.

• *Si $a_1 = 3$ et s'il n'y a pas de trou, on remplace a_k par $a_k + 1$.*

• *Si $a_1 = 3$ et s'il y a un trou,* i.e. *$n = 3 + 4 + \cdots + (m-2) + m$, la partition de $n+1$ est $n+1 = 2+3+4+\cdots+(m-2)+(m-1)$. C'est par exemple le cas pour $n = 13 = 3+4+6$; on obtient $14 = 2+3+4+5$.*

• *Si $a_1 = 2$ et s'il n'y a pas de trou, i.e. $n = 2+3+\cdots+m$, la partition de $n+1$ est $n+1 = 2+3+\cdots+(m-1)+(m+1)$, obtenu en remplaçant m par $m+1$. Pour obtenir $n+2$, on remplace $m-1$ par m, etc. Le trou se décale progressivement vers la gauche. Enfin si on a une décomposition de la forme $2+4+5+\cdots+m$ l'entier qui suit s'écrira $3+4+\cdots+m$ et on est ramené au premier point.*

L'étude de configurations d'ensembles finis obéissant à diverses contraintes forme ce qu'on appelle la théorie extrémale des ensembles. Cette définition étant assez vague, voici deux exemples simples de problèmes de ce type : quel est le nombre maximal de parties distinctes de $[\![1,n]\!]$ dont deux quelconques ont toujours une intersection non vide[4] *? Un autre exemple, moins simple, est le problème de Sperner : quel est le nombre maximal de parties distinctes de $[\![1,n]\!]$ telles qu'aucune des parties ne soit incluse dans une autre (on parle d'antichaîne pour la relation d'inclusion) ? Les problèmes posés sont souvent difficiles. En voici un exemple qui montre en outre comment un outil algébrique, ici les matrices d'incidence, permet de résoudre un problème combinatoire.*

1.12. Un problème de théorie extrémale des ensembles

Soit A un ensemble fini de cardinal $m \geqslant 2$ et $\mathrm{U}_1, \ldots, \mathrm{U}_n$ des parties non vides deux à deux distinctes de A. On suppose qu'il existe un entier $a \geqslant 0$ tel que si $i \neq j$, $\mathrm{Card}(\mathrm{U}_i \cap \mathrm{U}_j) = a$. Montrer que $n \leqslant m$.

(École normale supérieure)

▷ **Solution.**

• Notons $x_1, \ldots, x_m$ les éléments de A. On va traduire les relations d'appartenance des x_i aux ensembles U_j à l'aide d'une matrice, appelée matrice d'incidence. Soit $\mathrm{M} = (m_{ij}) \in \mathcal{M}_{m,n}(\mathbb{R})$ définie par $m_{ij} = 1$ si $x_i \in \mathrm{U}_j$ et $m_{ij} = 0$ sinon.

• Regardons comment se traduisent les hypothèses sur cette matrice M. Pour $i \neq j$ on a

$$\mathrm{Card}(\mathrm{U}_i \cap \mathrm{U}_j) = \sum_{k=1}^{m} m_{ki} m_{kj} = a$$

On reconnaît le terme (i,j) de la matrice ${}^t\mathrm{MM} \in \mathcal{M}_n(\mathbb{R})$. En notant d_i le cardinal de U_i, on a donc

$$\mathrm{H} = {}^t\mathrm{MM} = \begin{pmatrix} d_1 & a & \ldots & a \\ a & d_2 & \ldots & a \\ \vdots & & \ddots & \vdots \\ a & \ldots & a & d_n \end{pmatrix}.$$

4. Réponse 2^{n-1}. Cette valeur est atteinte en prenant par exemple la famille des parties qui contiennent un élément fixé. Et si $\mathrm{A}_1, \ldots, \mathrm{A}_p$ sont des parties deux à deux distinctes qui ont la propriété mentionnée, alors les complémentaires des A_i sont deux à deux distincts et distincts des A_i, de sorte que $2p \leqslant 2^n$.

On va montrer que cette matrice H est inversible. Cela permettra de conclure puisqu'alors $n = \operatorname{rg} H \leqslant \operatorname{rg}(M) \leqslant m$ fournit l'inégalité souhaitée.

Pour cela on va simplement calculer le déterminant de H, connu sous le nom de déterminant de Hurwitz. Notons $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{K}^n$ et C la colonne dont tous les coefficients sont égaux à a. On a

$$\det H = \det(C + (d_1 - a)e_1, \ldots, C + (d_n - a)e_n).$$

Si on développe $\det H$ par multilinéarité, on obtient $\prod_{i=1}^{n}(d_i - a) + \sum_{k=1}^{n} \det((d_1 - a)e_1, \ldots, (d_{k-1} - a)e_{k-1}, C, (d_{k+1} - a)e_{k+1}, \ldots, (d_n - a)e_n)$, puisque les déterminants où la colonne C intervient deux fois ou plus sont nuls. On a donc

$$\det H = \prod_{i=1}^{n}(d_i - a) + a \sum_{i=1}^{n} \prod_{j \neq i}(d_j - a)$$

Or, pour tout i, on a $a \leqslant d_i$ (car $a = \operatorname{Card}(U_i \cap U_j)$, pour tout $j \neq i$), avec égalité pour au plus un indice i (car les égalités $a = d_i = d_j$ impliquent $U_i = U_i \cap U_j = U_j$ et les ensembles U_i sont deux à deux distincts). Il en résulte que $\det H > 0$ et donc que H est inversible. ◁

L'exercice suivant est un autre exemple où l'on utilise un outil algébrique, cette fois des polynômes, pour traduire une situation ensembliste.

1.13. Ensembles définis par récurrence

On définit des sous-ensembles de $\mathbb{N}$ par $A_1 = \emptyset, B_1 = \{0\}$ et pour n entier $\geqslant 1$,

$$A_{n+1} = B_n + 1 = \{x + 1,\ x \in B_n\}$$
$$B_{n+1} = A_n \Delta B_n = (A_n \cup B_n) \setminus (A_n \cap B_n).$$

Montrer que si p est une puissance de 2, alors $B_p = B_1$.
Indication : on pourra considérer les polynômes $P_n = \sum_{k \in A_n} X^k$ et $Q_n = \sum_{k \in B_n} X^k$ et raisonner modulo 2.

(**École polytechnique**)

Solution.

• L'ensemble B_{n+1} est la différence symétrique de A_n et B_n. Regardons les ensembles pour les premières valeurs de n. On a

$$\begin{array}{ll} A_2 = \{1\} & B_2 = \{0\} \\ A_3 = \{1\} & B_3 = \{0,1\} \\ A_4 = \{1,2\} & B_4 = \{0\} \\ A_5 = \{1\} & B_5 = \{0,1,2\} \\ A_6 = \{1,2,3\} & B_6 = \{0,2\} \\ A_7 = \{1,3\} & B_7 = \{0,1,3\} \\ A_8 = \{1,2,4\} & B_8 = \{0\}. \end{array}$$

Il apparaît effectivement que $B_2 = B_4 = B_8 = B_1$.

• Suivons l'indication et posons $P_n = \sum_{k \in A_n} X^k$ et $Q_n = \sum_{k \in B_n} X^k$, en regardant ces polynômes dans $\mathbb{Z}/2\mathbb{Z}[X]$. Ces polynômes caractérisent parfaitement les parties A_n et B_n. Les relations de définition se traduisent alors par les relations

$$P_{n+1} = XQ_n \quad \text{et} \quad Q_{n+1} = P_n + Q_n, \quad \text{pour tout } n \in \mathbb{N}^*.$$

En particulier, la suite (Q_n) vérifie la relation de récurrence linéaire à deux termes : pour tout $n \geqslant 2$, $Q_{n+1} = Q_n + XQ_{n-1}$, avec $Q_1 = Q_2 = 1$. Par convention on pose $Q_0 = 0$, ce qui rend la relation encore vraie pour $n = 1$.

• On va regarder (Q_n) comme une suite du corps $K = \mathbb{Z}/2\mathbb{Z}(X)$ des fractions rationnelles à coefficients dans $\mathbb{Z}/2\mathbb{Z}$. L'équation caractéristique est $r^2 + r + X = 0$ (ne pas oublier qu'on est en caractéristique 2 donc les moins peuvent être remplacés par des plus). La théorie habituelle de l'équation du second degré ne s'applique pas en caractéristique 2 : il est impossible de passer à la forme canonique car 2 n'est pas inversible. Cela dit, il existe un sur-corps K' de K dans lequel l'équation admet deux racines α_1 et α_2. On a

$$\alpha_1 + \alpha_2 = 1 \text{ et } \alpha_1\alpha_2 = X.$$

En particulier $\alpha_1 \neq \alpha_2$ car sinon $1 = 2\alpha_1 = 0$ ce qui est impossible. On sait alors qu'il existe deux éléments A, B de K' tels que, pour tout $n \in \mathbb{N}$,

$$Q_n = A\alpha_1^n + B\alpha_2^n.$$

On détermine A et B à l'aide de Q_0 et Q_1. Il vient $A+B = 0$, c'est-à-dire $A = B$ et $A(\alpha_1 + \alpha_2) = 1$, soit finalement $A = B = 1$. On a donc, pour tout $n \in \mathbb{N}$,

$$Q_n = \alpha_1^n + \alpha_2^n.$$

• Il ne reste plus qu'à calculer α_i^n lorsque n est une puissance de 2 (pour $i = 1$ et 2). On a $\alpha_i^2 = \alpha_i + X$. Comme on est en caractéristique deux, le carré d'une somme est la somme des carrés (les doubles produits

disparaissent). On a donc $\alpha_i^4 = (\alpha_i + X)^2 = \alpha_i^2 + X^2 = \alpha_i + X + X^2$. De même, on a facilement par récurrence sur p,

$$\alpha_i^{2^p} = \alpha_i + X + X^2 + X^4 + \cdots + X^{2^{p-1}} = \alpha_i + \sum_{k=0}^{p-1} X^{2^k}.$$

On en déduit que $Q_{2^p} = \alpha_1^{2^p} + \alpha_2^{2^p} = \alpha_1 + \alpha_2 = 1$. Donc $B_{2^p} = \{1\}$. ◁

Dans l'énoncé suivant on s'intéresse aux applications d'un ensemble fini dans lui-même qui peuvent se décomposer en un produit d'applications idempotentes.

1.14. Théorème de Howie

Soit X un ensemble fini non vide et θ une application injective et non surjective d'une partie D de X dans X.

1. On suppose que $\theta(D) \subset D$. Montrer qu'il existe un nombre fini d'applications $p_1, \ldots, p_n$ de X dans X vérifiant $p_i \circ p_i = p_i$ pour tout i et telles que la restriction à D de $p_1 \circ \cdots \circ p_n$ coïncide avec θ.

2. Montrer que la conclusion de la question 1 reste vraie sans l'hypothèse $\theta(D) \subset D$.

3. Soit α une application non bijective de X dans X. Montrer qu'il existe un nombre fini d'applications $\pi_1, \ldots, \pi_n$ de X dans X vérifiant $\pi_i \circ \pi_i = \pi_i$ pour tout i et telles que $\alpha = \pi_1 \circ \cdots \circ \pi_n$. *C'est le théorème de Howie.*

(École normale supérieure)

Solution.

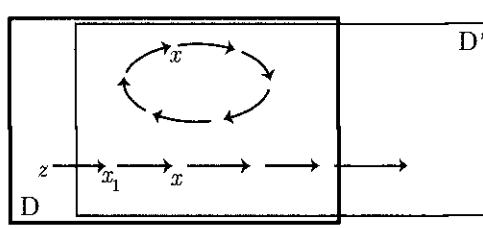
1. Comme θ est supposée injective et que $\theta(D) \subset D$, on a forcément $\theta(D) = D$ et θ établit une bijection de D sur D. Comme θ n'est pas surjective lorsqu'on la regarde comme une application de D dans X, on peut affirmer que D est une partie stricte de X. Observons qu'une application p de X dans X vérifie $p \circ p = p$ si et seulement si p fixe tous les points de son image. La question est triviale si D est vide et facile si D est un singleton $\{x\}$ (il suffit de prendre $n = 1$ et pour p_1 l'application constante égale à x). Nous supposerons que D contient au moins deux éléments. Nous savons que toute permutation de D peut se décomposer en un produit de transpositions. Il suffit donc de traiter le cas où θ est une transposition, disons $\theta = (b, c)$. Par hypothèse, il existe au moins un élément $a \in X \setminus D$. Notons p_1 l'application qui envoie b sur a et qui fixe tous les autres points de X, p_2 l'application qui envoie c sur b et qui fixe tous les autres points et enfin p_3 l'application qui envoie a sur c et qui fixe tout le reste. Il est clair que p_1, p_2, p_3 sont idempotentes (*i.e.* vérifient $p_i \circ p_i = p_i$). Et on a

$$\begin{array}{ccccccc} & p_1 & & p_2 & & p_3 & \\ b & \longrightarrow & a & \longrightarrow & a & \longrightarrow & c \\ c & \longrightarrow & c & \longrightarrow & b & \longrightarrow & b \end{array}$$

Comme par ailleurs $p_3 \circ p_2 \circ p_1$ fixe tout point de $D \setminus \{b, c\}$ on a le résultat : la restriction de $p_3 \circ p_2 \circ p_1$ à D est égale à la transposition qui échange b et c.

Notons qu'on aurait aussi pu utiliser la décomposition de la permutation θ en produit de cycles à supports disjoints avec la même idée. Pour obtenir la décomposition d'un cycle $(b_1, \ldots, b_m)$ on choisit à nouveau a en dehors de D et on considère p_1 qui envoie b_m sur a en fixant tous les autres points, p_2 qui envoie b_{m-1} sur b_m en fixant tous les autres points, jusqu'à p_m qui envoie b_1 sur b_2 en fixant le reste, et enfin p_{m+1} qui envoie a sur b_1 en fixant le reste. On vérifie comme avant que les p_i sont idempotentes et que la restriction de $p_{m+1} \circ p_m \circ \cdots \circ p_1$ à D est égale au cycle $(b_1, \ldots, b_m)$.

2. On ne suppose plus maintenant que $D' = \theta(D)$ est égal à D. On va utiliser une généralisation de la décomposition en cycles à support disjoints d'une permutation. Prolongeons momentanément θ à D' en disant que les points de $D' \setminus D$ sont fixes. On définit alors la relation binaire R sur $D \cup D'$ en disant que $x R y$ s'il existe un entier naturel k tel que $y = \theta^k(x)$ ou $x = \theta^k(y)$. Il s'agit clairement d'une relation d'équivalence sur $D \cup D'$. Décrivons la classe d'un élément. Soit $x \in D$. Si les images successives de x par θ restent toutes dans D alors, comme dans le cas des bijections, il existe nécessairement un plus petit entier k tel que $\theta^k(x) = x$ et la classe de x est égale à $\{x, \theta(x), \ldots, \theta^{k-1}(x)\}$. L'alternative est qu'il existe un entier k tel que $\theta^k(x)$ ne soit plus dans D. Les itérés de x par θ s'arrêtent alors à cet élément. On regarde alors si x est ou non dans D'. Si oui, il est l'image par θ d'un unique élément x_1 de D : la classe de x est la même que la classe de x_1. On regarde alors si x_1 est encore dans D' ou non. Bref, on peut trouver un élément z de $D \setminus D'$ tel que la classe de x soit celle de z, à savoir $\{z, \theta(z), \ldots, \theta^\ell(z) = x, \ldots, \theta^k(x) = \theta^{k+\ell}(z)\}$. La figure suivante résume ces possibilités :



Comme dans la question 1 il nous suffit de prouver le résultat pour la restriction de θ à une classe d'équivalence (intersectée avec D), restriction qui reste injective. C'est fait dans la question 1 pour les classes

circulaires. Il reste donc à traiter le cas d'une classe du second type. On suppose donc que $D = \{z, \theta(z), \ldots, \theta^{k-1}(z)\}$ avec $\theta^k(z) \notin D$ $(k \geqslant 2)$. Prenons alors l'application p_1 de X dans X qui envoie $\theta^{k-1}(z)$ sur $\theta^k(z)$ et qui fixe tous les autres points, p_2 qui envoie $\theta^{k-2}(z)$ sur $\theta^{k-1}(z)$ et qui fixe tous les autres points et ainsi de suite jusqu'à p_k qui envoie z sur $\theta(z)$ en fixant le reste. Il est clair que les p_i sont idempotentes et que la restriction de $p_k \circ \cdots \circ p_1$ à D est égale à θ.

3. Choisissons $D = \{x_1, \ldots, x_p\}$ un système de représentants des classes d'équivalence pour la relation d'équivalence canoniquement associée à α : la restriction θ de α à D établit une bijection entre D et l'image de α. Comme α n'est pas bijective, θ n'est pas surjective et on peut donc appliquer la question 2. Il existe des applications $\pi_1, \ldots, \pi_n$ de X dans X vérifiant $\pi_i \circ \pi_i = \pi_i$ pour tout i et telles que $\alpha(x_k) = \pi_1 \circ \cdots \circ \pi_n(x_k)$ pour tout k. Notons alors π l'application qui à tout x de X associe l'unique x_k tel que $\alpha(x) = \alpha(x_k)$. Il est clair que π est idempotente et on a $\alpha = \pi_1 \circ \cdots \circ \pi_n \circ \pi$. Le théorème de Howie est prouvé. ◁

Les deux exercices qui suivent concernent des évaluations asymptotiques.

1.15. Distribution du premier chiffre des puissances de 2

Pour $i \in [\![1, 9]\!]$, et $n \in \mathbb{N}^*$, on note $N_i(n)$ le nombre d'éléments de l'ensemble $\{2, 2^2, \ldots, 2^n\}$ dont le premier chiffre de l'écriture décimale est i. Calculer $\lim\limits_{n \to +\infty} \dfrac{N_i(n)}{n}$.

(École normale supérieure)

Solution.

Le but de l'exercice est de déterminer la fréquence d'apparition des chiffres $1, 2, \ldots, 9$ en première position dans la suite des puissances de 2. A priori on ne sait pas si tous les chiffres vont apparaître, et encore moins s'il vont apparaître une infinité de fois. Fixons i entre 1 et 9, et commençons par traduire le fait que i est le premier chiffre de 2^p.

• L'entier 2^p commence par i si et seulement s'il existe $k \in \mathbb{N}$ tel que $i.10^k \leqslant 2^p < (i+1).10^k$ c'est-à-dire, en prenant le logarithme, si et seulement s'il existe k tel que

$$\frac{\ln i}{\ln 10} + k \leqslant p\frac{\ln 2}{\ln 10} < k + \frac{\ln(i+1)}{\ln 10}.$$

Cela est encore équivalent à dire que le résidu modulo 1 de $p\theta$ est dans l'intervalle $\left[\dfrac{\ln i}{\ln 10}, \dfrac{\ln(i+1)}{\ln 10}\right[$, où $\theta = \dfrac{\ln 2}{\ln 10}$. Ainsi, $N_i(n)$ est exacte-

ment le nombre d'entiers $p \in [\![1, n]\!]$ tels que $p\theta$ modulo 1 appartienne à $\left[\frac{\ln i}{\ln 10}, \frac{\ln(i+1)}{\ln 10}\right[$.

• On est donc amené à étudier le comportement de la suite $(p\theta)_{p\geqslant 1}$ modulo 1. Il s'avère que θ est irrationnel. En effet, si $\frac{\ln 2}{\ln 10} = \frac{a}{b}$ avec $(a, b) \in (\mathbb{N}^*)^2$, il vient $2^b = 10^a$ ce qui est impossible puisque 2^b n'est pas divisible par 5. Montrons alors que la suite $(p\theta) \mod 1$ est dense dans l'intervalle $[0, 1]$. Cela prouvera déjà qu'il y a une infinité d'entiers p tels que 2^p commence par le chiffre i.

Il est bien entendu équivalent de montrer que l'ensemble $\mathbb{N}\theta + \mathbb{Z}$ est dense dans $\mathbb{R}$. On a presque un sous-groupe additif de $\mathbb{R}$. Or, nous savons qu'un sous-groupe additif de $\mathbb{R}$ est soit monogène, soit dense. Le sous-groupe $\mathbb{Z} + \theta\mathbb{Z}$ n'est pas monogène : si on avait $\mathbb{Z} + \theta\mathbb{Z} = a\mathbb{Z}$, il existerait des entiers α, β tels que $1 = \alpha a$ et $\theta = \beta a$. Mais en faisant le rapport des deux égalités, on aurait $\theta = \frac{\beta}{\alpha} \in \mathbb{Q}$ ce qui n'est pas. Ainsi, $\mathbb{Z} + \theta\mathbb{Z}$ est dense dans $\mathbb{R}$. Il est alors facile d'en déduire que $\mathbb{N}\theta + \mathbb{Z}$ reste encore dense dans $\mathbb{R}$. Soit $]a, b[\subset \mathbb{R}$. On peut trouver un élément z non nul de $\mathbb{Z} + \theta\mathbb{Z}$ tel que $|z| < b - a$. Quitte à prendre $-z$, on peut prendre z dans $\mathbb{N}\theta + \mathbb{Z}$. Supposons par exemple $z > 0$. Si $n_0 \in \mathbb{Z}$ est un entier inférieur à a, il est clair que l'un des éléments de la suite $(n_0 + kz)_{k\geqslant 0}$ va tomber dans l'intervalle $]a, b[$. Or, $n_0 + kz$ est dans $\mathbb{N}\theta + \mathbb{Z}$ pour tout $k \geqslant 0$. D'où le résultat. (Pour $z < 0$, on considère de même un entier $n_0 > b$.)

• La suite $(p\theta) \pmod 1$ est donc dense dans $[0, 1]$. Mais cela ne suffit pas pour déterminer un équivalent de $\mathrm{N}_i(n)$. Il nous faut étudier comment les résidus modulo 1 des termes $p\theta$ se répartissent dans l'intervalle $[0, 1[$ afin de déterminer la proportion de ceux qui vont tomber dans l'intervalle $\left[\frac{\ln i}{\ln 10}, \frac{\ln(i+1)}{\ln 10}\right[$ ou plus généralement dans un intervalle I fixé. Si la répartition est régulière, concept que nous allons préciser, il est légitime de penser que la proportion cherchée est exactement la longueur de l'intervalle I. Nous dirons qu'une suite (s_n) de réels de $[0, 1]$ est *équirépartie* si, pour tout intervalle $\mathrm{I} \subset [0, 1]$, la proportion

$$\frac{\mathrm{Card}\{k \in [\![1, n]\!], \quad s_k \in \mathrm{I}\}}{n}$$

tend vers la longueur de I lorsque n tend vers l'infini.

Topologiquement le quotient $\mathbb{R}/\mathbb{Z}$ est un cercle (c'est le segment $[0, 1]$ où on a identifié 0 et 1). En fait, l'application $\overline{x} \longmapsto e^{2i\pi x}$ établit une bijection (et même un homéomorphisme pour le lecteur connaissant la notion de topologie quotient) entre $\mathbb{R}/\mathbb{Z}$ et le cercle unité S^1 du plan complexe. La suite $(p\theta)$ modulo 1 devient la suite des puissances de $e^{2i\pi\theta}$. Cette suite tourne sur le cercle S^1, l'écart (angulaire) entre deux termes consécutifs étant constant. Il est alors assez naturel de penser que cette

suite est équirépartie et c'est effectivement le cas : c'est le théorème de Bohl-Sierpinski-Weyl[5] qui l'affirme. Pour notre problème on peut alors en déduire que

$$\boxed{\lim_{n\to+\infty} \frac{\mathrm{N}_i(n)}{n} = \frac{\ln(i+1) - \ln i}{\ln 10}}.$$

En particulier le chiffre i apparaît d'autant plus fréquemment qu'il est petit ! Pour $i = 1$ on a une fréquence de 30%, alors que pour $i = 9$, on obtient 4,5%.

• Évidemment il reste à prouver le théorème de Bohl-Sierpinski-Weyl. On peut en donner une preuve directe[6] mais c'est assez long. On renvoie le lecteur au tome 2 d'analyse (exercices 1.29 et 1.30) où figure un exercice qui démontre le critère d'équirépartition de Weyl qui implique facilement ce résultat. ◁

L'exercice suivant établit un résultat qui remonte à Gauss et qui donne une estimation du nombre de points à coordonnées entières dans un disque euclidien de rayon R.

1.16. Un théorème de Gauss

On munit $\mathbb{R}^n$, $n \geqslant 2$, de sa structure euclidienne canonique. Donner un équivalent lorsque r tend vers $+\infty$ du nombre $\mathrm{N}(r)$ de points de $\mathbb{Z}^n$ de norme inférieure ou égale à r.

(École polytechnique, École normale supérieure)

▷ **Solution.**

L'idée naturelle est la suivante. Le nombre de points entiers dans un domaine D de $\mathbb{R}^n$, suffisamment grand et assez régulier (par exemple convexe), est approximativement le volume de D. Dans le cas des boules fermées B_r de centre l'origine et de rayon $r > 0$ on va montrer rigoureusement que $\mathrm{N}(r)$ est effectivement équivalent au volume $\mathrm{V}(r)$ de cette boule.

À tout point $x = (x_1, \ldots, x_n)$ de $\mathbb{Z}^n$, on associe l'hypercube C_x de coté 1 centré en x défini par

$$\mathrm{C}_x = \left\{ t = (t_1, \ldots, t_n) \in \mathbb{R}^n, \quad \forall i \in [\![1, n]\!], \ |t_i - x_i| \leqslant \frac{1}{2} \right\}.$$

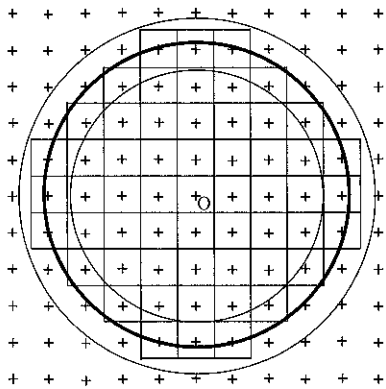
Ces hypercubes C_x sont de volume 1. Pour $r > 0$, considérons l'ensemble

5. Prouvé indépendamment par les trois mathématiciens vers 1910.

6. Le lecteur pourra trouver une preuve de ce type dans les deux ouvrages suivants : HAUZY (G.), *Propriétés statistiques de suites arithmétiques*, PUF, 1976 ; HARDY (G.H.) & WRIGHT (E.M), *An introduction to the theory of numbers*, Oxford University Press, 5ème éd., 1979.

des cubes C_x centrés aux points de $\mathbb{Z}^n$ qui sont dans la boule fermée B_r ; il y en a $N(r)$. Le volume de la réunion de ces hypercubes est donc $N(r)$.

• Voici une figure dans le plan :



Le réseau $\mathbb{Z}^2$ est représenté par des petites croix. En gras figure un cercle de rayon $r > 0$, les deux autres cercles ayant pour rayons respectifs $r + \frac{\sqrt{2}}{2}$ et $r - \frac{\sqrt{2}}{2}$. Commençons par expliquer l'idée dans le cas du plan. Il est clair que tout carré C_x centré en un point $x \in \mathbb{Z}^2 \cap B_r$ est inclus dans le disque de rayon $r + \frac{\sqrt{2}}{2}$. Cela provient de l'inégalité triangulaire et du fait qu'un point de C_x est à une distance inférieure à $\frac{\sqrt{2}}{2}$ de x. De même, on montre que le disque de rayon $r - \frac{\sqrt{2}}{2}$ est complètement inclus dans la réunion des petits carrés. Or, l'aire de la réunion des carrés est exactement $N(r)$, le nombre de carrés (car ils sont tous d'aire 1 et d'intérieurs disjoints). On a donc l'encadrement

$$\pi\left(r - \frac{\sqrt{2}}{2}\right)^2 \leqslant N(r) \leqslant \pi\left(r + \frac{\sqrt{2}}{2}\right)^2.$$

Cette double inégalité prouve que $N(r) \sim \pi r^2$.

• Passons au cas général. La démarche est la même. On fait intervenir le volume d'une boule de rayon r. A priori, on ne sait pas le calculer, mais si on note b_n le volume de la boule euclidienne fermée de rayon 1 (ainsi $b_2 = \pi$), par homogénéité une boule de rayon $r > 0$ a pour volume $V(r) = b_n r^n$. La distance maximale d'un point d'un de nos hypercubes à son centre est $\frac{\sqrt{n}}{2}$. L'inégalité précédente s'écrit alors

$$b_n\left(r - \frac{\sqrt{n}}{2}\right)^n \leqslant N(r) \leqslant b_n\left(r + \frac{\sqrt{n}}{2}\right)^n.$$

On en déduit que $N(r) \sim b_n r^n$ lorsque r tend vers $+\infty$. ◁

Voici une autre manière de voir les choses. Au lieu de compter les points de $\mathbb{Z}^n$ dans B_r, il revient au même de compter les points du réseau $\frac{1}{r}\mathbb{Z}^n$ dans la boule unité B_1. Alors $\dfrac{\mathrm{N}(r)}{r^n}$ est le volume de la réunion des hypercubes de côté $\dfrac{1}{r}$ centrés aux points de $\dfrac{1}{r}\mathbb{Z}^n \cap \mathrm{B}_1$ et il semble clair que ce volume va tendre vers le volume de B_1 : c'est pratiquement le procédé d'approximation utilisé pour définir la notion de partie quarrable dans le cas du plan.

On trouvera dans le tome 2 d'analyse (exercice 3-23) une application du résultat de cet exercice à la recherche d'un équivalent pour une série entière. Signalons enfin qu'il est facile d'obtenir une relation de récurrence entre b_n et b_{n-1} conduisant à la formule suivante : $b_n = \dfrac{\pi^{n/2}}{\Gamma\left(\dfrac{n+2}{2}\right)}$, *où Γ est la fonction d'Euler.*

Nous allons terminer par deux exercices concernant les entiers naturels, exercices de nature moins combinatoire, mais qui s'inscrivent tout de même naturellement dans ce chapitre. Le premier étudie à quelle condition deux suites pseudo-arithmétiques permettent de partitionner $\mathbb{N}^*$ (une suite pseudo-arithmétique est une suite d'entiers de la forme $(\mathrm{E}(n\alpha))_{n\geqslant 1}$ avec $\alpha > 1$).

1.17. Théorème de Beatty (1926)

Soit $a, b > 1$, $\mathrm{E} = \{\mathrm{E}(na), n \in \mathbb{N}^*\}$ et $\mathrm{F} = \{\mathrm{E}(nb), n \in \mathbb{N}^*\}$. Montrer qu'il y a équivalence entre :

(i) E et F forment une partition de $\mathbb{N}^*$;

(ii) $\dfrac{1}{a} + \dfrac{1}{b} = 1$ et a et b sont irrationnels.

(École normale supérieure)

Solution.

• Le résultat est assez intuitif si l'on introduit la notion de *densité* d'une partie A de $\mathbb{N}^*$. Pour tout $n \geqslant 1$, on note a_n le cardinal de $\mathrm{A} \cap [\![1, n]\!]$. Si la suite $\left(\dfrac{a_n}{n}\right)$ converge on dira que A admet une densité égale par définition à la limite $d(\mathrm{A})$ de cette suite. Par exemple, l'ensemble des entiers pairs est de densité $\dfrac{1}{2}$, l'ensemble des carrés est de densité nulle.

• Montrons que pour $\alpha > 1$, l'ensemble $\{\mathrm{E}(n\alpha), n \in \mathbb{N}^*\}$ est de densité α^{-1}. En effet, les entiers $\mathrm{E}(k\alpha)$ sont deux à deux distincts et il y en a $\mathrm{E}\left(\dfrac{n}{\alpha}\right)$ dans l'intervalle $[\![1, n]\!]$. Le résultat annoncé provient de ce que $\lim\limits_{n\to+\infty} \dfrac{1}{n}\mathrm{E}\left(\dfrac{n}{\alpha}\right) = \dfrac{1}{\alpha}$.

• $(i) \Longrightarrow (ii)$. Il est clair que si A, B sont des parties disjointes de $\mathbb{N}^*$ admettant toutes deux une densité, alors $A \cup B$ admet une densité et $d(A \cup B) = d(A) + d(B)$. Comme $d(\mathbb{N}^*) = 1$, on a donc $a^{-1} + b^{-1} = 1$ d'après le point précédent. De plus, a et b ne peuvent pas être tous les deux rationnels car si $a = \frac{p}{q}$ et $b = \frac{p'}{q'}$, alors $E(p'qa) = E(pq'b) = pp'$ est dans $E \cap F$. L'un des deux est donc irrationnel. Mais la relation $\frac{1}{a} + \frac{1}{b} = 1$ impose alors l'irrationalité du second.

• $(ii) \Longrightarrow (i)$. Commençons par montrer que E et F sont disjoints. On raisonne par l'absurde. Soit $k \in E \cap F$, $n \geqslant 1$ et $m \geqslant 1$ tels que $k = E(na) = E(mb)$. Par définition de la partie entière, $k \leqslant na < k + 1$ et $k \leqslant mb < k + 1$. En divisant la première inégalité par a et la seconde par b on a $\frac{k}{a} \leqslant n < \frac{k}{a} + \frac{1}{a}$ et $\frac{k}{b} \leqslant m < \frac{k}{b} + \frac{1}{b}$. Compte tenu de la relation qui lie a et b, on obtient en additionnant les deux inégalités $k \leqslant n+m < k+1$. Comme k, n et m sont entiers cela impose $n+m = k$. On a alors nécessairement égalité dans les deux inégalités précédentes, c'est-à-dire $na = mb = k$. C'est absurde car a et b sont irrationnels. On a donc $E \cap F = \emptyset$.

Montrons maintenant que $E \cup F = \mathbb{N}^*$. Soit $n \geqslant 1$, $p = E(na)$ et m l'unique entier (éventuellement nul) tel que $E(mb) < p < E((m+1)b)$ (les inégalités sont strictes car $E \cap F = \emptyset$). L'entier $E(mb)$ (resp. $p = E(na)$) est le plus grand élément de F (resp. de E) appartenant à l'intervalle $[\![1, p]\!]$. Les applications $k \longmapsto E(ka)$ et $k \longmapsto E(kb)$ étant injectives (car $a, b > 1$), l'intervalle $[\![1, p]\!]$ contient donc $m + n$ éléments de $E \cup F$ (car ces deux ensembles sont disjoints). Or, on a $\frac{p}{a} \leqslant n < \frac{p+1}{a}$ et $\frac{p}{b} - 1 < m < \frac{p+1}{b}$. En additionnant il vient $p - 1 < n + m < p + 1$, donc $p = n + m$: tous les entiers de $[\![1, p]\!]$ sont dans $E \cup F$. Le résultat suit car n est arbitraire et p est donc arbitrairement grand. ◁

Il a été démontré, dans les années 20, qu'il est impossible de partitionner $\mathbb{N}^*$ avec trois suites pseudo-arithmétiques ou plus.

Voici, pour finir, un exercice posé aux Olympiades Internationales en 1971 à Zilina (aujourd'hui en Slovaquie).

1.18. Un exercice d'Olympiades

Soit $A = (a_{ij})$ une matrice carrée de taille $n \geqslant 2$, à coefficients entiers naturels et telle que, pour $(k, l) \in [\![1, n]\!]^2$, si $a_{kl} = 0$ alors $\sum_{i=1}^{n} a_{il} + \sum_{j=1}^{n} a_{kj} \geqslant n$. Montrer que $\sum_{i,j} a_{ij} \geqslant \frac{n^2}{2}$. Donner un exemple où il y a égalité.

(École normale supérieure)

Solution.

On va essayer d'évaluer la somme des coefficients en utilisant une rangée (*i.e.* une ligne ou une colonne) où il y a beaucoup de zéros pour exploiter au mieux l'hypothèse. Cela invite à considérer une rangée dont la somme S des éléments est minimale. Quitte à prendre la transposée de A puis à permuter deux lignes on peut supposer que la rangée en question est la première ligne de A.

• Si $S \geqslant \frac{n}{2}$, le résultat est évident car la somme des coefficients de chacune des n lignes de A est minorée par $\frac{n}{2}$.

• Si $S < \frac{n}{2}$, il y a $\alpha \geqslant \frac{n}{2}$ zéros dans la première ligne de A. En notant $J = \{j \in [\![1, n]\!],\ a_{1j} = 0\}$, on a d'après l'hypothèse

$$\sum_{i,j} a_{ij} = \sum_{j \in J} \sum_{i=1}^{n} a_{ij} + \sum_{j \notin J} \sum_{i=1}^{n} a_{ij} \geqslant \alpha(n - S) + (n - \alpha)S.$$

En écrivant $\alpha(n - S) + (n - \alpha)S = \frac{n^2}{2} + (n - 2S)(\alpha - \frac{n}{2}) \geqslant \frac{n^2}{2}$, on obtient l'inégalité voulue.

Pour avoir égalité, il est bien entendu nécessaire de prendre n pair. La matrice «damier» définie par $a_{ij} = 1$ si $i \equiv j$ [2] et $a_{ij} = 0$ fournit alors un exemple où il y a égalité. ◁

Chapitre 2

Théorie des groupes

C'est au cours du XIX^{e} *siècle qu'émerge la notion abstraite de groupe. Deux sortes de lois de composition interne sont à l'origine du concept : celle relative aux classes d'équivalences en Arithmétique (introduites de manière implicite dans les* Disquisitiones arithmeticæ *de Gauss) et la composition des permutations sur un ensemble fini étudiée notamment par Galois et dont la théorie fut largement développée par Jordan dans son* Traité des substitutions. *Le premier à rapprocher les points de vue et à concevoir que la structure de groupe est indépendante de la nature des objets considérés est Cayley qui donne une première définition abstraite d'un groupe dans* On the theory of groups as depending on the symbolic equation $\theta^n = 1$ *(1854). Il dresse des tables de multiplication pour les groupes de petit ordre. Il connaît les deux groupes d'ordre* 6 *(à isomorphisme près), les cinq d'ordre* 8 *et fournit des exemples d'une grande variété : ensembles de matrices inversibles, de quaternions, groupe des racines n-ième de l'unité... Jordan approfondira ce travail et étudiera de manière très poussée le groupe linéaire et ses sous-groupes. Il est notamment à l'origine de la notion de groupe quotient et de la théorie des représentations linéaires (*i.e. *l'étude des morphismes d'un groupe dans le groupe linéaire d'un espace vectoriel). Cette réalisation géométrique des groupes permet de traduire les propriétés géométriques en propriétés algébriques et inversement. L'étude des groupes de transformations permet la classification des géométries de Felix Klein. Il faut toutefois reconnaître que durant tout ce siècle le terme de «groupe» fut employé de manière très vague, sans que les exemples classiques des groupes additif* $\mathbb{Z}$ *ou multiplicatif* $\mathbb{Q}^*$ *soient relevés et ce n'est qu'en 1898 que le mathématicien Weber en donne la définition connue aujourd'hui.*

Commençons par un exercice élémentaire sur les lois de composition, posé au concours de l'École polytechnique.

2.1. Existence d'un idempotent

Soit E un ensemble fini muni d'une loi de composition interne associative. Montrer qu'il existe $s \in \mathrm{E}$ tel que $s^2 = s$.

(École polytechnique)

▷ **Solution.**

Soit a un élément quelconque de E. L'associativité de la loi permet de donner un sens à a^k pour tout entier $k \in \mathbb{N}^*$. Comme E est fini, la suite $u_n = a^{2^n}$ (obtenue à partir de $u_0 = a$ par itération de $s \mapsto s^2$) ne peut pas être injective. On peut donc trouver $n \in \mathbb{N}^*$ et $p > 0$ tels que $u_{n+p} = u_n$. Si on pose $b = a^{2^n}$, on a alors $b^{2^p} = b$. Si $p = 1$ c'est gagné. Pour avoir des idées regardons ce qui se passe pour $p = 2$. On a $b^4 = b$. On obtient alors $(b^3)^2 = b^6 = b^4 b^2 = bb^2 = b^3$ et b^3 est idempotent. En fait de manière générale, si $x^m = x$ avec $m \geqslant 2$, alors x^{m-1} est idempotent puisque $(x^{m-1})^2 = x^{2m-2} = x^m \cdot x^{m-2} = x \cdot x^{m-2} = x^{m-1}$. Donc ici, $s = b^{2^p-1}$ convient. ◁

Les exercices suivants concernent les notions de sous-groupes et de morphismes de groupes.

2.2. Groupes dont l'ensemble des sous-groupes est fini

Caractériser les groupes dont l'ensemble des sous-groupes est fini.
(École normale supérieure)

▷ **Solution.**

Les groupes finis vérifient de manière évidente cette condition. Nous allons démontrer que ce sont les seuls. Soit G un groupe dont l'ensemble E des sous-groupes est fini. Tout x de G est d'ordre fini car un élément d'ordre infini engendre un sous-groupe isomorphe à $\mathbb{Z}$ et $\mathbb{Z}$ admet une infinité de sous-groupes. Si E′ désigne le sous-ensemble de E formé des groupes monogènes, on a $G = \bigcup_{H \in E'} H$. Comme E′ est fini et que les éléments de E′ sont des ensembles finis d'après ce qui précède, G est fini. ◁

2.3. Morphismes de $\mathbb{Q}$ dans $\mathbb{Z}$

Trouver tous les morphismes de groupes de $(\mathbb{Q}, +)$ dans $(\mathbb{Z}, +)$.
(École normale supérieure)

▷ **Solution.**

Soit f un morphisme de groupes de $(\mathbb{Q}, +)$ dans $(\mathbb{Z}, +)$. Son image est un sous-groupe de $\mathbb{Z}$, c'est-à-dire un certain $n\mathbb{Z}$, $n \in \mathbb{N}$. Si $n \geqslant 1$, on choisit un antécédent x de n. On obtient alors $2f(\frac{x}{2}) = f(x) = n$ et donc $\frac{n}{2} = f(\frac{x}{2}) \in n\mathbb{Z}$, ce qui est absurde. On a donc $n = 0$ et f est nul. ◁

L'exercice suivant doit être rapproché de l'exercice 6.15, plus classique, où il s'agit de démontrer la même équivalence pour un endomorphisme f d'un espace vectoriel E *de dimension finie. La démonstration est quasiment la même : on passe d'une inclusion à une égalité ensembliste en invoquant ici un argument de cardinal et là un argument de dimension ; une relation entre le cardinal de* $\operatorname{Im} f$ *et celui de* $\operatorname{Ker} f$ *remplace le théorème du rang.*

2.4. Équivalence $\operatorname{Ker} f = \operatorname{Ker} f^2 \Longleftrightarrow \operatorname{Im} f = \operatorname{Im} f^2$

Soit G un groupe fini et f un morphisme de G dans G. Montrer que

$$\operatorname{Ker} f = \operatorname{Ker} f^2 \Longleftrightarrow \operatorname{Im} f = \operatorname{Im} f^2.$$

(École normale supérieure)

Solution.

• On observe que l'on a $\operatorname{Ker} f \subset \operatorname{Ker} f^2$ et $\operatorname{Im} f^2 \subset \operatorname{Im} f$. Ces ensembles étant tous finis, on en déduit que $\operatorname{Ker} f = \operatorname{Ker} f^2$ équivaut à $|\operatorname{Ker} f| = |\operatorname{Ker} f^2|$ et $\operatorname{Im} f = \operatorname{Im} f^2$ à $|\operatorname{Im} f| = |\operatorname{Im} f^2|$. Il s'agit de démontrer que ces deux égalités de cardinaux sont équivalentes.

• Montrons que si $f : \mathrm{G} \to \mathrm{G}$ est un morphisme de groupes, alors $|\mathrm{G}| = |\operatorname{Im} f| \times |\operatorname{Ker} f|$. Soit $\mathcal{R}$ la relation d'équivalence associée à f : $x \mathcal{R} y \Longleftrightarrow f(x) = f(y)$. L'ensemble quotient $\mathrm{G}/\mathcal{R}$ est équipotent à $\operatorname{Im} f$. Mais d'autre part, $f(x) = f(y)$ équivaut à $f(x^{-1}y) = 1$, puisque f est un morphisme et donc à $x^{-1}y \in \operatorname{Ker} f$. La classe $\overline{x}$ d'un élément x pour $\mathcal{R}$ est donc la classe à gauche modulo $\operatorname{Ker} f$: $\overline{x} = x \operatorname{Ker} f$. Ainsi, $\overline{x}$ est équipotent à $\operatorname{Ker} f$ (par la bijection $g \in \operatorname{Ker} f \longmapsto xg \in \overline{x}$). Les classes ont donc toutes même cardinal, celui de $\operatorname{Ker} f$. On en déduit que

$$|\mathrm{G}| = |\mathrm{G}/\mathcal{R}| \times |\operatorname{Ker} f| = |\operatorname{Im} f| \times |\operatorname{Ker} f|.$$

On a donc, pour tout morphisme f de G,

$$|\mathrm{G}| = |\operatorname{Im} f| \times |\operatorname{Ker} f|.$$

On obtient de même

$$|\mathrm{G}| = |\operatorname{Im} f^2| \times |\operatorname{Ker} f^2|.$$

Ces deux égalités impliquent l'équivalence entre $|\operatorname{Im} f| = |\operatorname{Im} f^2|$ et $|\operatorname{Ker} f| = |\operatorname{Ker} f^2|$, ce qui établit le résultat demandé. ◁

Voici quelques rappels sur la notion de groupe quotient dans le cas abélien. Soit G *un groupe abélien, et* H *un sous-groupe de* G. *La relation*

$\mathrm{R_H}$ *(dite de congruence modulo* H*) définie sur* G *par* $x\mathrm{R_H}y \Leftrightarrow x^{-1}y \in \mathrm{H}$ *est une relation d'équivalence. On note* $\overline{x} = x\mathrm{H} = \mathrm{H}x$ *la classe d'un élément* x *de* G *et* G/H *l'ensemble quotient. La loi de* G *est alors compatible avec le passage au quotient, c'est-à-dire que la classe* $\overline{xy}$ *ne dépend pas du choix de* x *dans* $\overline{x}$ *et de* y *dans* $\overline{y}$*. Cela permet de définir une loi de composition interne sur* G/H *par* $\overline{x}.\overline{y} = \overline{xy}$*, loi qui munit* G/H *d'une structure de groupe, appelé groupe quotient de* G *par* H*. Les quotients de* $\mathbb{Z}$ *sont des exemples fondamentaux qui ont été étudiés en cours.*

Rappelons enfin le premier théorème d'isomorphisme dans le cas abélien. Si $f : \mathrm{G} \to \mathrm{G}'$ *est un morphisme de groupes, alors la relation d'équivalence associée à* f *et la relation de congruence modulo* $\operatorname{Ker} f$ *sont égales et la bijection canonique* $\overline{f}$ *de* $\mathrm{G}/\operatorname{Ker} f$ *sur* $\operatorname{Im} f$ *induite par* f *est un isomorphisme de groupes.*

2.5. Sous-groupes finis de $\mathbb{Q}/\mathbb{Z}$

Montrer que, pour tout $n \geqslant 1$, il existe un unique sous-groupe de $(\mathbb{Q}/\mathbb{Z}, +)$ de cardinal n.

(École polytechnique)

▷ **Solution.**

Soit $n \geqslant 1$ et G un sous-groupe de cardinal n de $\mathbb{Q}/\mathbb{Z}$. Si $x \in \mathbb{Q}$ est tel que $\overline{x} \in \mathrm{G}$, l'ordre de $\overline{x}$ divise n et donc $\overline{nx} = n\overline{x} = \overline{0}$. On en déduit qu'il existe $p \in \mathbb{Z}$ tel que $x = \dfrac{p}{n}$. Si r désigne le résidu de p modulo n, on a même $\overline{x} = \overline{\left(\dfrac{r}{n}\right)}$. Ceci montre que $\mathrm{G} \subset \left\{\overline{\left(\dfrac{r}{n}\right)}, 0 \leqslant r \leqslant n-1\right\}$. Les n éléments de cet ensemble sont distincts et ils forment un sous-groupe de $\mathbb{Q}/\mathbb{Z}$, le sous-groupe engendré par $\overline{\left(\dfrac{1}{n}\right)}$. D'après ce qui précède, c'est le seul sous-groupe de $(\mathbb{Q}/\mathbb{Z}, +)$ de cardinal n. ◁

Les exercices suivants concernent la notion d'ordre d'un élément dans un groupe. L'ordre d'un élément x *d'un groupe* G *est le cardinal (fini ou infini) du sous-groupe de* G *engendré par* x*. Si l'ordre de* x *est fini égal à* m*, ce sous-groupe est égal à* $\{e, x, x^2, \ldots, x^{m-1}\}$ *et isomorphe à* $\mathbb{Z}/m\mathbb{Z}$*. L'entier* m *est le plus petit* $k \in \mathbb{N}^*$ *tel que* $x^k = e$*. Dans le cas où l'ordre de* x *est infini, le sous-groupe qu'il engendre est isomorphe à* $\mathbb{Z}$*.*

Lorsque G *est fini, l'ordre de tout* x *divise le cardinal de* G *(c'est un corollaire du théorème de Lagrange). Bien entendu il n'y a pas, en général, d'élément d'ordre* d *pour tout diviseur* d *de* $|\mathrm{G}|$ *(sans quoi tous les groupes finis seraient cycliques). Le lemme de Cauchy, qui fera plus loin l'objet d'un exercice, montre que si* G *est fini de cardinal* n*, alors*

G admet, pour tout diviseur premier p de n, des éléments d'ordre p. Les deux exercices qui suivent deviennent très simples dès lors que l'on dispose de ce lemme. Nous en donnons toutefois des solutions n'y faisant pas appel et qui sont sans doute plus dans l'esprit de ce qu'attendait l'examinateur.

2.6. Groupes abéliens de cardinal pq

Soit G un groupe abélien d'ordre pq, où p et q sont deux nombres premiers distincts. Montrer que G est cyclique.

(École normale supérieure)

▷ **Solution.**

Nous noterons la loi du groupe multiplicativement. Comme p et q sont premiers, les éléments de G différents de 1 sont d'ordre pq, p ou q d'après le théorème de Lagrange.

• Montrons que si G admet un élément x d'ordre p et un élément y d'ordre q, alors xy est d'ordre pq. On a $xy \neq 1$, car sinon $y = x^{-1}$ a même ordre que x, $(xy)^p = y^p \neq 1$ car q ne divise pas p et, de même, $(xy)^q \neq 1$. L'ordre de xy ne peut être que pq. À l'aide du lemme de Cauchy (voir 2.10) on pourrait donc conclure ici.

• Si G n'est pas cyclique, *i.e.* ne possède pas d'élément d'ordre pq, on en déduit que les éléments de G distincts de 1 sont tous d'ordre p ou tous d'ordre q. Supposons-les par exemple tous d'ordre p. Soit x un élément d'ordre p et H le sous-groupe qu'il engendre. On considère le groupe quotient G/H de cardinal q. Comme q est premier, le cours assure que G/H est cyclique. Soit $z \in$ G tel que $\overline{z}$ engendre G/H ; $\overline{z}$ est donc d'ordre q. Mais on a aussi $\overline{z^p} = \overline{z}^p = \overline{1}$. Comme $\overline{z}$ est d'ordre q, on en déduit q divise p, ce qui est absurde.

Conclusion. G est cyclique. ◁

Le résultat n'est bien entendu plus vrai si G *n'est pas supposé abélien : le groupe $\mathcal{S}_3$ est d'ordre $6 = 2\times 3$ et n'est pas abélien (donc non cyclique).*

2.7. Un cas particulier du lemme de Cauchy

Soit G un groupe de cardinal $2p$ avec p premier. Montrer que G contient un élément d'ordre p.

(École normale supérieure)

▷ **Solution.**

D'après le théorème de Lagrange, les éléments de G sont d'ordre 1, 2, p ou $2p$. Raisonnons par l'absurde et supposons qu'il n'y ait pas

d'élément d'ordre p. Dans ces conditions, G n'est pas cyclique (car si x est un générateur de G, alors x^2 est d'ordre p) et si $x \in G$, x est d'ordre 1 (si x est l'élément neutre) ou 2. En particulier, $p \geqslant 3$ et pour tout $x \in G$, $x^2 = 1$. On peut alors prouver le lemme suivant, intéressant en soi :

Lemme. *Soit* G *un groupe tel que pour tout* $x \in G$, $x^2 = 1$. *Alors* G *est abélien et* Card G *est une puissance de* 2.

Démonstration.

• Soit x et y dans G. Pour tout $z \in G$ on a $z^2 = 1$, *i.e.* $z = z^{-1}$. On en déduit que

$$xy = (xy)^{-1} = y^{-1}x^{-1} = yx$$

et G est abélien.

• Montrons par récurrence sur Card G que Card G est une puissance de 2.

Il n'y a rien à vérifier si Card G = 1.

Supposons Card G $\geqslant 2$. On considère les sous-groupes de G distincts de G ; il en existe, $\{1\}$ par exemple. On en choisit un de cardinal maximal que l'on note H. D'après l'hypothèse de récurrence, Card H est une puissance de 2. Soit $a \in G \backslash H$. Montrons que $H \cup aH$ est un sous-groupe de G ; en effet, il est égal au sous-groupe H′ engendré par a et H :

∗ on a clairement $H \cup aH \subset H'$;

∗ réciproquement, a étant d'ordre 2, tout élément de H′ s'écrit $a^\alpha h$, avec $\alpha \in \{0, 1\}$ et $h \in H$. D'où le résultat annoncé.

Les ensembles H et aH sont disjoints car sinon il existerait $h \in H$ qui s'écrirait $h = ak$ avec $k \in H$ et on aurait $a = hk^{-1} \in H$, ce qui est exclu. On en déduit que $\mathrm{Card}(H \cup aH) = 2\,\mathrm{Card}\,H$. Par maximalité du cardinal de H, $H \cup aH = G$. En particulier, $\mathrm{Card}\,G = 2\,\mathrm{Card}\,H$ est une puissance de 2. ◇

Le cardinal de notre groupe G devrait être une puissance de 2. Mais Card G = $2p$, avec $p \geqslant 3$ et premier. C'est contradictoire et G possède donc un élément d'ordre p. ◁

Un élégant argument d'algèbre linéaire permet de mieux saisir encore la structure de G *lorsque pour tout* $x \in G$, $x^2 = e$: G *est alors naturellement muni d'une structure d'espace vectoriel sur le corps* $\mathbb{Z}/2\mathbb{Z}$, *la loi externe étant définie par* $0.x = e$ *(neutre de* G*) et* $1.x = x$ *pour tout* x *de* G, *et la loi de groupe étant bien entendu celle de* G. *La propriété vérifiée par* G *intervient dans l'axiome de distributivité :* $(1 + 1).x = 0.x = e$ *et* $(1.x)(1.x) = x^2 = e$. *On laisse au lecteur le soin de vérifier les autres axiomes. Comme* G *est fini, il est de dimension finie en tant que* $\mathbb{Z}/2\mathbb{Z}$*-espace vectoriel. Si on note* n *cette dimension,* G *est isomorphe à* $(\mathbb{Z}/2\mathbb{Z})^n$ *en tant qu'espace vectoriel et* a fortiori *en tant que groupe.*

Le fait qu'un groupe abélien fini d'exposant 2 (voir ci-après) est isomorphe à $(\mathbb{Z}/2\mathbb{Z})^k$ résulte aussi du théorème de structure des groupes abéliens finis (cf. exercice 2.20 pour l'unicité et 7.22 pour l'existence).

2.8. Exposant d'un groupe abélien fini

Soit G un groupe abélien fini. Pour tout $x \in G$, on note $O(x)$ l'ordre de x.

1. Soit $(x, y) \in G^2$, $m = O(x)$, $n = O(y)$. On suppose que m et n sont premiers entre eux. Montrer que $O(xy) = mn$. Si on ne suppose plus m premier avec n, a t-on $O(xy) = \text{ppcm}(m, n)$?

2. Soit $(m, n) \in (\mathbb{N}^*)^2$. Montrer l'existence de $(m', n') \in (\mathbb{N}^*)^2$ tel que $m'|m$, $n'|n$, $\text{pgcd}(m', n') = 1$ et $\text{ppcm}(m, n) = m'n'$.

3. Montrer qu'il existe $z \in G$ tel que $O(z)$ soit le ppcm des ordres des éléments de G (ce ppcm est appelé l'exposant du groupe G).

4. Soit K un corps commutatif, G un sous-groupe fini du groupe multiplicatif K^*. Montrer que G est cyclique.

(**École normale supérieure**)

Solution.

1. Supposons m et n premiers entre eux et considérons $k \in \mathbb{Z}$ tel que $(xy)^k = 1$. Élevons cette égalité à la puissance n. Comme G est abélien il vient $x^{kn} = 1$. Donc m divise kn. Comme $n \wedge m = 1$, le lemme de Gauss garantit que m divise k. On montre de la même manière que n divise k, puis que nm divise k, à nouveau parce que $n \wedge m = 1$. Comme $(xy)^{mn} = 1$, il en résulte que $O(xy) = mn$.

L'hypothèse «m et n sont premiers entre eux» est essentielle : en effet, si on prend par exemple $y = x^{-1}$, l'élément $xy = 1$ est d'ordre 1 alors que $O(x) = O(y)$ et donc que $\text{ppcm}(O(x), O(y)) = n$.

2. Écrivons $m = \prod_{p \in \mathcal{P}} p^{\nu_p(m)}$ et $n = \prod_{p \in \mathcal{P}} p^{\nu_p(n)}$, où $\mathcal{P}$ désigne l'ensemble des nombres premiers. On a alors

$$\text{ppcm}(m, n) = \prod_{p \in \mathcal{P}} p^{\max(\nu_p(m), \nu_p(n))}.$$

Soit $(m', n') \in \mathbb{N}^*$. Pour que $m'|m$ et $n'|n$, il faut et il suffit que pour tout $p \in \mathcal{P}$, $\nu_p(m') \leqslant \nu_p(m)$ et $\nu_p(n') \leqslant \nu_p(n)$. Pour que m' et n' soient premiers entre eux, il faut et il suffit que pour tout $p \in \mathcal{P}$, $\nu_p(m') = 0$ ou $\nu_p(n') = 0$. Enfin, pour que $m'n' = \text{ppcm}(m, n)$, il faut et il suffit que pour tout $p \in \mathcal{P}$, $\nu_p(m') + \nu_p(n') = \max(\nu_p(m), \nu_p(n))$. Tout cela nous invite à poser

$$m' = \prod_{p \in \mathcal{P}} p^{\alpha_p} \quad \text{et} \quad n' = \prod_{p \in \mathcal{P}} p^{\beta_p}$$

avec, pour tout $p \in \mathcal{P}$,

$$\alpha_p = \begin{cases} \nu_p(m) & \text{si } \nu_p(m) > \nu_p(n) \\ 0 & \text{si } \nu_p(n) \geqslant \nu_p(m) \end{cases} \quad \text{et} \quad \beta_p = \begin{cases} 0 & \text{si } \nu_p(m) > \nu_p(n) \\ \nu_p(n) & \text{si } \nu_p(n) \geqslant \nu_p(m). \end{cases}$$

Les entiers m' et n' vérifient alors toutes les conditions demandées.

3. Considérons un élément z de G d'ordre maximal m. Soit x un élément de G d'ordre n. On choisit m' et n' comme à la question 2. L'élément $z^{m/m'}$ est d'ordre m' : en effet, on a $(z^{m/m'})^{m'} = z^m = 1$ et si $(z^{m/m'})^k = 1$, avec $k > 0$, on obtient $z^{mk/m'} = 1$ et nécessairement $mk/m' \geqslant m$, *i.e.* $k \geqslant m'$. De même, $x^{n/n'}$ est d'ordre n'. D'après la question 1, $z^{m/m'}x^{n/n'}$ est d'ordre $m'n' = \operatorname{ppcm}(m,n)$. Or, par construction, cet ordre est inférieur à m. On a donc $\operatorname{ppcm}(m,n) \leqslant m$, c'est-à-dire $\operatorname{ppcm}(m,n) = m$ et n divise m.

L'ordre m de z est donc multiple de tous les ordres des éléments de G. C'est évidemment le plus petit, m étant lui-même l'ordre de z. C'est donc le ppcm des ordres des éléments de G.

4. Soit m le ppcm des ordres des éléments de G et n le cardinal de G. Comme $x^m = 1$ pour tout x de G, on a

$$\mathrm{G} \subset \{x \in \mathrm{K},\ x^m - 1 = 0\}.$$

Or, l'ensemble des racines de $\mathrm{X}^m - 1 \in \mathrm{K}[\mathrm{X}]$ est fini, de cardinal au plus m. On a donc $n \leqslant m$. Mais l'inégalité $m \leqslant n$ est également vérifiée puisqu'il existe $z \in \mathrm{G}$ d'ordre m. Finalement on obtient $m = n$ et z est un générateur de G, qui est cyclique. ◁

Il résulte notamment de la question 4 que, pour p premier, le groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$ est cyclique. Le lecteur trouvera une autre preuve de ce résultat très important dans l'exercice 4.13.

2.9. Puissances dans un groupe abélien d'exposant fini

Soit G un groupe abélien. On suppose qu'il existe $n \in \mathbb{N}^*$ tel que $x^n = 1$ pour tout $x \in \mathrm{G}$.

1. On suppose que $n = ab$, avec $\operatorname{pgcd}(a,b) = 1$, et on considère $\mathrm{G}_a = \{x^a,\ x \in \mathrm{G}\}$. Montrer que G_a est un sous-groupe de G. On définit de même G_b. Montrer que pour tout $x \in \mathrm{G}$, il existe un unique $(u,v) \in \mathrm{G}_a \times \mathrm{G}_b$ tel que $x = uv$.

2. Soit k un entier naturel premier avec n. Montrer que l'application $x \longmapsto x^k$ est un automorphisme de G. Déterminer l'application réciproque.

(École normale supérieure)

Solution.

1. Pour tout entier k, l'application $f_k : x \longmapsto x^k$ est un morphisme de groupes de G dans G, car G est abélien. Cela montre que $G_a = \operatorname{Im} f_a$ et $G_b = \operatorname{Im} f_b$ sont des sous-groupes de G.

Considérons le morphisme de groupes $\psi : G_a \times G_b \to G$ qui au couple (u, v) associe uv. Il s'agit de prouver que ψ est un isomorphisme.

• Injectivité. Soit $(u, v) \in G_a \times G_b$ tel que $uv = e$. Comme u s'écrit sous la forme $u = w^a$, on a $u^b = w^{ab} = e$. Donc l'ordre de u divise b. De même, l'ordre de v divise a. Or, $v = u^{-1}$ a le même ordre que u. Cet ordre divise donc a et b et comme a et b sont premiers entre eux, c'est 1. Ainsi $u = v = e$ et $\operatorname{Ker} \psi = \{(e, e)\}$.

• Surjectivité. Soit $x \in G$. Par le théorème de Bezout on peut trouver α et β tels que $1 = \alpha a + \beta b$. On a alors $x = (x^\alpha)^a (x^\beta)^b = \psi((x^\alpha)^a, (x^\beta)^b)$. D'où le résultat.

2. Montrons que f_k est injective en cherchant son noyau. Soit x dans $\operatorname{Ker} f_k$. On a $x^k = e$ et $x^n = e$ par hypothèse. L'ordre de x divise k et n ; c'est donc 1 puisque $\operatorname{pgcd}(n, k) = 1$. Ainsi, $x = e$ et $\operatorname{Ker} f_k = \{e\}$. Pour la surjectivité, on écrit une relation de Bezout entre k et n : $rk + sn = 1$. On a alors, pour tout $x \in G$, $x = x^{rk+sn} = (x^r)^k$. Donc f_k est bijective et $f_k^{-1} = f_r$. ◁

Les actions de groupe jouent un rôle essentiel dans bien des domaines des mathématiques. Tout d'abord en théorie des groupes. En regardant diverses actions d'un groupe G*, on peut obtenir beaucoup d'informations sur* G *lui-même (voir par exemple la preuve du lemme de Cauchy en 2.10). La théorie des représentations linéaires (*i.e. *des opérations linéaires d'un groupe sur un espace vectoriel) forme par exemple un pan entier de la théorie des groupes. Mais cette notion peut aussi se rencontrer en combinatoire lorsqu'on cherche à dénombrer les configurations d'un objet sur lequel un groupe agit : on se reportera par exemple à l'exercice 1.10 ou encore à l'exercice 6.21 qui donne une preuve de la formule de Burnside. Enfin, la notion d'opération de groupe est au cœur de la géométrie. Depuis le programme d'Erlangen de Felix Klein (1872), on comprend une géométrie comme l'étude de propriétés invariantes sous l'action d'un certain groupe.*

Dans le programme de Spéciales, le lecteur pourra avec un peu de recul se rendre compte que beaucoup de résultats sont des descriptions d'orbites pour certaines opérations de groupes : définition des angles orientés, recherche des classes de similitude de $\mathcal{M}_n(\mathrm{K})$*, classification des formes quadratiques...*

Voici un petit rappel de l'essentiel.

• *Une opération d'un groupe* G *sur un ensemble non vide* E *est un morphisme* φ *de* G *dans* $\mathcal{S}_\mathrm{E}$ *le groupe symétrique de* E*. On note usuellement* $g \cdot x$ *au lieu de* $\varphi(g)(x)$ *pour* $(g, x) \in \mathrm{G} \times \mathrm{E}$*.*

• *Si* G *opère sur* E *la relation* $x\mathrm{R}y \iff \exists g \in \mathrm{G},\ g \cdot x = y$ *est une relation d'équivalence sur* E *dont les classes sont appelées orbites. L'opération est dite transitive s'il n'y a qu'une seule orbite. Si* $x \in \mathrm{E}$ *on note* $\mathrm{G}_x = \{g \in \mathrm{G},\ g \cdot x = x\}$ *le stabilisateur de* x*. C'est un sous-groupe de* G*.*

• *Dans les exercices on utilisera plusieurs fois l'important résultat suivant : si* G *est fini et si* x *est un élément de* E*, le cardinal de l'orbite de* x *sous l'action de* G *est égal à l'indice du stabilisateur de* x*. Démontrons-le. Notons* Ω_x *l'orbite de* x*. L'application* $f : \mathrm{G} \to \Omega_x$ *qui à* g *associe* $g \cdot x$ *est surjective par définition. Étudions la relation d'équivalence* R_f *associée à* f *:*

$$g\mathrm{R}_f g' \iff f(g) = f(g') \iff g \cdot x = g' \cdot x \iff (g^{-1}g') \cdot x = x$$
$$\iff g^{-1}g' \in \mathrm{G}_x \iff g' \in g\mathrm{G}_x.$$

La classe d'équivalence de $g \in \mathrm{G}$ *est donc la classe à droite* $g\mathrm{G}_x$ *; elle est de cardinal* $|\mathrm{G}_x|$*. Toutes les classes d'équivalence ont donc le même cardinal. Comme l'ensemble quotient* G/R_f *est en bijection avec* Ω_x *on a bien*

$$\boxed{\operatorname{Card} \Omega_x = \frac{\operatorname{Card} \mathrm{G}}{\operatorname{Card} \mathrm{G}_x}}.$$

• *L'équation aux classes consiste à écrire, lorsque* E *est fini, que le cardinal de* E *est égal à la somme des cardinaux des différentes orbites. Elle permet souvent des raisonnements combinatoires ou arithmétiques comme dans la preuve suivante du lemme de Cauchy donnée par MacKay en 1959.*

2.10. Lemme de Cauchy

1. Soit G un groupe fini de cardinal p^m (avec p premier et m dans $\mathbb{N}^*$) qui opère sur un ensemble fini non vide E. On considère

$$\mathrm{E}^{\mathrm{G}} = \{x \in \mathrm{E},\ \forall g \in \mathrm{G},\ g \cdot x = x\}.$$

Montrer que $|\mathrm{E}^{\mathrm{G}}| \equiv |\mathrm{E}| \pmod{p}$.

2. Soit H un groupe fini d'ordre n et p un diviseur premier de n. Montrer que H contient un élément d'ordre p (*lemme de Cauchy*). On pourra, pour ce faire, utiliser une opération de $\mathbb{Z}/p\mathbb{Z}$ sur l'ensemble E des $(x_1, \ldots, x_p) \in \mathrm{H}^p$ tels que $x_1 x_2 \ldots x_p = e$.

3. Soit H un groupe fini d'ordre n et m dans $\mathbb{N}^*$ tel que, pour tout $x \in \mathrm{H}$, $x^m = e$. Montrer que n divise une puissance de m.

(École normale supérieure)

Solution.

1. Si $x \in \mathrm{E}$, on notera $\mathrm{O}(x)$ l'orbite de x sous l'action de G. Les éléments de E^{G} sont exactement les éléments x de E tels que $\mathrm{O}(x) = \{x\}$. Notons $w_1, w_2, \ldots, w_p$ les orbites de E de cardinal strictement supérieur à 1. On sait que si x_i est un élément de w_i, le cardinal de w_i est l'indice du stabilisateur de x_i dans G : c'est donc une puissance de p. Il résulte de l'équation aux classes que

$$|\mathrm{E}| = |\mathrm{E}^{\mathrm{G}}| + \sum_{i=1}^{p} |w_i| \equiv |\mathrm{E}^{\mathrm{G}}| \quad [p].$$

2. Soit $(x_1, \ldots, x_p)$ un élément de E. On a donc $x_1 x_2 \ldots x_p = e$. En multipliant à gauche par x_1^{-1} et à droite par x_1 il vient $x_2 x_3 \ldots x_p x_1 = e$, c'est-à-dire $(x_2, \ldots, x_p, x_1) \in \mathrm{E}$.

Notons c le cycle $(1, 2, \ldots, p)$ élément du groupe symétrique $\mathcal{S}_p$. Il s'agit d'un élément d'ordre p qui engendre un groupe cyclique K isomorphe à $\mathbb{Z}/p\mathbb{Z}$. On définit une opération de K sur l'ensemble H^p par $c \cdot (x_1, \ldots, x_p) = (x_{c(1)}, x_{c(2)}, \ldots, x_{c(p)}) = (x_2, \ldots, x_p, x_1)$. La remarque ci-dessus montre que E est stable par cette opération. Appliquons alors le résultat de la question précédente à l'opération induite sur E. On a : $|\mathrm{E}| \equiv |\mathrm{E}^{\mathrm{K}}| \ [p]$. Le cardinal de E est n^{p-1} : on peut choisir $x_1, \ldots, x_{p-1}$ quelconques ; x_p est alors déterminé de manière unique. Comme p divise n, $|\mathrm{E}^{\mathrm{K}}|$ est nul modulo p. Or, les éléments de E^{K} sont justement les p-uplets $(x, x, \ldots, x)$ avec $x^p = e$. Comme E^{K} est non vide puisqu'il contient le p-uplet $(e, e, \ldots, e)$, il a un cardinal supérieur à p. Il y a donc au moins $p - 1$ éléments d'ordre p dans H.

3. Il suffit de montrer que tous les facteurs premiers de n sont des facteurs premiers de m. Soit p premier divisant n. Le lemme de Cauchy garantit l'existence d'un élément $x \in \mathrm{H}$ d'ordre p. Or, par hypothèse $x^m = e$. C'est donc que p divise m. ◁

Le défaut de commutativité d'un groupe G *se mesure à l'aide de son centre. Il s'agit du sous-groupe* $\mathrm{Z}(\mathrm{G})$ *formé des éléments de* G *qui commutent avec tous les autres. Le lecteur montrera par exemple que le centre du groupe symétrique* $\mathcal{S}_n$ *est réduit à l'identité pour* $n \geqslant 3$. *L'exercice suivant se propose de prouver que le centre d'un* p*-groupe ne peut pas être trivial.*

2.11. Centre d'un p-groupe

Soit G un groupe fini et $a \in \mathrm{G}$.

1. Montrer que le nombre de conjugués de a est égal à l'indice dans G du centralisateur de a.

On suppose que G est de cardinal p^m avec p premier et $m \geqslant 1$.
2. Montrer que le centre de G est d'ordre p^k avec $0 < k \leqslant m$.
3. On suppose $m = 2$. Montrer que G est abélien.
(École polytechnique)

▷ **Solution.**

1. On considère l'action de G sur lui-même par conjugaison : si g et x sont dans G, $g \cdot x = gxg^{-1}$. L'ensemble Ω_a des conjugués de a est l'orbite de a et le centralisateur de a $\{g \in \mathrm{G},\ ga = ag\}$ n'est autre que le stabilisateur $\{g \in \mathrm{G}, gag^{-1} = a\} = \mathrm{G}_a$ de a pour cette opération. On a donc

$$\boxed{\operatorname{Card} \Omega_a = \frac{\operatorname{Card} \mathrm{G}}{\operatorname{Card} \mathrm{G}_a}}.$$

2. Notons $\mathrm{Z(G)} = \{a \in \mathrm{G},\ \forall g \in \mathrm{G},\ ag = ga\}$ le centre de G. Pour $a \in \mathrm{G}$, nous avons les équivalences :

$$a \in \mathrm{Z(G)} \iff \forall g \in \mathrm{G},\ gag^{-1} = a \iff \Omega_a = \{a\}$$

Écrivons que G est réunion disjointe des classes de conjugaison. La question précédente assure que le cardinal d'une classe de conjugaison est un diviseur de $\operatorname{Card} \mathrm{G} = p^m$. Si elle n'est pas réduite à un singleton, son cardinal est de la forme p^k avec $k \geqslant 1$ donc est nul modulo p. Ainsi, on obtient

$$\operatorname{Card} \mathrm{G} = \sum_{\Omega \text{ orbite}} \operatorname{Card} \Omega = \operatorname{Card} \mathrm{Z(G)} + \sum_{\substack{\Omega \text{ orbite} \\ \operatorname{Card} \Omega \geqslant 2}} \operatorname{Card} \Omega$$

et, modulo p, il reste $0 \equiv \operatorname{Card} \mathrm{G} \equiv \operatorname{Card} \mathrm{Z(G)}$. Donc p divise $\operatorname{Card} \mathrm{Z(G)}$. Or, Z(G) est un sous-groupe de G, donc son cardinal est de la forme p^k avec $0 \leqslant k \leqslant m$, d'après le théorème de Lagrange. Comme p divise $\operatorname{Card} \mathrm{Z(G)}$, k est nécessairement supérieur ou égal à 1.

3. Soit $\varphi : g \in \mathrm{G} \longmapsto \varphi_g \in \mathcal{S}_\mathrm{G}$ où $\mathcal{S}_\mathrm{G}$ est le groupe des permutations de G et φ_g le morphisme de conjugaison $x \longmapsto gxg^{-1}$. On a pour tout $g, g' \in \mathrm{G}$, $\varphi_{gg'} = \varphi_g \circ \varphi'_g$ et l'application φ est un morphisme[1] de G sur un sous-groupe de $\mathcal{S}_\mathrm{G}$, noté Int G et appelé groupe des automorphismes intérieurs de G. Le noyau de φ est exactement le centre de G. Comme ce centre n'est pas réduit à l'élément neutre d'après la question précédente, la relation $|\mathrm{G}| = |\operatorname{Im} \varphi| . |\operatorname{Ker} \varphi|$ (voir exercice 2.4) donne $|\operatorname{Int}\ \mathrm{G}| = 1$ ou p. Comme p est premier, Int G est cyclique dans les deux cas. Il existe

1. Cela traduit le fait que la conjugaison est une opération de groupes.

donc $g_0 \in G$ tel que, pour tout $g \in G$, il existe $k \in \mathbb{Z}$ vérifiant $\varphi_g = \varphi_{g_0}^k$, c'est-à-dire

$$\forall x \in G, \quad gxg^{-1} = g_0^k x g_0^{-k}.$$

Mais alors, g_0 est dans le centre de G puisque $gg_0g^{-1} = g_0^k g_0 g_0^{-k} = g_0$ et $gg_0 = g_0g$. Ainsi φ_{g_0} est l'identité et $|\operatorname{Int} G| = 1$. Donc, $Z(G) = G$ et G est un groupe abélien. ◁

Il existe en revanche des groupes non abéliens d'ordre p^3 et par suite des groupes non abéliens d'ordre p^m pour tout $m \geqslant 3$.

L'exercice suivant évalue la proportion d'éléments d'un groupe non abélien qui commutent, proportion qui est reliée au nombre de classes de conjugaison.

2.12. Nombre de classes de conjugaison

Soit G un groupe fini non abélien. On note $n(G)$ le nombre de couples d'éléments de G qui commutent divisé par le nombre total de couples de G. Montrer que $n(G) \leqslant \frac{5}{8}$.

(École polytechnique)

▷ **Solution.**

• On va commencer par interpréter $n(G)$. Pour tout $a \in G$, on note C_a le centralisateur de a, c'est-à-dire l'ensemble des éléments de G qui commutent avec a. On obtient alors

$$n(G) = \frac{1}{|G|^2} \sum_{a \in G} |C_a|.$$

Or, on sait que l'indice de C_a dans G est égal au cardinal de la classe de conjugaison de a. En notant $w_1, \ldots, w_k$ les classes de conjugaison de G, il vient alors

$$n(G) = \frac{1}{|G|^2} \sum_{i=1}^{k} \sum_{a \in w_i} |C_a| = \frac{1}{|G|^2} \sum_{i=1}^{k} \sum_{a \in w_i} \frac{|G|}{|w_i|} = \frac{1}{|G|^2} \sum_{i=1}^{k} |G| = \frac{k}{|G|}.$$

On est ramené à prouver que pour un groupe fini non commutatif, on a $k \leqslant \frac{5}{8}|G|$. Le résultat est bien entendu faux si G est abélien, puisque dans ce cas il y a $|G|$ classes de conjugaison (réduites à des singletons).

• De manière générale, la classe de conjugaison d'un élément a de G est un singleton si et seulement si a est dans le centre de G, que l'on notera $Z(G)$. Les autres classes ayant au moins deux éléments, on a l'inégalité

$$|Z(G)| + 2(k - |Z(G)|) \leqslant |G|.$$

Celle-ci, divisée par $|G|$ montre que

$$n(G) = \frac{k}{|G|} \leqslant \frac{1}{2} + \frac{|Z(G)|}{2|G|}.$$

Il nous suffit donc, pour conclure, de prouver que $\frac{|Z(G)|}{2|G|} \leqslant \frac{1}{8}$, c'est-à-dire que l'indice de $Z(G)$ dans G est supérieur ou égal à 4.

Nous raisonnons par l'absurde. L'indice de $Z(G)$ dans G est supérieur ou égal à 2, car G n'est pas commutatif.

Supposons d'abord que $Z(G)$ soit d'indice 2 : $|G| = 2|Z(G)|$. On a alors $G = Z(G) \cup aZ(G)$, où a est un élément quelconque n'appartenant pas à $Z(G)$. Par définition, a commute avec les éléments de $Z(G)$. Soit $b \in aZ(G)$, que l'on écrit az, où z est central. Alors a commute avec a lui-même et avec z, donc avec b. Ainsi, a commute avec tous les éléments de G : c'est absurde.

Supposons maintenant que $Z(G)$ soit d'indice 3 : $|G| = 3|Z(G)|$. On considère les classes à droite modulo $Z(G)$: $Z(G)$, $aZ(G)$ et $bZ(G)$. L'élément a commute avec tout élément de $Z(G)$ et avec tout élément de $aZ(G)$ (comme précédemment). Montrons qu'il commute aussi avec les éléments de $bZ(G)$. Pour cela, il suffit de prouver qu'il commute avec b. Or, ab n'est pas dans $aZ(G)$, car b n'est pas central. S'il est dans $bZ(G)$, alors $b^{-1}ab$ est central. On en déduit que $ab = bb^{-1}ab = b^{-1}ab^2$ et donc que $ba = ab$. Sinon, ab est dans $Z(G)$. Mais alors, $aba^{-1} = a^{-1}ab = b$ et donc $ab = ba$. Dans tous les cas, on obtient que a commute avec b, donc est central, ce qui est faux. ◁

En fait, le lecteur connaissant la notion de groupe quotient pourra montrer plus généralement que si $G/Z(G)$ est un groupe cyclique alors G est abélien (et le quotient est donc trivial). Comme tout groupe d'ordre 2 ou 3 est cyclique, il en résulte que dans un groupe non abélien le centre est au moins d'indice 4. On retrouve aussi qu'un groupe d'ordre p^2 avec p premier est cyclique (question 3 de l'exercice 2.11).

2.13. Sous-groupe d'indice infini

Soit G un groupe et H un sous-groupe de G d'indice infini, c'est-à-dire qu'il existe une suite infinie (x_n) d'éléments de G telle que les x_nH soient deux à deux distincts. Soit $H_1, ..., H_p$ des sous-groupes de G tels que $G = H \cup H_1 \cup \cdots \cup H_p$. Montrer que $G = H_1 \cup \cdots \cup H_p$.

(École normale supérieure)

1 - **Solution.**

Raisonnons par l'absurde et supposons l'existence d'un élément h dans $H\setminus(H_1\cup\ldots\cup H_p)$. On considère une suite $(x_n)_{n\in\mathbb{N}}$ de G telle que les x_nH sont deux à deux distincts. En particulier, au plus un des x_nH peut être égal à H et dans ce cas, on peut supposer que c'est le premier x_0H. On supposera donc que, pour $n\geqslant 1$, les x_n ne sont pas dans H.

Pour tout $n\geqslant 1$, il existe $i\in[\![1,p]\!]$ tel que $x_n\in H_i$. D'après le lemme des tiroirs, il existe $i_1\in[\![1,p]\!]$ tel que $x_n\in H_{i_1}$ pour une infinité d'indices n. On peut donc extraire de $(x_n)_{n\geqslant 1}$ une suite $(x_n^{(1)})_{n\geqslant 0}$ d'éléments de $H_{i_1}\setminus H$. Pour $n\neq m$, on a $x_n^{(1)}H\neq x_m^{(1)}H$.

Considérons pour $n\in\mathbb{N}$, $y_n=hx_n^{(1)}$ qui n'est ni dans H (sinon $x_n^{(1)}\in H$), ni dans H_{i_1} (sinon $h\in H_{i_1}$). Pour tout n, y_n est dans l'un des H_i avec $i\neq i_1$. Il existe donc une sous-suite $(z_n)_{n\in\mathbb{N}}$ de $(y_n)_{n\in\mathbb{N}}$ dont les éléments sont dans le même sous-groupe H_{i_2} avec $i_2\neq i_1$. On pose $x_n^{(2)}=z_0^{-1}z_n$ pour $n\in\mathbb{N}$. On a $x_n^{(2)}\in H_{i_2}$ puisque H_{i_2} est un sous-groupe. D'autre part, de $z_0=hx_k^{(1)}$ et $z_n=hx_\ell^{(1)}$ avec $k,\ell\in\mathbb{N}$, on déduit $x_n^{(2)}=\left[x_k^{(1)}\right]^{-1}x_\ell^{(1)}\in H_{i_1}$. De plus, si $m\in\mathbb{N}$ et $z_m=hx_r^{(1)}$, on a

$$x_m^{(2)}H=x_n^{(2)}H\iff z_mH=z_nH\iff x_r^{(1)}H=x_\ell^{(1)}H\iff r=\ell$$

et finalement,

$$x_m^{(2)}H=x_n^{(2)}H\iff m=n.$$

On dispose donc d'une suite $\left(x_n^{(2)}\right)_{n\geqslant 0}$ de $H_{i_1}\cap H_{i_2}$ pour laquelle les $x_n^{(2)}H$ sont deux à deux distincts.

On réitère le procédé : supposons construite pour $q<p$ une suite $\left(x_n^{(q)}\right)_{n\geqslant 0}$ de $H_{i_1}\cap\ldots\cap H_{i_q}$ pour laquelle les $x_n^{(q)}H$ sont deux à deux distincts. Au plus un terme de la suite peut être dans H et on peut supposer que c'est le premier $x_0^{(q)}$.

Considérons pour $n\in\mathbb{N}^*$, $y_n=hx_n^{(q)}$ qui n'est ni dans H (sinon $x_n^{(q)}\in H$), ni dans $H_{i_1}\cap\ldots\cap H_{i_q}$ par hypothèse sur h. Pour tout n, y_n est dans l'un des H_i avec $i\notin\{i_1,\ldots i_q\}$. Il existe donc une sous-suite $(z_n)_{n\in\mathbb{N}}$ de $(y_n)_{n\in\mathbb{N}}$ dont les éléments sont dans le même sous-groupe $H_{i_{q+1}}$ avec $i_{q+1}\notin\{i_1,\ldots,i_q\}$. On pose $x_n^{(q+1)}=z_0^{-1}z_n$ pour $n\in\mathbb{N}$. On a $x_n^{(q+1)}\in H_{i_{q+1}}$ puisque $H_{i_{q+1}}$ est un sous-groupe. De manière analogue à ce qui a été fait précédemment, $x_n^{(q+1)}$ s'écrit après simplification comme produit de deux éléments de $H_{i_1}\cap\ldots\cap H_{i_q}$ et les $x_n^{(q+1)}H$ sont des parties deux à deux distinctes. On dispose donc d'une suite $\left(x_n^{(q+1)}\right)_{n\geqslant 0}$ de $H_{i_1}\cap\ldots\cap H_{i_q}\cap H_{i_{q+1}}$ pour laquelle les $x_n^{(q+1)}H$ sont deux à deux distincts.

Au final, on obtient une suite $\left(x_n^{(p)}\right)_{n\geqslant 0}$ de $\mathrm{H}_{i_1}\cap\ldots\cap\mathrm{H}_{i_p}$, c'est-à-dire de $\mathrm{H}_1\cap\ldots\cap\mathrm{H}_p$ avec les $x_n^{(p)}\mathrm{H}$ deux à deux distincts. On choisit $n\in\mathbb{N}$ tel que $x=x_n^{(p)}\notin\mathrm{H}$. Dans ces conditions, xh ne peut appartenir ni à H, ni à aucun des H_i. Comme $\mathrm{G}=\mathrm{H}\cup\mathrm{H}_1\cup\ldots\cup\mathrm{H}_p$, on aboutit à une contradiction. $\triangleleft$

2.14. Un théorème de Frobenius (1895)

Soit G un groupe fini, H un sous-groupe de G. On suppose que $\mathrm{Card\,G}=p\,\mathrm{Card\,H}$, où p est le plus petit diviseur premier de Card G. Montrer que H est distingué dans G, c'est-à-dire que l'on a, pour tout $x\in\mathrm{G}$, $x\mathrm{H}x^{-1}=\mathrm{H}$.

(**École polytechnique**)

$\triangleright$ **Solution.**

• On considère les classes à gauche modulo H *i.e.* les gH où g parcourt G. On notera G/H l'ensemble de ces classes. Comme elles ont toutes le même cardinal que H, on obtient $\mathrm{Card\,G/H}=\dfrac{\mathrm{Card\,G}}{\mathrm{Card\,H}}$. Le groupe G agit naturellement sur G/H par translation à gauche : si $g\in\mathrm{G}$ et $x\in\mathrm{G}$, $g\cdot x\mathrm{H}=(gx)\mathrm{H}$. Mais on ne peut pas avoir d'information intéressante dans l'écriture de l'équation aux classes pour cette action, puisqu'il n'y qu'une seule orbite de cardinal p. On va plutôt regarder la restriction de cette action à H définie, pour $h\in\mathrm{H}$ et $x\in\mathrm{G}$, par $h\cdot x\mathrm{H}=(hx)\mathrm{H}$.

• On note $(\Omega_i)_{1\leqslant i\leqslant n}$ les orbites de G/H pour cette action. On sait que le cardinal de chaque Ω_i divise celui de H, donc celui de G. Écrivons alors l'équation aux classes qui exprime le fait que G/H est réunion disjointe des orbites sous l'action de H :

$$p=\mathrm{Card\,G/H}=\sum_{i=1}^{n}\mathrm{Card}\,\Omega_i.$$

L'orbite de H est bien sûr réduite au singleton $\{\mathrm{H}\}$. Les orbites qui ne sont pas des singletons ont un cardinal divisant $|\mathrm{G}|$; par hypothèse, un tel cardinal est supérieur ou égal à p. S'il existait une telle orbite, on aurait $\sum_{i=1}^{n}\mathrm{Card}\,\Omega_i\geqslant p+1$, ce qui est impossible. Toutes les orbites sont donc réduites à des singletons.

• Ainsi, pour tout $x\in\mathrm{G}$ et tout $h\in\mathrm{H}$, on a

$$h\cdot x\mathrm{H}=(hx)\mathrm{H}=x\mathrm{H}$$

et donc $x^{-1}hx\in\mathrm{H}$. On obtient, pour tout $x\in\mathrm{G}$, l'inclusion $x^{-1}\mathrm{H}x\subset\mathrm{H}$

et donc l'égalité $x^{-1}\mathrm{H}x = \mathrm{H}$ puisque H et $x^{-1}\mathrm{H}x$ ont le même cardinal, ce qui montre que H est distingué dans G. ◁

L'exercice suivant montre que dans un groupe fini, un sous-groupe propre ne peut pas couper toutes les classes de conjugaison.

2.15. Classes de conjugaison

1. Soit G un groupe fini. H un sous-groupe strict de G. Montrer qu'il existe $x \in \mathrm{G}$ tel que la classe de conjugaison de x ne rencontre pas H.

2. Donner un contre-exemple si G n'est pas fini.

(École normale supérieure)

▷ **Solution.**

1. Pour x et g dans G, on a $gxg^{-1} \in \mathrm{H} \iff x \in g^{-1}\mathrm{H}g$. Le problème revient donc à démontrer que la réunion $\bigcup_{g\in\mathrm{G}} g\mathrm{H}g^{-1}$ des conjugués de H, n'est pas égale à G. Pour cela on va majorer le cardinal de cette réunion et montrer qu'elle contient strictement moins d'éléments que G (la seconde question peut nous faire songer à un argument combinatoire). On observe que si g et g' sont dans la même classe à gauche modulo H, c'est-à-dire s'il existe $h \in \mathrm{H}$ tel que $g' = gh$, alors $g'\mathrm{H}g'^{-1} = g(h\mathrm{H}h^{-1})g^{-1} = g\mathrm{H}g^{-1}$. Dans la réunion ci-dessus, on peut donc se contenter de prendre un système de représentants des classes à gauche modulo H. Soit $g_1, \ldots, g_k$ un tel système de représentants, $k = |\mathrm{G}|/|\mathrm{H}|$ étant l'indice de H dans G. Les conjugués de H ayant au moins l'élément neutre en commun, il vient

$$\left|\bigcup_{g\in\mathrm{G}} g\mathrm{H}g^{-1}\right| = \left|\bigcup_{i=1}^{k} g_i\mathrm{H}g_i^{-1}\right| \leqslant 1 + (|\mathrm{H}|-1)k = |\mathrm{G}| + 1 - \frac{|\mathrm{G}|}{|\mathrm{H}|} < |\mathrm{G}|,$$

car $|\mathrm{H}| < |\mathrm{G}|$ par hypothèse.

2. La preuve précédente utilise fortement la finitude de G. Le résultat ne s'étend pas à un groupe infini. Prenons par exemple $\mathrm{G} = \mathrm{GL}_n(\mathbb{C})$ et H le sous-groupe de G formé des matrices triangulaires supérieures inversibles. Toute matrice de $\mathrm{M}_n(\mathbb{C})$ étant trigonalisable, la classe de conjugaison de toute matrice A de G rencontre H. ◁

Une manière très féconde pour obtenir des groupes consiste à regarder le groupe de symétrie d'un «objet géométrique». Ainsi, dans l'espace euclidien de dimension 3, la recherche des sous-groupes finis du groupe des

rotations est-elle liée aux polyèdres réguliers. Ces derniers, au nombre de 5, étaient déjà connus de Platon. L'exercice suivant propose la partie analyse du travail et détermine les caractéristiques que doit avoir un sous-groupe fini de $\mathrm{SO}_3(\mathbb{R})$.

2.16. Sous-groupes finis de $\mathrm{SO}_3(\mathbb{R})$

Soit X un ensemble et G un sous-groupe fini du groupe $\mathcal{S}_\mathrm{X}$ des bijections de X. Pour $y \in \mathrm{X}$, on note $\mathrm{X}_y = \{g(y),\ g \in \mathrm{G}\}$ l'orbite de y sous l'action de G et $\mathrm{G}_y = \{g \in \mathrm{G},\ g(y) = y\}$ le stabilisateur de y dans G.

1. Pour tout $y \in \mathrm{X}$, prouver que $|\mathrm{G}| = |\mathrm{G}_y| \times |\mathrm{X}_y|$.

2. Montrer que pour $(x, y) \in \mathrm{X}^2$, si $x \in \mathrm{X}_y$, alors $\mathrm{X}_x = \mathrm{X}_y$.

3.a. Désormais, X est la sphère unité de $\mathbb{R}^3$ muni de sa structure euclidienne canonique et G un sous-groupe fini du groupe des rotations. On note $\mathcal{T} = \{\mathrm{X}_y, y \in \mathrm{X}\}$ et pour $\mathrm{T} = \mathrm{X}_y \in \mathcal{T}$, $\nu(\mathrm{T}) = |\mathrm{G}_y|$. Prouver que

$$2|\mathrm{G}| - 2 = \sum_{\mathrm{T} \in \mathcal{T}} (\nu(\mathrm{T}) - 1)|\mathrm{T}|.$$

b. Quelles sont les valeurs possibles de $\nu(\mathrm{T})$?

(École polytechnique)

▷ **Solution.**

1. Cela a été rappelé dans le préambule aux exercices sur les opérations de groupes.

2. Soit $(x, y) \in \mathrm{X}^2$, avec $x \in \mathrm{X}_y$. On écrit $x = g_0(y)$ avec $g_0 \in \mathrm{G}$. On a alors $y = g_0^{-1}(x)$ et $y \in \mathrm{X}_x$. Si $z \in \mathrm{X}_x$, il existe $g \in \mathrm{G}$ tel que $z = g(x)$, d'où résulte $z = gg_0(y)$ et $z \in \mathrm{X}_y$. Ainsi $\mathrm{X}_x \subset \mathrm{X}_y$. Comme $y \in \mathrm{X}_x$, par symétrie du problème, on a $\mathrm{X}_y \subset \mathrm{X}_x$ et finalement l'égalité demandée.

3.a. Avant tout, remarquons que $|\mathrm{G}_y|$ ne dépend pas du choix de l'élément y, ce qui justifie la définition de $\nu(\mathrm{T})$. En effet, si $\mathrm{T} = \mathrm{X}_x = \mathrm{X}_y$, on a

$$|\mathrm{G}_x| = |\mathrm{G}|/|\mathrm{X}_x| = |\mathrm{G}|/|\mathrm{X}_y| = |\mathrm{G}_y|.$$

Rappelons qu'une rotation de $\mathbb{R}^3$ qui n'est pas l'identité admet une droite vectorielle (son axe) comme ensemble de points invariants. Par conséquent, si on note $\mathrm{G}' = \mathrm{G}\backslash\{\mathrm{I}\}$, où I désigne l'identité de $\mathbb{R}^3$, tout élément de G′ admet exactement deux points fixes dans X (les points d'intersection de l'axe de la rotation avec la sphère X). Considérons

$$\mathrm{E} = \{(g, x) \in \mathrm{G}' \times \mathrm{X},\ g(x) = x\}$$

D'après ce que nous venons de dire, on a

$$|\mathrm{E}| = 2|\mathrm{G}'| = 2(|\mathrm{G}| - 1) = 2|\mathrm{G}| - 2.$$

Si on note $\mathrm{X}' = \{x \in \mathrm{X},\ \exists g \in \mathrm{G}',\ g(x) = x\}$, on obtient

$$\mathrm{E} = \coprod_{x \in \mathrm{X}'} \{(g, x) \in \mathrm{G}' \times \mathrm{X},\ g(x) = x\},$$

le symbole $\coprod$ signifiant que la réunion est disjointe.

Or, si $x \in \mathrm{X}$ est fixé, $\{(g, x) \in \mathrm{G}' \times \mathrm{X},\ g(x) = x\} = (\mathrm{G}_x \backslash \{\mathrm{I}\}) \times \{x\}$ est de cardinal $\nu(\mathrm{X}_x) - 1$. En particulier, $x \in \mathrm{X}'$ si $\nu(\mathrm{X}_x) > 1$. On en déduit que

$$|\mathrm{E}| = \sum_{x \in \mathrm{X}'} (\nu(\mathrm{X}_x) - 1).$$

Puisque, pour $x \in \mathrm{X} \setminus \mathrm{X}'$, on a $\nu(\mathrm{X}_x) - 1 = 0$, on peut encore écrire

$$|\mathrm{E}| = \sum_{x \in \mathrm{X}} (\nu(\mathrm{X}_x) - 1).$$

Sachant que X est réunion disjointe des orbites T et que, quand x varie dans une orbite T, $\nu(\mathrm{X}_x)$ reste égal à $\nu(\mathrm{T})$, on obtient

$$|\mathrm{E}| = \sum_{\mathrm{T} \in \mathcal{T}} (\nu(\mathrm{T}) - 1)|\mathrm{T}|.$$

On conclut que

$$2|\mathrm{G}| - 2 = \sum_{\mathrm{T} \in \mathcal{T}} (\nu(\mathrm{T}) - 1)|\mathrm{T}|.$$

b. La dernière relation peut s'écrire

$$(*) \quad 2 - \frac{2}{|\mathrm{G}|} = \sum_{\mathrm{T} \in \mathcal{T}} \left(1 - \frac{1}{\nu(\mathrm{T})}\right)$$

puisque $|\mathrm{G}| = \nu(\mathrm{T})|\mathrm{T}|$, pour toute orbite T.

La quantité $2 - \dfrac{2}{|\mathrm{G}|}$ est strictement inférieure à 2 et pour tout $\mathrm{T} \in \mathcal{T}$ tel que $\nu(\mathrm{T}) \geqslant 2$, on a $1 - \dfrac{1}{\nu(\mathrm{T})} \geqslant \dfrac{1}{2}$. Par conséquent, il ne peut y avoir plus de trois termes non nuls dans la somme $\displaystyle\sum_{\mathrm{T} \in \mathcal{T}} \left(1 - \frac{1}{\nu(\mathrm{T})}\right)$.

• Imaginons qu'il n'y a qu'une seule orbite avec $\nu(\mathrm{T}) \geqslant 2$. Alors on a

$$0 < \frac{1}{\nu(\mathrm{T})} = \frac{2}{|\mathrm{G}|} - 1 \leqslant 0,$$

ce qui est impossible. Il y a donc deux ou trois orbites T telles que $\nu(\mathrm{T}) \geqslant 2$.

• Supposons qu'il y a deux orbites T avec $\nu(T) \geqslant 2$: T_1 et T_2. N'oublions pas que $\nu(T_1)$ et $\nu(T_2)$ divisent $|G|$ d'après le théorème de Lagrange. Écrivons $|G| = \nu(T_1)d_1 = \nu(T_2)d_2$. Alors $(*)$ devient

$$2 - \frac{2}{|G|} = 2 - \frac{d_1}{|G|} - \frac{d_2}{|G|},$$

d'où résulte $2 = d_1 + d_2$ et nécessairement $d_1 = d_2 = 1$. Les orbites T_1 et T_2 ont un seul élément.

Ainsi, on a deux points de X qui sont invariants par tous les éléments de G, ce qui signifie qu'ils sont sur les axes des rotations de G'. Ces deux points sont opposés et finalement toutes les rotations de G' admettent le même axe D. Le groupe G est alors abélien. Les restrictions des éléments de G au plan $P = D^{\perp}$ forment un sous-groupe fini de $SO(P)$. Nous savons qu'un tel sous-groupe est cyclique, engendré par une rotation d'angle $\frac{2\pi}{n}$ (P étant muni d'une orientation arbitraire). G est donc constitué des rotations d'axe fixe D et d'angle $\frac{2k\pi}{n}$, $0 \leqslant k \leqslant n-1$.

• Supposons qu'il y a trois orbites T avec $\nu(T) \geqslant 2$: T_1, T_2 et T_3. Notons $\nu_i = \nu(T_i)$ pour tout $1 \leqslant i \leqslant 3$. On peut supposer $\nu_1 \leqslant \nu_2 \leqslant \nu_3$. L'égalité $(*)$ s'écrit alors

$$1 + \frac{2}{|G|} = \frac{1}{\nu_1} + \frac{1}{\nu_2} + \frac{1}{\nu_3}.$$

On ne peut avoir $\nu_1 \geqslant 3$ car sinon

$$1 < 1 + \frac{2}{|G|} = \frac{1}{\nu_1} + \frac{1}{\nu_2} + \frac{1}{\nu_3} \leqslant 3.\frac{1}{3} = 1.$$

On a nécessairement $\nu_1 = 2$ et $(*)$ s'écrit $\frac{1}{2} + \frac{2}{|G|} = \frac{1}{\nu_2} + \frac{1}{\nu_3}$.
On ne peut avoir $\nu_2 \geqslant 4$ car sinon

$$\frac{1}{2} < \frac{1}{2} + \frac{2}{|G|} = \frac{1}{\nu_2} + \frac{1}{\nu_3} \leqslant 2.\frac{1}{4} = \frac{1}{2}.$$

Donc ν_2 vaut 2 ou 3.

⋆ Supposons $\nu_2 = 2$. Alors $\nu_3 = |G|/2$ et $|G|$ est pair.

⋆ Supposons $\nu_2 = 3$. Alors $(*)$ devient

$$\frac{1}{6} + \frac{2}{|G|} = \frac{1}{\nu_3}$$

et on a nécessairement $\nu_3 < 6$ et ν_3 vaut 3, 4 ou 5. La formule donne alors la valeur de $|G|$.

On conclut donc sur les valeurs possibles de $\nu(\mathrm{T})$ dans le cas de trois orbites :

Cas	ν_1	ν_2	ν_3	$\lvert\mathrm{G}\rvert$
1	2	2	$n/2$	n
2	2	3	3	12
3	2	3	4	24
4	2	3	5	60

Pour achever l'étude, il reste à effectuer un important travail de synthèse. On peut prouver qu'il existe bien un sous-groupe correspondant à ces quatre derniers cas-là comme on l'a fait dans le cas où on avait deux orbites T *avec* $\nu(\mathrm{T}) \geqslant 2$. *Le cas* 1 *est réalisé par le groupe des rotations laissant stables un polygone régulier à* $n/2$ *cotés, contenu dans un plan. On parle alors de groupe diédral. Le cas* 2 *est réalisé par le groupe des rotations laissant stable un tétraèdre régulier ; il est isomorphe à* $\mathcal{A}_4$. *Le cas* 3 *est réalisé par le groupe des rotations laissant stable le cube, qui est isomorphe à* $\mathcal{S}_4$ *et enfin le dernier cas est obtenu par le groupe des rotations laissant stable le dodécaèdre régulier, qui est isomorphe à* $\mathcal{A}_5$. *Les deux polyèdres réguliers restants sont l'octaèdre et l'icosaèdre qui ont respectivement le même groupe que le cube et le dodécaèdre.*

2.17. Groupe dérivé

Soit G un groupe (noté multiplicativement). On note D le sous-groupe de G engendré par les éléments de la forme $xyx^{-1}y^{-1}$ (avec x, y dans G) et C le sous-groupe de G engendré par les éléments de la forme x^2 ($x \in \mathrm{G}$).

1. Montrer que $\mathrm{D} \subset \mathrm{C}$.

2. On suppose que G est engendré par les éléments x qui vérifient $x = x^{-1}$. Montrer que $\mathrm{C} = \mathrm{D}$.

3. Soit $\mathrm{G} = \mathrm{O}_2(\mathbb{Q})$ le groupe multiplicatif des matrices $(2,2)$ qui sont orthogonales et à coefficients dans $\mathbb{Q}$. Montrer que dans ce cas D est un sous-groupe strict de $\mathrm{SO}_2(\mathbb{Q})$.

(**École normale supérieure**)

Solution.

1. Pour montrer que D est inclus dans C il suffit de prouver que tout élément $z \in \mathrm{G}$ de la forme $z = xyx^{-1}y^{-1}$ est dans C. Il suffit pour cela d'observer que

$$z = xyx^{-1}y^{-1} = x^2(x^{-1}y)^2(y^{-1})^2.$$

Indiquons qu'un tel élément z s'appelle un commutateur et que le groupe D est appelé le groupe dérivé de G.

2. Si x, y sont dans G, nous noterons $[x, y] = xyx^{-1}y^{-1}$ le commutateur associé. Comme précédemment, pour prouver que C = D, il nous suffit de montrer que tout carré de G est dans D. Soit $x \in$ G. Par hypothèse, il existe des involutions $x_1, \ldots, x_p$ (c'est-à-dire des éléments vérifiant $x_i = x_i^{-1}$) tels que $x = x_1 x_2 \ldots x_p$. On va travailler par récurrence sur p. Si $p = 1$, on a $x^2 = x_1^2 = e$, où e est le neutre de G. Si $p = 2$, on a

$$x^2 = x_1 x_2 x_1 x_2 = x_1 x_2 x_1^{-1} x_2^{-1} = [x_1, x_2] \in \mathrm{D}.$$

Supposons que le carré d'un produit de $p-1$ involutions soit dans D. On a alors,

$$\begin{aligned} x^2 &= (x_1 \ldots x_p)(x_1 \ldots x_p) = x_1(x_2 \ldots x_p)x_1(x_2 \ldots x_p)^{-1}(x_2 \ldots x_p)^2 \\ &= [x_1, x_2 \ldots x_p](x_2 \ldots x_p)^2 \in \mathrm{D} \end{aligned}$$

en tant que produit d'un commutateur et d'un élément qui est dans D par hypothèse de récurrence.

3. Notons que $\mathrm{G} = \mathrm{O}_2(\mathbb{Q})$ est bien un sous-groupe du groupe orthogonal $\mathrm{O}_2(\mathbb{R})$: il est non vide, stable par le produit puisque le produit de deux matrices à coefficients dans $\mathbb{Q}$ est encore une matrice à coefficients dans $\mathbb{Q}$, et stable par inverse puisque l'inverse d'une matrice orthogonale est simplement sa transposée. Les éléments de G de déterminant -1 sont des involutions (des symétries axiales géométriquement). Nous savons de plus, que n'importe quelle matrice de $\mathrm{SO}_2(\mathbb{R})$ peut se décomposer comme un produit de deux symétries, l'une des deux pouvant être choisie arbitrairement. En particulier, si $\mathrm{A} \in \mathrm{SO}_2(\mathbb{Q})$ on peut décomposer A comme un produit de deux involutions de G puisqu'il suffit d'écrire $\mathrm{A} = \mathrm{S}(\mathrm{SA})$ avec $\mathrm{S} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. La condition de la question 2 est donc vérifiée : G est bien engendré par les involutions.

Il est évidemment clair que D est inclus dans $\mathrm{SO}_2(\mathbb{Q})$ (un commutateur à un déterminant égal à 1). Supposons par l'absurde que $\mathrm{D} = \mathrm{C} = \mathrm{SO}_2(\mathbb{Q})$. Comme tout élément de $\mathrm{G} \setminus \mathrm{SO}_2(\mathbb{Q})$ a un carré égal au neutre, C est engendré par les carrés des éléments de $\mathrm{SO}_2(\mathbb{Q})$. Mais comme $\mathrm{SO}_2(\mathbb{Q})$ est commutatif, un produit de carrés est encore un carré et tout élément de $\mathrm{SO}_2(\mathbb{Q})$ est alors un carré. On va voir que ce n'est pas le cas. Considérons $\mathrm{M} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ la matrice qui correspond à la rotation d'angle $\frac{\pi}{2}$ et supposons qu'il existe $\mathrm{N} \in \mathrm{G}$ telle que $\mathrm{N}^2 = \mathrm{M}$. On a forcément $\mathrm{N} \in \mathrm{SO}_2(\mathbb{Q})$ et on peut écrire $\mathrm{N} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ avec $(a, b) \in \mathbb{Q}^2$.

Mais on a alors $(a+ib)^2 = i$, ce qui donne $a+ib = \pm\left(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right)$. C'est absurde car $\sqrt{2}$ est irrationnel. ◁

Les exercices qui suivent concernent l'étude de quelques exemples de groupes.

2.18. Groupes quasi-cycliques de Prüfer

Soit $p \in \mathbb{N}^*$ un entier premier et U_p le groupe multiplicatif de $\mathbb{C}^*$ engendré par l'ensemble des nombres $\exp\left(\frac{2i\pi}{p^\alpha}\right)$ où α décrit $\mathbb{N}$.

1. Quels sont les sous-groupes de U_p ?

2. Montrer que U_p est indécomposable, c'est-à-dire n'est pas isomorphe à un produit direct de deux groupes non triviaux.

(**École normale supérieure**)

▷ **Solution.**

1. Pour $\alpha \in \mathbb{N}$, notons G_α le sous-groupe de $\mathbb{C}^*$ engendré par $\exp\left(\frac{2i\pi}{p^\alpha}\right)$. C'est le groupe des racines p^α-ièmes de l'unité. On a, pour tout $\alpha \in \mathbb{N}$, $G_\alpha \subset G_{\alpha+1}$, car $\exp\left(\frac{2i\pi}{p^\alpha}\right) = \left(\exp\left(\frac{2i\pi}{p^{\alpha+1}}\right)\right)^p$ et *a fortiori*, $G_\beta \subset G_\alpha$ si $\beta \leqslant \alpha$. On en déduit que $\bigcup_{\alpha\in\mathbb{N}} G_\alpha$ est un sous-groupe de $\mathbb{C}^*$. En effet, deux éléments de cette réunion appartiennent à un même G_α, qui contient aussi leur produit. On a donc

$$U_p = \bigcup_{\alpha\in\mathbb{N}} G_\alpha.$$

Examinons les sous-groupes de U_p. Il y a évidemment les G_α, qui sont finis et U_p lui-même. On va montrer que ce sont les seuls.

Tout d'abord, pour tout $\alpha \in \mathbb{N}$, les sous-groupes de G_α sont les G_β avec $\beta \leqslant \alpha$. Il a déjà été noté que ce sont des sous-groupes de G_α. Inversement, si G est un sous-groupe de G_α, le cardinal de G divise celui de G_α d'après le théorème de Lagrange, et il existe donc $\beta \leqslant \alpha$ tel que $|G| = p^\beta$. Mais alors, on a $x^{p^\beta} = 1$ pour tout x de G (toujours d'après le théorème de Lagrange) et donc $G \subset G_\beta$. Les deux groupes ayant même cardinal, l'inclusion est une égalité.

Soit maintenant un sous-groupe G de U_p qui n'est pas un G_α. D'après ce qui précède, G n'est contenu dans aucun G_α. Pour tout $\alpha \in \mathbb{N}$, il existe $x \in G$ tel que $x \notin G_\alpha$. Considérons le plus petit entier naturel n tel que $x \in G_n$. On note que $\alpha < n$ (car sinon, on aurait $x \in G_n \subset G_\alpha$).

On écrit $x = \exp\left(\frac{2ik\pi}{p^n}\right)$, avec $k \in \mathbb{Z}$. Puisque, par définition de n, x n'appartient pas à G_{n-1}, on a $x^{p^{n-1}} = \exp\left(\frac{2ik\pi}{p}\right) \neq 1$. On en déduit que p ne divise pas k ; p étant un nombre premier, k est premier avec p et donc premier avec p^n. Mais ceci est la condition pour que x engendre G_n. Comme $x \in G$, G_n est contenu dans G. On a, *a fortiori*, $G_\alpha \subset G$, car $\alpha < n$ implique $G_\alpha \subset G_n$. Puisque α est un entier quelconque, on en déduit que G, qui contient tous les G_α, est égal à U_p.

Il en résulte que l'ensemble des sous-groupes de U_p est totalement ordonné par l'inclusion.

2. Supposons qu'il existe deux groupes H, K et un isomorphisme φ de $H \times K$ sur U_p. Posons $\tilde{H} = H \times \{1\}$: c'est un sous-groupe de $H \times K$ isomorphe à H. De même $\tilde{K} = \{1\} \times K$ est isomorphe à K et $\tilde{H} \cap \tilde{K} = \{(1,1)\}$. Les images par φ des sous-groupes $\tilde{H}$ et $\tilde{K}$ sont des sous-groupes de U_p d'intersection égale à $\{1\}$. D'après le point précédent l'un des deux est réduit à $\{1\}$. Donc soit H, soit K est trivial. ◁

2.19. Le groupe modulaire

Soit $P = \{z \in \mathbb{C},\ \operatorname{Im} z > 0\}$ le *demi-plan de Poincaré*. À toute matrice $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de $SL_2(\mathbb{Z})$ on associe la fonction homographique f_A définie pour $z \in P$ par $f_A(z) = \frac{az+b}{cz+d}$.

1. Montrer que l'application $\psi : A \longmapsto f_A$ est un morphisme de groupes de $SL_2(\mathbb{Z})$ dans le groupe des permutations de P. Déterminer $\operatorname{Ker}\psi$. On note G l'image de ψ ; on l'appelle le *groupe modulaire*.

2. Pour $z \in P$, on pose $I_z = \{\operatorname{Im} g(z),\ g \in G\}$. Montrer que I_z est une partie de $\mathbb{R}_+^*$ admettant un plus grand élément. (On pourra prouver que pour tout réel $\alpha > 0$, l'ensemble des $g \in G$ tels que $\operatorname{Im} g(z) > \alpha$ est fini).

3. On note G_0 le sous-groupe de G engendré par $u : z \longmapsto z+1$ et $v : z \longmapsto -\frac{1}{z}$. Soit $z \in P$. Montrer qu'il existe $g_0 \in G_0$ tel que, pour tout $g \in G_0$, $\operatorname{Im} g(z) \leqslant \operatorname{Im} z_0$, où $z_0 = g_0(z)$.

4. Soit $D = \{z \in P,\ |\operatorname{Re} z| \leqslant \frac{1}{2},\ |z| \geqslant 1\}$ le *domaine fondamental du groupe modulaire*. Montrer qu'on peut choisir z_0 dans D.

5. En déduire que $G_0 = G$.

(École normale supérieure)

Solution.

1. Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$, $B = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in SL_2(\mathbb{Z})$ et $z \in P$. On a

$$f_A(z) = \frac{az+b}{cz+d} = \frac{(az+b)(c\overline{z}+d)}{|cz+d|^2} = \frac{ac|z|^2 + bd + adz + bc\overline{z}}{|cz+d|^2}$$

Il en résulte, en prenant la partie imaginaire, que

$$\mathrm{Im}(f_A(z)) = \frac{(ad-bc)\,\mathrm{Im}\,z}{|cz+d|^2} = \frac{\mathrm{Im}\,z}{|cz+d|^2} > 0.$$

On en déduit que $f_A(P) \subset P$. On a ensuite

$$f_A(f_B(z)) = \frac{a\dfrac{a'z+b'}{c'z+d'} + b}{c\dfrac{a'z+b'}{c'z+d'} + d} = \frac{(aa'+bc')z + (ab'+bd')}{(ca'+dc')z + cb' + dd'} = f_{AB}(z).$$

Comme $f_I = \mathrm{Id}_P$, l'application f_A est bijective et sa bijection réciproque est $f_{A^{-1}}$. Le résultat précédent prouve alors que ψ est un morphisme de groupes de $SL_2(\mathbb{Z})$ dans le groupe des permutations de P.

Supposons que $A \in \mathrm{Ker}\,\psi$. On a $f_A(z) = z$ pour tout $z \in P$ ce qui donne pour tout $z \in P$, $az + b = cz^2 + dz$. D'où l'on déduit $c = b = 0$ et $a = d$. Comme $ad - bc = 1$ on a $a = \pm 1$. Il en résulte que $\mathrm{Ker}\,\psi = \{\pm I\}$.

2. Soit $z \in P$ et $\alpha > 0$. On suit l'indication de l'énoncé. Pour $g \in G$, $\mathrm{Im}\,g(z) \geqslant \alpha$ équivaut à $|cz+d|^2 \leqslant \dfrac{\mathrm{Im}\,z}{\alpha}$ où c, d sont les coefficients de la seconde ligne d'une matrice $A \in SL_2(\mathbb{Z})$ telle que $f_A = g$ (cette matrice est unique au signe près d'après la question précédente). Il n'y a qu'un nombre fini de couples $(c,d) \in \mathbb{Z}^2$ vérifiant cette inégalité. En effet, on a

$$c^2(\mathrm{Im}\,z)^2 \leqslant |cz+d|^2 \leqslant \frac{\mathrm{Im}\,z}{\alpha} \text{ et donc } c^2 \leqslant \frac{1}{\alpha\,\mathrm{Im}\,z}, \text{ puis}$$
$$d^2 \leqslant |cz|^2 + |cz+d|^2 \leqslant c^2|z|^2 + \frac{\mathrm{Im}\,z}{\alpha}.$$

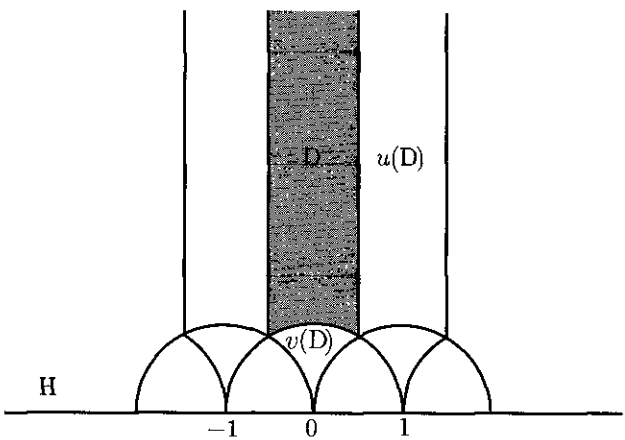
Il en résulte que $I_z \cap [\alpha, +\infty[$, qui est l'ensemble des nombres réels qui s'écrivent $\dfrac{\mathrm{Im}\,z}{|cz+d|^2}$, avec $(c,d) \in \mathbb{Z}^2$ vérifiant l'inégalité ci-dessus, est fini (éventuellement vide).

On en déduit que I_z admet un plus grand élément : si on choisit $\alpha > 0$ tel que $I_z \cap [\alpha, +\infty[$ soit non vide (on prend $\alpha = g(z)$ avec g élément quelconque de G), le plus grand élément de I_z est le plus grand élément de $I_z \cap [\alpha, +\infty[$.

3. Posons $J_z = \{\mathrm{Im}\,g(z), \quad z \in G_0\}$. Comme $J_z \subset I_z$, pour tout $\alpha > 0$, $J_z \cap [\alpha, +\infty[$ est aussi fini. Donc J_z admet également un plus grand élément. D'où l'existence de $g_0 \in G_0$ tel que l'on ait, pour tout $g \in G_0$, $\mathrm{Im}\,g(z) \leqslant \mathrm{Im}\,z_0 = \mathrm{Im}\,g_0(z)$.

4. Supposons que $\operatorname{Re} z_0 = m + t$ où $m \in \mathbb{Z}$ et $t \in \left]-\frac{1}{2}, \frac{1}{2}\right]$. Alors $u^{-m}(z_0)$ admet t comme partie réelle et possède la même partie imaginaire que z_0. Comme $u^{-m}(z_0)$ est égal à $(u^{-m} \circ g_0)(z)$, on peut donc en remplaçant $g_0(z)$ par $(u^{-m} \circ g_0)(z)$, supposer $|\operatorname{Re} z_0| \leqslant \frac{1}{2}$. Mais alors, on a obligatoirement $|z_0| \geqslant 1$. En effet, on observe que $\operatorname{Im} v(z_0) = \dfrac{\operatorname{Im} z_0}{|z_0|^2}$ de sorte que si $|z_0| < 1$, on a $\operatorname{Im}(v \circ g_0)(z) > \operatorname{Im} z_0$, ce qui contredit le choix de g_0.

5. Représentons D.



Les groupes G et G_0 opèrent naturellement sur P. La question précédente a montré que la G_0-orbite de tout point z de P rencontre ce domaine D. On va maintenant regarder si deux points z et z' de D peuvent être dans la même G-orbite.

Supposons donc qu'il existe $g \in G$ tel que $z' = g(z)$. Quitte à remplacer le couple (z, z') par le couple (z', z), où $z = g^{-1}(z')$, on peut supposer que $\operatorname{Im} g(z) \geqslant \operatorname{Im} z$. Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z})$ telle que $g = f_A$. On a $|cz + d|^2 \leqslant 1$ car $\operatorname{Im} g(z) \geqslant \operatorname{Im} z$. En posant $z = x + iy$, avec $(x, y) \in \mathbb{R}^2$, on obtient

$$1 \geqslant (cx + d)^2 + c^2y^2 \geqslant c^2y^2 \geqslant \frac{3}{4}c^2,$$

car $z \in D$. On en déduit que $|c| < 2$, ce qui donne $c \in \{1, 0, -1\}$. Comme il est loisible de changer les signes des coefficients de A, le cas $c = -1$ se ramène au cas $c = 1$.

• Supposons $c = 0$. On a alors $ad = 1$ et donc $a = d = \pm 1$. Dans ce cas, g est une translation : $g(z) = z + k$, avec $k \in \mathbb{Z}$. Si $k = 0$, $g = \mathrm{Id}_P$ et $z' = z$. Si $k \neq 0$, $\operatorname{Re} z$ et $\operatorname{Re} z'$ étant tous les deux dans l'intervalle

$\left]-\frac{1}{2}, \frac{1}{2}\right]$, on a obligatoirement $k = \pm 1$; l'un des deux nombres $\mathrm{Re}\, z$ et $\mathrm{Re}\, z'$ doit être égal à $-\frac{1}{2}$ et l'autre à $\frac{1}{2}$.

• Supposons $c = 1$. De $|z + d| \leqslant 1$, on déduit que

$$1 \cdot (x + d)^2 + y^2 = |z|^2 + 2dx + d^2 \geqslant 1 - 2|d||x| + d^2 \geqslant 1 - |d| + d^2$$

et donc $d^2 \leqslant |d|$. Cela impose $d \in \{0, -1, 1\}$. On a alors $1 - |d| + d^2 = 1$ et les inégalités précédentes sont des égalités. Cela entraîne

$$|z| = 1 \quad \text{et} \quad 2dx = -2|d||x| = -|d|.$$

⋆ Si $d = 0$, on obtient $b = -1$ et $z' = g(z) = a - \frac{1}{z} = a - \overline{z}$. En particulier $\mathrm{Re}(z') = a - \mathrm{Re}(z)$. Si $\mathrm{Re}(z) \neq \pm\frac{1}{2}$, c'est-à-dire si $z \neq j$ et $z \neq -j^2$, alors $a = 0$ et $z' = -\overline{z}$. Si $z = j$, on peut avoir $a = 0$ ou -1 ; on obtient alors $z' = -j^2$ ou j. De même, si $z = -j^2$, alors $z' = j$ ou j^2. Dans tous les cas, on note que $|z'| = 1$.

⋆ Si $d = 1$, on a $|z| = 1$ et $x = -\frac{1}{2}$, c'est-à-dire $z = j$. On obtient alors $b = 1$ et $z' = g(z) = a - \frac{1}{z+1} = a - \frac{1}{j+1} = a + j$. Ceci nécessite $a = 0$ ou 1 et $z' = j$ ou $-j^2$.

⋆ Si $d = -1$, on obtient de même, $z = -j^2$ et $z' = j$ ou $-j^2$.

On observe que, dans tous les cas, si $z \neq z'$, z et z' sont sur la *frontière* de D. De plus, si z n'est pas sur la frontière de D, la seule application g telle que $z = g(z)$ est Id_P. Nous allons déduire de cela que $G = G_0$.

En effet, prenons $g \in G$ et z un point intérieur à D (par exemple $z = 2i$). Puisque $g(z) \in P$, la question 4 montre qu'il existe $g_0 \in G_0$ tel que $g_0(g(z)) \in D$. D'après ce qu'on vient de voir, cela impose $g_0 \circ g(z) = z$ et $g_0 \circ g = \mathrm{Id}_P$. On a donc $g = g_0^{-1}$ et $g \in G_0$. ◁

On déduit de ce résultat que le groupe $\mathrm{SL}_2(\mathbb{Z})$ *est engendré par les deux matrices* $U = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ *et* $V = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. *On trouvera une autre preuve de ce résultat par des manipulations élémentaires dans le chapitre sur le groupe linéaire dans le tome 2 d'algèbre.*

Un problème majeur de la théorie des groupes finis, qui n'est pas encore totalement résolu de nos jours, est celui de la classification à isomorphisme près. Après un travail colossal de plusieurs dizaines d'années la classification est terminée pour les groupes finis simples, c'est-à-dire sans sous-groupe distingué non trivial. Avec les outils du programme des classes préparatoires on peut s'attaquer à un problème incomparablement

plus facile : la classification des groupes finis abéliens. L'énoncé ci-après démontre la partie unicité du théorème suivant :

Théorème. *Si* G *est un groupe abélien fini avec* $|\mathrm{G}| \geqslant 2$, *il existe une unique suite* $(a_1, \ldots, a_n)$ *telle que* $2 \leqslant a_1 | a_2 | \cdots | a_n$ *et* $\mathrm{G} \simeq \prod_{i=1}^{n} \mathbb{Z}/a_i\mathbb{Z}$.

L'entier a_n *n'est autre que l'exposant du groupe* G *(cf. exercice 2.8). Pour la partie existence nous renvoyons le lecteur à l'exercice 7.22.*

2.20. Unicité dans le théorème de structure des groupes abéliens finis

Soit $(a_1, \ldots, a_n, b_1, \ldots, b_m) \in (\mathbb{N}^*)^{m+n}$ vérifiant

$$2 \leqslant a_1 | a_2 | \cdots | a_n \quad \text{et} \quad 2 \leqslant b_1 | b_2 | \cdots | b_m.$$

On suppose qu'il existe un isomorphisme de groupes

$$\prod_{i=1}^{n} \mathbb{Z}/a_i\mathbb{Z} \simeq \prod_{i=1}^{m} \mathbb{Z}/b_i\mathbb{Z}.$$

Montrer que $m = n$ et que $a_i = b_i$ pour tout $i \in [\![1, n]\!]$.

(École normale supérieure)

▷ **Solution.**

Notons $\mathrm{G}_1 = \prod_{i=1}^{n} \mathbb{Z}/a_i\mathbb{Z}$ et $\mathrm{G}_2 = \prod_{i=1}^{m} \mathbb{Z}/b_i\mathbb{Z}$. Démontrons le résultat d'unicité par récurrence sur $n \geqslant 0$.

• Supposons $n = 0$. On a alors

$$\mathrm{G}_1 = \{0\} \simeq \prod_{i=1}^{m} \mathbb{Z}/b_i\mathbb{Z}.$$

Ceci implique $1 = \operatorname{Card} \mathrm{G}_1 = \operatorname{Card} \mathrm{G}_2 = b_1 \ldots b_m$. D'où $m = 0$.

• Supposons $n \geqslant 1$ et la propriété établie jusqu'au rang $n - 1$. Pour faire chuter n, nous allons considérer le groupe $a_1 \cdot \mathrm{G}_1$. Il est constitué des éléments de la forme

$$a_1 \cdot \mathrm{X} = a_1 \cdot (x_1, \ldots, x_n) = (a_1 \cdot x_1, \ldots, a_1 \cdot x_n),$$

où $\mathrm{X} = (x_1, \ldots, x_n)$ est un élément quelconque de G_1 et où $a_1 \cdot x$ désigne l'élément $x + \cdots + x$ avec a_1 termes. En fait $a_1 \cdot \mathrm{G}_1$ est l'image de G_1 par le morphisme de groupes $\mathrm{X} \longmapsto a_1 \cdot \mathrm{X}$. C'est donc bien un sous-groupe de G_1 et on a clairement,

$$a_1 \cdot \mathrm{G}_1 = (a_1 \cdot (\mathbb{Z}/a_1\mathbb{Z})) \times (a_1 \cdot (\mathbb{Z}/a_2\mathbb{Z})) \times \cdots \times (a_1 \cdot (\mathbb{Z}/a_n\mathbb{Z})) \text{ et}$$
$$a_1 \cdot \mathrm{G}_2 = (a_1 \cdot (\mathbb{Z}/b_1\mathbb{Z})) \times (a_1 \cdot (\mathbb{Z}/b_2\mathbb{Z})) \times \cdots \times (a_1 \cdot (\mathbb{Z}/b_m\mathbb{Z})).$$

Comme $G_1 \simeq G_2$, on a $a_1 \cdot G_1 \simeq a_1 \cdot G_2$. Regardons d'un peu plus près les groupes $a_1 \cdot (\mathbb{Z}/a_i\mathbb{Z})$ et $a_1 \cdot (\mathbb{Z}/b_j\mathbb{Z})$ qui interviennent ci-dessus. Pour cela, établissons un petit lemme.

Lemme. *Si n, p sont des entiers naturels non nuls, le groupe $p \cdot (\mathbb{Z}/n\mathbb{Z})$ est cyclique, isomorphe au groupe $\mathbb{Z}/\left(\dfrac{n}{\operatorname{pgcd}(p,n)}\right)\mathbb{Z}$.*

Démonstration. On note $\pi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ le morphisme surjectif de groupes qui à un entier $x \in \mathbb{Z}$ associe $\overline{x}$ sa classe modulo n. Pour tout x dans $\mathbb{Z}$, on a

$$p\pi(x) = p\overline{x} = \overline{px} = \pi(px).$$

Il en résulte que $p \cdot (\mathbb{Z}/n\mathbb{Z})$ est l'image par π du sous-groupe $p\mathbb{Z}$ de $\mathbb{Z}$. C'est donc le sous-groupe cyclique de $\mathbb{Z}/n\mathbb{Z}$ engendré par $\pi(p) = \overline{p}$. Or, l'ordre de $\overline{p}$ est le plus petit entier naturel non nul k tel que kp soit divisible par n, c'est-à-dire $\dfrac{n}{\operatorname{pgcd}(p,n)}$, ce qui démontre le lemme. $\diamondsuit$

Comme pour tout $i \in [\![1, n]\!]$, $a_1 | a_i$, on a $\operatorname{pgcd}(a_1, a_i) = a_1$ et, d'après le lemme, $a_1 \cdot (\mathbb{Z}/a_i\mathbb{Z}) \simeq \mathbb{Z}/\dfrac{a_i}{a_1}\mathbb{Z}$. Ainsi on obtient

$$a_1 \cdot G_1 \simeq \prod_{1 \leqslant i \leqslant n} \mathbb{Z}/\frac{a_i}{a_1}\mathbb{Z}.$$

Du lemme, on tire également que

$$a_1 \cdot G_2 \simeq \prod_{1 \leqslant i \leqslant m} \mathbb{Z}/\left(\frac{b_i}{\operatorname{pgcd}(a_1, b_i)}\right)\mathbb{Z}.$$

Comme $\operatorname{Card} G_1 = \operatorname{Card} G_2$ on a $a_1 \dots a_n = b_1 \dots b_m$. Comme $a_1 \cdot G_1$ et $a_1 \cdot G_2$ sont isomorphes, on a également

$$\prod_{1 \leqslant i \leqslant n} \frac{a_i}{a_1} = \prod_{1 \leqslant i \leqslant m} \frac{b_i}{\operatorname{pgcd}(a_1, b_i)},$$

ce qui donne

$$\frac{\prod\limits_{1 \leqslant i \leqslant n} a_i}{a_1^n} = \frac{\prod\limits_{1 \leqslant i \leqslant m} b_i}{\prod\limits_{1 \leqslant i \leqslant m} \operatorname{pgcd}(a_1, b_i)}.$$

On a donc $a_1^n = \prod\limits_{1 \leqslant i \leqslant m} \operatorname{pgcd}(a_1, b_i) \leqslant a_1^m$ car $\operatorname{pgcd}(a_1, b_i) \leqslant a_1$, ce qui implique $n \leqslant m$. Comme G_1 et G_2 jouent des rôles symétriques (si $n \geqslant 1$, alors nécessairement $m \geqslant 1$), on peut démontrer de même que $m \leqslant n$ et ainsi $n = m$. Les majorations des $\operatorname{pgcd}(a_1, b_i)$ par a_1 effectuées ci-dessus sont alors nécessairement des égalités. On a donc pour tout

$i \in [\![1,n]\!]$, $a_1 = \operatorname{pgcd}(a_1, b_i)$, *i.e.* $a_1 | b_i$. En particulier, $a_1 | b_1$. Toujours pour des raisons de symétrie, $b_1 | a_1$ et finalement $a_1 = b_1$.

Revenons alors à l'isomorphisme $a_1 \cdot \mathrm{G}_1 \simeq a_1 \cdot \mathrm{G}_2 = b_1 \cdot \mathrm{G}_2$. Il s'écrit

$$\prod_{1 \leqslant i \leqslant n} \mathbb{Z}/\frac{a_i}{a_1}\mathbb{Z} \simeq \prod_{1 \leqslant i \leqslant n} \mathbb{Z}/\frac{b_i}{a_1}\mathbb{Z}.$$

Si $a_i = a_1$ $(1 \leqslant i \leqslant n)$, alors $\mathbb{Z}/\frac{a_i}{a_1}\mathbb{Z} = \{0\}$. C'est vrai en particulier pour $i = 1$. Nous avons fait chuter n et allons pouvoir appliquer l'hypothèse de récurrence. Si tous les a_i sont égaux à a_1, $a_1 \cdot \mathrm{G}_1 = \{0\} = a_1 \cdot \mathrm{G}_2$ et tous les b_i sont égaux à a_1. L'unicité est alors prouvée.

Dans le cas contraire, on peut considérer r le plus petit entier de $[\![2,n]\!]$ tel que $a_r > a_1$. Nécessairement, il existe $i \in [\![1,n]\!]$ tel que $b_i > b_1 = a_1$. On note s le plus petit de ces entiers. On a $s \geqslant 2$. Dans ces conditions, l'isomorphisme devient

$$\mathbb{Z}/\frac{a_r}{a_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/\frac{a_n}{a_1}\mathbb{Z} \simeq \mathbb{Z}/\frac{b_s}{a_1}\mathbb{Z} \times \cdots \times \mathbb{Z}/\frac{b_n}{a_1}\mathbb{Z}.$$

Par hypothèse de récurrence, on a $r = s$ (et si $i < r$, $a_i = a_1 = b_i$) et

$$\frac{a_r}{a_1} = \frac{b_r}{b_1}, \ldots, \frac{a_n}{a_1} = \frac{b_n}{a_1}.$$

On conclut donc à l'égalité $a_i = b_i$ pour tout $1 \leqslant i \leqslant n$, ce qui termine la démonstration par récurrence. ◁

Le groupe symétrique d'un ensemble E, *noté* $\mathcal{S}_\mathrm{E}$, *est le groupe des permutations de* E. *On note* $\mathcal{S}_n$ *le groupe symétrique de* $[\![1,n]\!]$ *pour tout* $n \geqslant 1$. *L'étude des groupes de permutation finis se ramène à l'étude des* $\mathcal{S}_n$ *car si* E *est de cardinal* n, $\mathcal{S}_\mathrm{E}$ *est isomorphe à* $\mathcal{S}_n$.

On rappelle que toute permutation $\sigma \in \mathcal{S}_n$ *admet une décomposition en cycles à supports disjoints et que cette décomposition est unique à l'ordre des cycles près. Un cycle* c *de longueur* p *ou encore* p*-cycle, sera noté* $c = (a_1, a_2, \ldots, a_p)$ *où pour tout* i, $a_{i+1} = c(a_i)$ *(lire les indices modulo* p*). Une transposition est un cycle de longueur* 2. *Pour tout* $n \geqslant 2$, *le groupe* $\mathcal{S}_n$ *est engendré par les transpositions. En fait, les transpositions* $(1,i)$ *pour* $2 \leqslant i \leqslant n$ *suffisent, ce qui peut se voir par récurrence sur* n *ou simplement à partir du résultat précédent puisque pour* $i \neq j$, $(i,j) = (1,i)(1,j)(1,i)$. *L'exercice suivant s'intéresse au nombre minimal de transpositions permettant de générer* $\mathcal{S}_n$.

2.21. Génération du groupe symétrique

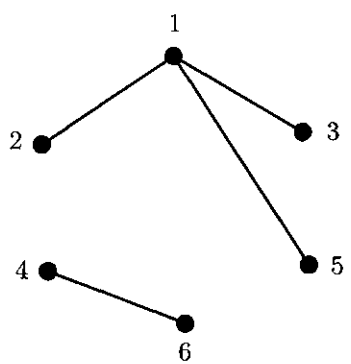
Soit $n \geqslant 2$. Quel est le nombre minimal de transpositions engendrant le groupe symétrique $\mathcal{S}_n$?

(École normale supérieure)

Solution.

On a rappelé plus haut que les transpositions $(1,2),\ldots,(1,n)$ engendrent $\mathcal{S}_n$. Ainsi, $n-1$ transpositions suffisent et intuitivement, il semble difficile de faire mieux, ce que semble confirmer l'étude des premières valeurs de n. En effet, pour $n=2$ il nous faut au moins une transposition (la seule !) et pour $n=3$, il nous en faut au moins 2 puisque une seule transposition engendre un sous-groupe à deux éléments.

Donnons-nous des transpositions $\tau_1,\ldots,\tau_k$, avec $k \leqslant n-2$. Pour visualiser le problème nous allons représenter les entiers de 1 à n par des points du plan. On fait en sorte que trois quelconques des points ne soient pas alignés. Si la transposition (a,b) apparaît dans la liste $\tau_1,\ldots,\tau_k$ on joint par un segment les points correspondant aux entiers a et b. On obtient ce qu'on appelle un graphe[2]. Par exemple, pour $n=6$, $k=4$, $\tau_1=(1,2)$, $\tau_2=(1,3)$, $\tau_3=(1,5)$ et $\tau_4=(4,6)$, on obtient le graphe suivant :



Il est assez clair qu'une condition nécessaire pour que $\tau_1,\ldots,\tau_k$ engendrent $\mathcal{S}_n$ est qu'on puisse passer de n'importe quel sommet $i \in [\![1,n]\!]$ à n'importe quel autre sommet $j \in [\![1,n]\!]$ en suivant les arêtes du graphe. Un graphe vérifiant cette propriété est dit *connexe*. Il est aisé de dessiner des graphes connexes sur un ensemble de cardinal n qui contiennent $n-1$ arêtes. On peut par exemple choisir un sommet et le relier à tous

[2] Formellement, un graphe G est un couple (S, A) où S est un ensemble fini, et A un ensemble de paires d'éléments de S. Les éléments de S sont appelés les sommets du graphe, ceux de A les arêtes.

les autres. Mais il semble impossible d'y arriver avec seulement $n-2$ arêtes. On a effectivement le résultat suivant :

Lemme. *Si un graphe* G *sur un ensemble* E *de cardinal* $n \geqslant 2$ *est connexe, alors il a au moins* $n-1$ *arêtes.*

Démonstration. On procède par récurrence sur n, le résultat étant clair pour $n=2$. Supposons le résultat prouvé au rang $n-1$. Soit G un graphe à n sommets. Si x est un sommet de G on note $\delta(x)$ sa *valence* c'est-à-dire le nombre d'arêtes qui arrivent en x. On a aisément

$$\sum_{x\in \mathrm{E}} \delta(x) = 2a(\mathrm{G}),$$

où $a(\mathrm{G})$ est le nombre d'arêtes de G (car une arête correspond à deux sommets). La connexité de G impose évidemment $\delta(x) \geqslant 1$ pour tout x. Deux cas se présentent :

- Pour tout x, $\delta(x) \geqslant 2$. Alors on a directement grâce à l'égalité ci-dessus $a(\mathrm{G}) \geqslant n$.
- Il existe un sommet x_0 de valence 1. Retirons ce sommet x_0 et l'unique arête qui y aboutit. On obtient un nouveau graphe G′ sur l'ensemble $\mathrm{E}' = \mathrm{E}\backslash\{x_0\}$ qui est encore connexe. Par hypothèse de récurrence, $a(\mathrm{G}') \geqslant |\mathrm{E}'| - 1 = n-2$ et donc $a(\mathrm{G}) = a(\mathrm{G}') + 1 \geqslant n-1$. ◁

2.22. Plongement de $\mathcal{S}_n$ dans $\mathcal{A}_{n+2}$

Soit $n \in \mathbb{N}^*$. Montrer qu'il existe un morphisme injectif de $\mathcal{S}_n$ dans $\mathcal{A}_{n+2}$.

(École normale supérieure)

▷ **Solution.**

On considère l'application $\psi : \mathcal{S}_n \to \mathcal{A}_{n+2}$ définie par $\psi(\sigma) = \sigma$ si σ est une permutation paire et $\psi(\sigma) = \sigma \circ (n+1, n+2)$ si σ est impaire. L'application ψ est injective par unicité de la décomposition en cycles à supports disjoints. Et il est aisé de constater que ψ est un morphisme de groupes. ◁

Le lecteur qui a un peu plus de connaissances en théorie des groupes (simplicité du groupe alterné $\mathcal{A}_n$ *pour* $n \geqslant 5$*) pourra démontrer qu'il est impossible d'injecter le groupe* $\mathcal{S}_n$ *dans* $\mathcal{A}_{n+1}$ *pour* $n \geqslant 2$.

Les deux exercices suivants montrent l'intérêt qu'il y a à connaître des parties génératrices d'un groupe. Si X *est une partie génératrice d'un groupe* G*, la connaissance de la restriction d'un morphisme de groupes*

$f : G \to G'$ à X caractérise f de manière unique. Cette idée permettra au lecteur de montrer facilement que les morphismes de groupes de $\mathbb{Z}/n\mathbb{Z}$ dans $\mathbb{Z}/n\mathbb{Z}$ sont les applications $x \longmapsto ax$: il suffit de regarder l'image du générateur $\overline{1}$. Dans le second tome algèbre on trouvera d'autres utilisations des parties génératrices dans l'étude du groupe linéaire.

2.23. Morphismes d'un sous-groupe de $\mathcal{S}_p$ dans $\mathbb{Z}/p\mathbb{Z}$

Soit p un nombre premier et G un sous-groupe du groupe symétrique $\mathcal{S}_p$. Déterminer le cardinal de l'ensemble des homomorphismes de G dans $\mathbb{Z}/p\mathbb{Z}$.

(École normale supérieure)

Solution.

Soit f un morphisme de $(G, \circ)$ dans $(\mathbb{Z}/p\mathbb{Z}, +)$. La composée de deux éléments σ_1 et σ_2 sera notée $\sigma_1\sigma_2$ et l'application identité 1.

On sait que tout élément σ de $\mathcal{S}_p$ peut se décomposer en produit (commutatif) de cycles à supports disjoints : on écrit $\sigma = c_1c_2\cdots c_k$ et l'on note n_i l'ordre du cycle c_i. On a donc $n_1 + \cdots + n_k \leqslant p$. L'ordre de σ est $n = \text{ppcm}(n_1, \ldots, n_k)$. On a $\sigma^n = 1$ donc $n.f(\sigma) = f(\sigma^n) = 0$. L'ordre de tout élément non nul de $\mathbb{Z}/p\mathbb{Z}$ est p car il divise l'ordre p du groupe. Donc si $f(\sigma) \neq 0$, p divise $n = \text{ppcm}(n_1, \ldots, n_k)$ et comme p est premier, il divise un des n_i. On a donc alors $n_i \geqslant p$ et comme $n_i \leqslant n_1 + \cdots + n_k \leqslant p$, $n_i = p$. Ainsi $k = 1$ et σ est un p-cycle. On a donc montré que si $\sigma \in \mathcal{S}_p$ et n'est pas un p-cycle, $f(\sigma) = 0$. Cela montre déjà que si G ne contient pas de p-cycle, $f = 0$.

Si G contient le p-cycle c, alors G contient le sous-groupe engendré par c, $gr(c) = \{c^k,\ k \in [\![0, p-1]\!]\}$. Examinons le cas où G contient de plus un élément σ qui n'appartient pas à $gr(c)$.

Supposons que σ est un p-cycle c' que l'on note $c' = (a_1, \ldots, a_p)$. Il existe $k \in [\![1, p-1]\!]$ tel que $c^k(a_1) = a_2$, par définition d'un p-cycle. On en déduit que $c^{-k}c'(a_1) = a_1$, donc $c^{-k}c'$ n'est pas un p-cycle donc $f(c^{-k}c') = -kf(c) + f(c') = 0$. Comme $c' \notin gr(c)$, $c^{-k}c' \neq 1$ et il existe $j \in [\![2, p]\!]$ tel que $c^{-k}c'(a_j) \neq a_j$; soit $i \in [\![1, p-1]\!]$ tel que $c^{-k}c'(a_j) = a_{i+j} = c'^i(a_j)$ (les indices sont notés modulo p). On a alors $c'^{-i}c^{-k}c'(a_j) = a_j$ et l'on montre que $-kf(c) + (1-i)f(c') = 0$ comme précédemment . On en déduit que $if(c') = 0$ et comme i est premier avec p, $f(c') = 0$, puis $-kf(c) = 0$ et comme k est premier avec p, $f(c) = 0$.

Si σ n'est pas un p-cycle, on le décompose en produit de cycles à supports disjoints $c'c_2 \ldots c_i$, où l'ordre de c' est supérieur ou égal à 2 (car σ n'est pas l'identité). Si $c' = (a_1a_2 \ldots a_j)$, on considère $k \in [\![1, p-1]\!]$ tel que $c^k(a_1) = a_2$. On a alors $c^{-k}c'(a_1) = a_1$. Comme a_1 n'est pas dans le

support de $c_2 \dots c_i$, on a également $c^{-k}\sigma(a_1) = a_1$. Ainsi $c^{-k}\sigma$ n'est pas un p-cycle donc $f(c^{-k}\sigma) = -kf(c) + f(\sigma) = 0$ et $f(c) = 0$ car $f(\sigma) = 0$.

Le seul cas où il existe un morphisme non nul de G dans $\mathbb{Z}/p\mathbb{Z}$ est celui où $G = gr(c)$ où c est un p-cycle. Mais alors G est un groupe cyclique d'ordre p. On peut choisir comme image de c un élément quelconque de $\mathbb{Z}/p\mathbb{Z}$. On obtient p morphismes de G dans $\mathbb{Z}/p\mathbb{Z}$.

Conclusion. Si G est un sous-groupe de $\mathcal{S}_p$ engendré par un p-cycle, il y a p morphismes de G dans $\mathbb{Z}/p\mathbb{Z}$. Dans tous les autres cas, il n'y en a qu'un : l'application nulle. ◁

2.24. Morphismes de $\mathcal{S}_4$ dans $\mathcal{S}_3$

Déterminer les morphismes du groupe $\mathcal{S}_4$ dans le groupe $\mathcal{S}_3$.
(École normale supérieure)

▷ **Solution.**

Dans la suite , nous noterons Id_4 et Id_3 les éléments neutres respectifs de $\mathcal{S}_4$ et $\mathcal{S}_3$. Il est bien connu que $\mathcal{S}_4$ est engendré par les transpositions $(1,2),(1,3),(1,4)$. On observera aussi que l'image par un morphisme de groupes f d'un élément x d'ordre n est un élément dont l'ordre est un diviseur de n puisque $(f(x))^n = f(x^n) = e$. Soit f un morphisme de groupes de $\mathcal{S}_4$ dans le groupe $\mathcal{S}_3$. Pour $2 \leqslant i \leqslant 4$, notons $\tau_i = f((1,i))$. Comme $(1,i)$ est d'ordre 2, τ_i est d'ordre 1 ou 2. Les seuls éléments de $\mathcal{S}_3$ d'ordre 2 étant les transpositions, τ_i est soit une transposition, soit l'identité. À partir de là, plusieurs cas sont possibles.

• Soit il existe $i \in [\![2,4]\!]$ tel que $\tau_i = \mathrm{Id}_3$. Pour $j \in [\![2,4]\!]$, $j \neq i$, on a $(1,j)(1,i) = (1,i,j)$. On en déduit : $f((1,i,j)) = f((1,j))f((1,i)) = \tau_j$. Mais $(1,i,j)$ est d'ordre 3, donc $\tau_j^3 = \mathrm{Id}_3$ et comme $\tau_j^2 = \mathrm{Id}_3$, on obtient que $\tau_j = \mathrm{Id}_3$. On a donc $f((1,j)) = \mathrm{Id}_3$, pour tout $j \in [\![2,4]\!]$. Ces transpositions engendrant $\mathcal{S}_4$, f est le morphisme constant :

$$s \in \mathcal{S}_4 \longmapsto \mathrm{Id}_3 \in \mathcal{S}_3.$$

Dans tous les autres cas, τ_2, τ_3, τ_4 seront des transpositions.

• Supposons qu'il existe une transposition τ de $\mathcal{S}_3$ telle que l'on ait $\tau_2 = \tau_3 = \tau_4 = \tau$. Toute permutation paire (resp. impaire) étant produit d'un nombre pair (resp. impair) d'éléments de $\{(1,2),(1,3),(1,4)\}$, on en déduit que :

$$\begin{cases} f(s) = \mathrm{Id}_3 & \text{si } s \text{ est une permutation paire de } \mathcal{S}_4 \\ f(s) = \tau & \text{si } s \text{ est une permutation impaire de } \mathcal{S}_4. \end{cases}$$

Réciproquement, si τ est une transposition quelconque de $\mathcal{S}_3$, toute application de $\mathcal{S}_4$ dans $\mathcal{S}_3$ ainsi définie est un morphisme de groupes qui s'iden-

lifie à la signature. On peut définir 3 tels morphismes, car $\mathcal{S}_3$ possède 3 transpositions.

• Supposons que parmi les trois transpositions τ_2, τ_3, τ_4, deux soient égales et la troisième distincte des deux autres. On peut donc déterminer (i, j, k) et (a, b, c) tels que $\{i, j, k\} = \{2, 3, 4\}$, $\{a, b, c\} = \{1, 2, 3\}$, $\tau_i = \tau_j = (a, b)$ et $\tau_k = (a, c)$. On obtient alors

$$f((1, i, j)) = f((1, j)(1, i)) = (a, b)(a, b) = \mathrm{Id}_3$$
$$f((1, i, k)) = f((1, k)(1, i)) = (a, c)(a, b) = (a, b, c).$$

On remarque que $(1, i, k)(1, i, j) = (1, k)(i, j)$, d'où l'on déduit que $f((1, k)(i, j)) = f((1, i, k))f((1, i, j)) = (a, b, c)$. Mais le produit de transpositions de supports disjoints $(1, k)(i, j)$ est d'ordre 2. On obtient ainsi,

$$(a, c, b) = (a, b, c)^2 = f(((1, k)(i, j))^2) = f(\mathrm{Id}_4) = \mathrm{Id}_3 .$$

C'est impossible. Ce troisième cas ne peut jamais être obtenu.

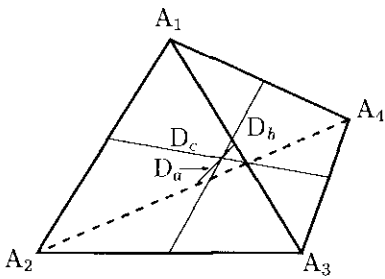
• Supposons enfin que τ_2, τ_3, τ_4 sont trois transpositions distinctes. On obtient à l'ordre près les trois transpositions de $\mathcal{S}_3$, $(1, 2), (2, 3), (1, 3)$. Il existe a, b, c tels que $\{a, b, c\} = \{1, 2, 3\}$ et $\tau_2 = (a, b), \tau_3 = (b, c)$ et $\tau_4 = (c, a)$. Étant donnés a, b, c, il existe au plus un morphisme f de $\mathcal{S}_4$ dans $\mathcal{S}_3$ ayant ces valeurs pour τ_2, τ_3 et τ_4, puisque $(1, 2), (1, 3)$ et $(1, 4)$ engendrent $\mathcal{S}_4$.

Pour montrer l'existence d'un tel morphisme f, nous en donnerons une réalisation géométrique. Il y a plusieurs manières d'obtenir $\mathcal{S}_4$ comme groupe d'isométries. On peut, par exemple l'identifier aux isométries laissant un tétraèdre régulier invariant ou aux isométries positives laissant un cube invariant. Dans chacun des cas, ces isométries induisent une action sur un ensemble à 3 éléments : perpendiculaires communes aux couples d'arêtes opposées pour le tétraèdre ; axes des faces pour le cube. On définit ainsi une application de $\mathcal{S}_4$ dans $\mathcal{S}_3$. Montrons cela de manière plus précise.

⋆ Soit $\mathcal{T}$ un tétraèdre régulier de sommets $\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{A}_4$ et $\mathrm{Is}(\mathcal{T})$ les isométries de l'espace laissant $\mathcal{T}$ invariant. L'action de $\mathrm{Is}(\mathcal{T})$ sur les sommets du tétraèdre définit un morphisme de groupes $\varphi : \mathrm{Is}(\mathcal{T}) \longmapsto \mathcal{S}_4$. La réflexion par rapport au plan médiateur de $[\mathrm{A}_i, \mathrm{A}_j]$ $(1 \leqslant i < j \leqslant 4)$ a pour image par φ la transposition de A_i et A_j dans l'ensemble des sommets. Les transpositions étant dans $\mathrm{Im}\,\varphi$ et engendrant $\mathcal{S}_4$, φ est surjective. Comme $(\mathrm{A}_1, \mathrm{A}_2, \mathrm{A}_3, \mathrm{A}_4)$ est un repère affine de l'espace, la seule application laissant fixes les sommets du tétraèdre $\mathcal{T}$ est l'identité. Le noyau de φ est donc réduit à l'identité ; ainsi φ est injective et finalement φ est un isomorphisme de $\mathrm{Is}(\mathcal{T})$ sur $\mathcal{S}_4$.

Dans un tétraèdre régulier, la perpendiculaire commune à deux arêtes opposées est le segment qui joint les milieux de ces arêtes. L'ensemble des

perpendiculaires communes aux trois couples d'arêtes opposées sera noté $\mathcal{D} = \{D_a, D_b, D_c\}$. Elles sont numérotées de telle façon que D_a (resp. D_b, resp. D_c) soit la perpendiculaire commune à $[A_1, A_3]$ et $[A_2, A_4]$ (resp. $[A_1, A_4]$ et $[A_2, A_3]$, resp. $[A_1, A_2]$ et $[A_3, A_4]$).

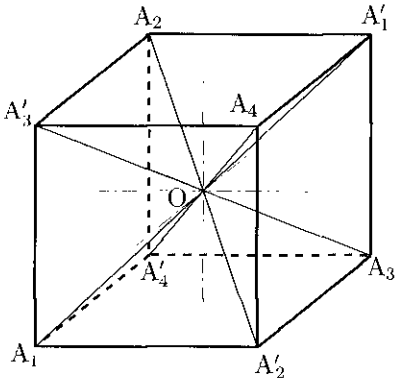


Soit $u \in \mathcal{S}_4$, $s = \varphi^{-1}(u)$ l'isométrie de $\mathcal{T}$ associée. L'isométrie s transforme un couple d'arêtes opposées en un autre couple d'arêtes opposées, ceci de manière injective. Elle transforme donc un élément de $\mathcal{D}$ en un élément de $\mathcal{D}$, de manière injective. Elle induit donc une permutation u' de $\mathcal{D}$ qui s'identifie à un élément de $\mathcal{S}_3$. Soit f l'application qui à $u \in \mathcal{S}_4$ associe $u' \in \mathcal{S}_3$. Par construction, l'application qui à s associe u' est un morphisme (pour la loi $\circ$). Puisque φ^{-1} est aussi un morphisme, on en déduit que f est un morphisme de groupes de $\mathcal{S}_4$ dans $\mathcal{S}_3$.

Pour $i \in \{2,3,4\}$, posons $\{j,k\} = \{2,3,4\} \setminus \{i\}$. Alors $\varphi^{-1}((1,i))$ est la réflexion par rapport au plan médiateur de $[A_1, A_i]$. Cette réflexion échange A_1 et A_i et laisse invariants A_j et A_k ; elle laisse invariants les milieux de $[A_1, A_i]$ et $[A_j, A_k]$ et échange les milieux de $[A_1, A_j]$ et $[A_i, A_k]$ et les milieux de $[A_i, A_j]$ et $[A_i, A_k]$. Elle laisse donc invariante la perpendiculaire commune à $[A_1, A_i]$ et $[A_j, A_k]$ et échange les deux autres. Ces remarques permettent de déterminer les images par $\varphi^{-1}((1,i))$ des éléments de $\mathcal{D}$ et donc $\tau_i = f((1,i))$. On trouve $\tau_2 = (a,b), \tau_3 = (b,c)$ et $\tau_4 = (c,a)$.

Ceci démontre l'existence d'un morphisme de groupes f de $\mathcal{S}_4$ dans $\mathcal{S}_3$ ayant les propriétés voulues. Puisque (a,b,c) étant une permutation quelconque de $(2,3,4)$, on trouve 6 morphismes de ce type.

⋆ Venons-en au cas du cube. Soit $\mathcal{C}$ un cube de centre O, de sommets A_1, A_2, A_3, A_4, A'_1, A'_2, A'_3, A'_4, les sommets étant numérotés de telle façon que A'_i soit le symétrique de A_i par rapport à O et que $A_1A'_2A_3A'_4$ soit un carré. Notons $\mathrm{Is}^+(\mathcal{C})$ l'ensemble des isométries positives laissant $\mathcal{C}$ invariant. Les éléments de $\mathrm{Is}^+(\mathcal{C})$ sont des rotations dont l'axe contient O. Soit Δ l'ensemble des diagonales du cube, c'est-à-dire l'ensemble des quatre droites $\delta_i = (A_iA'_i)$ $(1 \leqslant i \leqslant 4)$.



Tout élément de $\mathrm{Is}^+(\mathcal{C})$ transforme une diagonale du cube en une autre diagonale. Étant bijectif, il réalise une permutation des éléments de Δ. On obtient donc un morphisme de groupes $\Psi : \mathrm{Is}^+(\mathcal{C}) \longrightarrow \mathcal{S}_4$.

Si $s \in \mathrm{Ker}\,\Psi$, s laisse les quatre droites de Δ invariantes. La restriction de s à chaque droite de Δ est l'identité ou la symétrie par rapport à O. L'application vectorielle associée à s est donc $\pm \mathrm{Id}$. La symétrie par rapport au point O n'étant pas une isométrie positive, $s = \mathrm{Id}$. L'application Ψ est donc injective.

Soit δ_{12} la droite passant par O qui joint les milieux des arêtes $[\mathrm{A}_1, \mathrm{A}'_2]$ et $[\mathrm{A}'_1, \mathrm{A}_2]$ et s_{12} le retournement d'axe δ_{12}. Il est immédiat que s_{12} échange les diagonales δ_1 et δ_2. La droite δ_{12} est orthogonale au plan P contenant A_3, A_4, A'_3 et A'_4. La restriction de s_{12} à P est donc la symétrie de centre O. L'application s_{12} laisse donc invariantes les diagonales δ_3 et δ_4. L'image de s_{12} par Ψ est la transposition des diagonales δ_1 et δ_2. On montre de même que $\Psi(\mathrm{Is}^+)$ contient toutes les transpositions et on conclut que ψ est surjective. Finalement Ψ est un isomorphisme de Is^+ sur $\mathcal{S}_4$.

Considérons maintenant les trois axes des faces, c'est-à-dire les trois droites passant par O et orthogonales à deux des faces du cube. On notera $\mathcal{A} = \{d_a, d_b, d_c\}$ l'ensemble de ces axes. Ils sont numérotés de telle façon que d_a (resp. d_b, resp. d_3) soient respectivement les axes des faces $\mathrm{A}_1\mathrm{A}'_2\mathrm{A}_3\mathrm{A}'_4$, $\mathrm{A}_1\mathrm{A}'_2\mathrm{A}_4\mathrm{A}'_3$ et $\mathrm{A}_1\mathrm{A}'_3\mathrm{A}_2\mathrm{A}'_4$.

Tout élément de $\mathrm{Is}^+(\mathcal{C})$ transformant une face du cube en une autre face et laissant O invariant, transforme un élément de $\mathcal{A}$ en un élément de $\mathcal{A}$. Elle opère donc une permutation de $\mathcal{A}$. En associant à tout élément u de $\mathcal{S}_4$ l'isométrie $\Psi^{-1}(u)$, puis à celle-ci une permutation de $\mathcal{A}$, identifiée à un élément de $\mathcal{S}_3$, on obtient comme précédemment pour le tétraèdre un morphisme f de $\mathcal{S}_4$ dans $\mathcal{S}_3$. Sachant qu'avec les notations précédentes, on a $\Psi(s_{1i}) = (1, i)$, on peut déterminer $\tau_i = f((1, i))$, pour $1 \leqslant i \leqslant 3$. On trouve de nouveau $\tau_2 = (a, b)$, $\tau_3 = (b, c)$ et $\tau_4 = (c, a)$ et on conclut de la même façon.

Conclusion. Au total, on trouve 10 morphismes de groupes de $\mathcal{S}_4$ dans $\mathcal{S}_3$. Le premier est constant, les trois suivants ont pour images les trois groupes de cardinal 2 de $\mathcal{S}_3$, les 6 autres sont surjectifs. $\triangleleft$

L'ensemble $\mathrm{Aut}(\mathrm{G})$ *des automorphismes d'un groupe* G, *muni de la composition est un groupe. Les conjugaisons* $f_a : x \longmapsto axa^{-1}$, *pour* $a \in \mathrm{G}$, *sont des automorphismes remarquables appelés automorphismes intérieurs de* G *(les autres seront dits extérieurs). L'ensemble des automorphismes intérieurs forme un sous-groupe de* $\mathrm{Aut}(\mathrm{G})$. *Ce sous-groupe est bien entendu trivial lorsque* G *est abélien. La connaissance de* $\mathrm{Aut}(\mathrm{G})$ *est importante en théorie des groupes pour les questions d'extension. Le lecteur pourra montrer à titre d'exercice que pour tout* $n \geqslant 2$, $\mathrm{Aut}(\mathbb{Z}/n\mathbb{Z})$ *est isomorphe au groupe des inversibles de l'anneau* $\mathbb{Z}/n\mathbb{Z}$. *L'exercice suivant étudie les automorphismes du groupe symétrique.*

2.25. Automorphismes de $\mathcal{S}_n$

Soit $n \in \mathbb{N}^*$, $n \neq 6$. Montrer que les automorphismes de $\mathcal{S}_n$ sont intérieurs *i.e.* de la forme $s \longmapsto \sigma \circ s \circ \sigma^{-1}$ où $\sigma \in \mathcal{S}_n$. On s'intéressera aux images des transpositions.

(**École normale supérieure**)

$\triangleright$ **Solution.**

• Soit φ un automorphisme du groupe symétrique. Nous savons que l'ensemble T des transpositions forme une partie génératrice de $\mathcal{S}_n$. L'automorphisme φ sera donc uniquement déterminé si l'on connaît les images par φ des transpositions. Pour commencer, examinons l'image d'une transposition et même l'image d'un élément quelconque de $\mathcal{S}_n$ par un automorphisme intérieur. Soit $\sigma \in \mathcal{S}_n$ et $(a_1, \ldots, a_k)$ un cycle de longueur k (on dira k-cycle). On a alors

$$\boxed{\sigma \circ (a_1, \ldots, a_k) \circ \sigma^{-1} = (\sigma(a_1), \ldots, \sigma(a_k))}. \quad (*)$$

Le lecteur vérifiera cette égalité sans grande difficulté. Un k-cycle est donc transformé en un k-cycle par tout automorphisme intérieur. En particulier une transposition est transformée en une transposition.

Plus généralement, si s est un élément de $\mathcal{S}_n$ et $s_1, \ldots, s_k$ les cycles apparaissant dans la décomposition de s en produit de cycles à support disjoint ($s = s_1 s_2 \ldots s_k$), on obtient

$$\boxed{\sigma s \sigma^{-1} = (\sigma s_1 \sigma^{-1})(\sigma s_2 \sigma^{-1}) \ldots (\sigma s_k \sigma^{-1})}. \quad (**)$$

Pour tout i, $\sigma s_i \sigma^{-1}$ est un cycle de même longueur que s_i d'après la formule $(*)$. Si $\{a_1, \ldots, a_p\}$ est le support de s_i, $\{\sigma(a_1), \ldots, \sigma(a_p)\}$ est le

support de $\sigma s_i \sigma^{-1}$. En particulier les $\sigma s_i \sigma^{-1}$ sont des cycles à supports disjoints.

Nous allons démontrer, pour commencer, que φ transforme toute transposition en une transposition. Un automorphisme de $\mathcal{S}_n$ conserve les propriétés algébriques des éléments de $\mathcal{S}_n$. Si τ est une transposition, τ est d'ordre 2, donc $\varphi(\tau)$ est aussi d'ordre 2. Cependant, cela ne suffit pas pour affirmer que $\varphi(\tau)$ est une transposition puisqu'il existe d'autres éléments d'ordre 2, les produits de transpositions à support disjoints (il n'y a qu'elles, puisque l'ordre d'une permutation est le ppcm des longueurs des cycles intervenant dans sa décomposition en produit de cycles à supports disjoints). Notons T_k l'ensemble des permutations de $\mathcal{S}_n$ qui sont produit d'exactement k transpositions à supports disjoints (pour $2k \leqslant n$). On a en particulier $\mathrm{T}_1 = \mathrm{T}$. Nous proposons deux démonstrations différentes du fait qu'un élément de T_1 est transformé par φ en un élément de T_1.

$\star$ D'après la relation (**), chaque ensemble T_k est une classe de conjugaison de $\mathcal{S}_n$. L'image par φ d'une classe de conjugaison est encore une classe de conjugaison (car $\varphi(\sigma \circ \tau \circ \sigma^{-1}) = \varphi(\sigma) \circ \varphi(\tau) \circ \varphi(\sigma)^{-1}$). L'ensemble $\varphi(\mathrm{T}_1)$ est donc l'un des T_k.

On va montrer par un argument de cardinalité que, pour $n \neq 6$, on ne peut pas avoir $\varphi(\mathrm{T}_1) = \mathrm{T}_k$ avec $k \neq 1$. Pour cela, on va calculer le cardinal de T_k. On a $|\mathrm{T}_1| = \mathrm{C}_n^2 = \dfrac{n(n-1)}{2}$, car une transposition est déterminée par son support. Pour $k \geqslant 2$, on obtient $|\mathrm{T}_k| = \dfrac{\mathrm{C}_n^2 \mathrm{C}_{n-2}^2 \dots \mathrm{C}_{n-2k+2}^2}{k!}$: on choisit les supports des k transpositions et on divise par $k!$ car l'ordre n'importe pas. En simplifiant, il vient

$$|\mathrm{T}_k| = \frac{n(n-1)\dots(n-2k+1)}{2^k k!}.$$

Nous allons montrer que si $n \neq 6$, l'équation $|\mathrm{T}_k| = |\mathrm{T}_1|$, c'est-à-dire

$$k! 2^{k-1} = (n-2)(n-3)\dots(n-2k+1),$$

n'a pas de solution k strictement supérieure à 1 (avec $2k \leqslant n$).

En effet, pour $k = 2$, on obtient $(n-2)(n-3) = 4$, équation qui n'a pas de solution dans $\mathbb{N}$. Pour $k \geqslant 3$, il vient $\mathrm{C}_{n-k}^k (n-2)\dots(n-k+1) = 2^{k-1}$, ce qui ne peut arriver que si $n - k + 1 = n - 2$ (sinon le terme de gauche a un facteur impair), soit $k = 3$. Mais on a alors $(n-2)\mathrm{C}_{n-3}^3 = 4$ qui entraîne $n = 6$, cas qui est justement écarté. On a donc $\varphi(\mathrm{T}_1) = \mathrm{T}_1$

$\star$ Pour montrer qu'une transposition est transformée en une

transposition, on peut[3] s'intéresser au centralisateur d'une transposition τ, qui est transformé en celui de $\varphi(\tau)$ par l'automorphisme φ.

Soit s un élément de T_k qu'on écrit $s = \tau_1\tau_2\dots\tau_k$, où $\tau_1,\dots,\tau_k$ sont des transpositions à supports disjoints. La permutation σ est dans le centralisateur de s si elle vérifie $\sigma s\sigma^{-1} = s$. Nous avons vu que $\sigma s\sigma^{-1} = (\sigma\tau_1\sigma^{-1})\dots(\sigma s_k\sigma^{-1})$, où les $\sigma\tau_i\sigma^{-1}$ sont des transpositions à supports disjoints. Par unicité de la décomposition en produit de cycles à supports disjoints, on en déduit que les $\sigma\tau_i\sigma^{-1}$ doivent s'obtenir par permutation des τ_i. Il y a $k!$ possibilités pour σ de permuter les τ_i. Supposons une telle permutation des transpositions fixée, et par exemple qu'une transposition (a,b) est envoyé sur (a',b'). On doit avoir

$$(a',b') = \sigma(a,b)\sigma^{-1} = (\sigma(a),\sigma(b)),$$

c'est-à-dire $\{\sigma(a),\sigma(b)\} = \{a',b'\}$. On a donc 2 possibilités différentes pour définir σ sur l'ensemble $\{a,b\}$ sachant que cet ensemble est envoyé sur $\{a',b'\}$. Il reste ensuite à déterminer σ sur les éléments qui n'appartiennent pas au support d'une des transpositions τ_i. Pour ces $n-2k$ éléments, il y a $(n-2k)!$ choix possibles. Au total, il y a $2^k k!(n-2k)!$ permutations qui commutent avec s.

Soit maintenant une transposition τ. Son centralisateur est de cardinal $2(n-2)!$. Si $\varphi(\tau)$ appartient à T_k, son centralisateur possède $(n-2k)!k!2^k$ éléments. Nous avons donc $2(n-2)! = (n-2k)!k!2^k$, c'est-à-dire

$$k!2^{k-1} = (n-2)(n-3)\dots(n-2k+1).$$

Nous obtenons la même équation que dans la méthode précédente et la conclusion est la même.

Le fait que les deux méthodes conduisent à la même équation n'a rien de mystérieux. En effet le cardinal de la classe de conjugaison d'un élément est égal à l'indice de son centralisateur dans $\mathcal{S}_n$. Il revient donc au même de dire que deux éléments de $\mathcal{S}_n$ ont des classes de conjugaison de même cardinal et des centralisateurs de même cardinal (cf. exercice 2.11)

• On a donc $\varphi(\mathrm{T}) = \mathrm{T}$. On va s'intéresser aux images des transpositions $(1,k)$ pour $k \in [\![2,n]\!]$. Celles-ci suffisent à générer $\mathcal{S}_n$. Il existe a_1 et a_2 distincts tels que $\varphi((1,2)) = (a_1,a_2)$. De même, il existe a et b distincts tels que $\varphi((1,3)) = (a,b)$. Comme $(1,2)$ et $(1,3)$ ne commutent pas, (a_1,a_2) et (a,b) non plus, ce qui nécessite $\{a_1,a_2\}\cap\{a,b\} \neq \emptyset$. On peut supposer que $a = a_1$ et on note $a_3 = b$. On a donc $\varphi((1,3)) = (a_1,a_3)$. L'injectivité de φ assure que $a_2 \neq a_3$.

3. Comme le fait Daniel Perrin dans son *Cours d'algèbre*, Ellipses, 1996, p. 30-33. On y trouvera une étude précise des centralisateurs des éléments d'ordre 2 qui explique mieux en quoi le cas $n = 6$ se distingue.

Montrons par récurrence sur $i \in [\![2, n]\!]$ que $\varphi((1,i))$ s'écrit (a_1, a_i) où a_i est distinct des a_k pour $k < i$. C'est vrai pour $i = 2$ et 3. Supposons $i > 3$. Alors $\varphi((1,i))$ est une transposition dont le support rencontre celui de $\varphi((1,k)) = (a_1, a_k)$ pour $2 \leqslant k \leqslant i-1$ (par hypothèse de récurrence), car $(1,i)$ et $(1,k)$ ne commutent pas (la démonstration est la même que pour $i = 3$). Si a_1 n'était pas dans le support de $\varphi((1,i))$, les a_k pour $k \in [\![2, i-1]\!]$ y seraient. Comme ils sont au nombre de $i - 2 \geqslant 2$, cela ne peut se produire que si $i = 4$. Dans ces conditions $\varphi((1,4)) = (a_2, a_3) = (a_1, a_3)(a_2, a_1)(a_1, a_3) = \varphi((3,1)(1,2)(1,3))$ et par injectivité, on a $(1,4) = (3,1)(1,2)(1,3)$, ce qui est faux.

Donc a_1 est dans le support de $\varphi((1,i))$. Cette dernière s'écrit (a_1, a_i) et par injectivité de φ, a_i est bien distincts des a_k pour $k < i$.

• On se retrouve donc avec n éléments de $[\![1, n]\!]$, $a_1, \ldots, a_n$ deux à deux distincts tels que $\varphi((1,i)) = (a_1, a_i)$ pour tout $2 \leqslant i \leqslant n$. Si on considère la permutation σ qui à $i \in [\![1, n]\!]$ associe a_i, on a d'après la remarque préliminaire

$$\sigma \circ (1,i) \circ \sigma^{-1} = (a_1, a_i).$$

Autrement dit, φ coïncide sur l'ensemble des transpositions $(1,i)$ (où $2 \leqslant i \leqslant n$), avec l'automorphisme intérieur $\varphi_\sigma : s \longmapsto \sigma \circ s \circ \sigma^{-1}$. Comme φ est un morphisme et que les $(1,i)$ forment un système générateur de $\mathcal{S}_n$, on a $\varphi = \varphi_\sigma$.

Conclusion. Pour $n \neq 6$, tout automorphisme de $\mathcal{S}_n$ est intérieur. ◁

Dans le cas exceptionnel $n = 6$, on peut montrer que $|\operatorname{Aut}(\mathcal{S}_6)| = 1440$, mais qu'il n'y a que 720 automorphismes intérieurs[4].

Nous terminons ce chapitre avec un exercice concernant la structure de groupe ordonné. Un groupe abélien $(\mathrm{G}, +)$ est dit ordonné s'il est muni d'un ordre total $\leqslant$ compatible avec la loi $+$, c'est-à-dire vérifiant

$$\forall (x,y,z) \in \mathrm{G}^3, \quad x \leqslant y \implies x + z \leqslant y + z.$$

Dans ce cas, on montre facilement que l'ensemble $\mathrm{P} = \{x \in \mathrm{G},\ 0 \leqslant x\}$ des éléments positifs de G possèdent les trois propriétés suivantes :

$$\mathrm{P} \cup (-\mathrm{P}) = \mathrm{G}, \quad \mathrm{P} \cap (-\mathrm{P}) = \{0\} \quad \textit{et} \quad \mathrm{P} + \mathrm{P} \subset \mathrm{P}.$$

Réciproquement, si P est une partie de G qui vérifient ces trois propriétés, alors la relation binaire $\leqslant$ définie par $x \leqslant y \iff y - x \in \mathrm{P}$ est une relation d'ordre total qui est compatible avec la loi $+$. Dans l'exercice suivant il s'agit de déterminer toutes les structures de groupe ordonné sur $\mathbb{Z}^2$.

4. On pourra se reporter à tout bon traité de théorie des groupes, par exemple le livre de Daniel Perrin déjà mentionné ou ROTMAN (J.), *An Introduction to the Theory of Groups*, GTM 148, 4e éd., Springer, 1995. p. 160-162.

2.26. Structure de groupe ordonné sur $\mathbb{Z}^2$

Si P est une partie non vide de $\mathbb{Z}^2$ on note

$$-\mathrm{P} = \{-p,\ p \in \mathrm{P}\} \quad \text{et} \quad \mathrm{P} + \mathrm{P} = \{p+q,\ (p,q) \in \mathrm{P}^2\}.$$

Trouver les parties P de $\mathbb{Z}^2$ vérifiant les points suivants :

(*i*) $\mathrm{P} \cup (-\mathrm{P}) = \mathbb{Z}^2$;

(*ii*) $\mathrm{P} \cap (-\mathrm{P}) = \{0\}$;

(*iii*) $\mathrm{P} + \mathrm{P} \subset \mathrm{P}$.

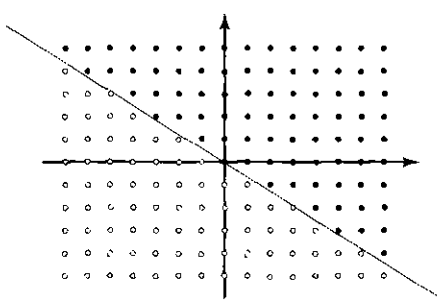
(École normale supérieure)

▷ **Solution.**

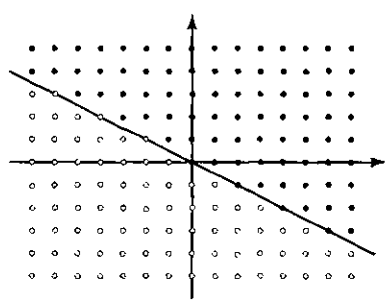
Il est clair que l'intersection de $\mathbb{Z}^2$ avec un demi-plan fermé de $\mathbb{R}^2$ dont la frontière contient $(0,0)$ convient si cette frontière ne contient pas d'autre point de $\mathbb{Z}^2$ que l'origine. On obtient donc une solution au problème posé en prenant pour P l'ensemble

$$\mathrm{P}_{a,b} = \{(x,y) \in \mathbb{Z}^2,\ ax + by \geqslant 0\},$$

avec $\dfrac{b}{a}$ irrationnel.



$\mathrm{P}_{a,b}$ avec b/a irrationnel



$\mathrm{P}^2_{1,2}$

Si la droite $ax + by = 0$ contient des points de $\mathbb{Z}^2$, donc si $\dfrac{b}{a}$ est rationnel, ou si $a = 0$, ces points sont tous les multiples entiers d'un même point (p,q), et on obtient deux ensembles P qui conviennent en ajoutant à l'ensemble

$$\{(x,y) \in \mathbb{Z}^2,\ ax + by > 0\}$$

soit les multiples entiers positifs, soit les multiples entiers négatifs de (p,q).

Il y a deux choix possibles pour (p,q). On choisira celui pour lequel la base $((a,b),(p,q))$ est orientée positivement. Les ensembles P ci-dessus sont donc

$$\mathrm{P}^1_{a,b} = \{(x,y) \in \mathbb{Z}^2,\ ax+by>0\} \cup \mathbb{N}(p,q),$$
$$\mathrm{P}^2_{a,b} = \{(x,y) \in \mathbb{Z}^2,\ ax+by>0\} \cup \mathbb{N}(-p,-q)$$

Soit maintenant une partie P de $\mathbb{Z}^2$ vérifiant (i), (ii), (iii). Montrons que P est de l'une des formes ci-dessus. Pour trouver une droite pouvant servir de «frontière» à P, il est naturel de considérer

$$\mathrm{P}^+ = \{(x,y) \in \mathrm{P},\ x>0\} \quad \text{et} \quad m = \sup\left\{\frac{y}{x},\ (x,y) \in \mathrm{P}^+\right\}.$$

Afin d'éviter d'avoir à distinguer plusieurs cas de figure, on suppose que $(1,0)$ et $(0,-1)$ sont dans P. On peut toujours se ramener à ce cas, d'abord parce que soit $(1,0)$, soit son opposé est dans P, et de même pour $(0,-1)$, ensuite parce que les symétries par rapport à l'axe des x et par rapport à l'axe des y conservent les propriétés (i), (ii), (iii), ainsi que celle d'être de la forme $\mathrm{P}_{a,b}$ ou $\mathrm{P}^i_{a,b}$ $(i=1$ ou $2)$.

La solution repose sur deux observations.

• Soit $(x,y) \in \mathbb{Z}^2$ et k un entier >0. Alors $(x,y) \in \mathrm{P}$ si et seulement si $(kx,ky) \in \mathrm{P}$.

Le sens direct résulte de (ii). D'autre part, pour $(x,y) \neq 0$, on a

$$(x,y) \notin \mathrm{P} \Rightarrow (-x,-y) \in \mathrm{P} \Rightarrow (-kx,-ky) \in \mathrm{P} \Rightarrow (kx,ky) \notin \mathrm{P}.$$

• Soit $(x_0,y_0) \in \mathbb{Z}^2$, $(x,y) \in \mathbb{Z}^2$. Si $x_0>0$, $(x,y) \in \mathrm{P}^+$ et si

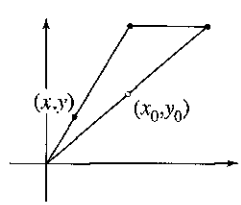
$$\frac{y_0}{x_0} < \frac{y}{x},$$

alors $(x_0,y_0) \in \mathrm{P}^+$.

En effet, tout point entier (x_0,y_0) tel que $y_0<0$ est dans P, puisque c'est une combinaison linéaire à coefficients positifs de $(1,0)$ et $(0,-1)$. On peut donc supposer $y_0>0$ et $y>0$. Il suffit alors de trouver des entiers k, ℓ, $n>0$ tels que

$$k(x_0,y_0) = \ell(x,y) + n(1,0).$$

On peut prendre $k=y$, $\ell=y_0$, $n=yx_0-y_0x$ (on vérifie qu'on a bien $n>0$).



Premier cas : m est fini. Soit $(x_0, y_0) \in \mathbb{Z}^2$ tel que $y_0 < mx_0$.

Si $x_0 > 0$, il existe, par définition de m, un point $(x, y) \in \mathrm{P}^+$ tel que $\dfrac{y_0}{x_0} < \dfrac{y}{x}$ et, d'après ce qui précède, $(x_0, y_0) \in \mathrm{P}$.

Si $x_0 < 0$, on a $\dfrac{-y_0}{-x_0} > m$, le point $(-x_0, -y_0)$ n'est pas dans P, et donc $(x_0, y_0) \in \mathrm{P}$.

Si $x_0 = 0$, alors (x_0, y_0) est un multiple entier positif de $(0, -1)$, et donc $(x_0, y_0) \in \mathrm{P}$.

Autrement dit, P contient $\mathrm{P}_{m,-1}$. Si m est irrationnel, l'ensemble P ne contient aucun des points entiers tels que $y > mx$, parce que leurs opposés sont dans P, et $\mathrm{P} = \mathrm{P}_{m,-1}$.

Si m est rationnel, disons $m = \dfrac{q}{p}$ (fraction irréductible, $p > 0$), le réseau $\mathbb{Z}^2$ se partitionne en trois parties : $\mathrm{P}_{m,-1}$, $-\mathrm{P}_{m,-1}$, et l'ensemble des points entiers situés sur la droite $y = mx$, qui est aussi l'ensemble des multiples entiers de (p, q). L'ensemble P contient soit (p, q) (et tous ses multiples positifs), soit $(-p, -q)$ (et tous ses multiples positifs), il est donc de la forme $\mathrm{P}^i_{m,-1}$, $i = 1$ ou 2.

Deuxième cas : $m = +\infty$. Pour tout point $(x_0, y_0) \in \mathbb{Z}^2$ tel que $x_0 > 0$, il existe $(x, y) \in \mathrm{P}^+$ tel que $\dfrac{y_0}{x_0} < \dfrac{y}{x}$, et, par, conséquent, $(x_0, y_0) \in \mathrm{P}$. On en déduit que P est formé des points entiers (x, y) tels que $x > 0$ ou $x = 0$ et $y \geqslant 0$. Autrement dit, $\mathrm{P} = \mathrm{P}^2_{1,0}$. ◁

On a ainsi montré qu'une structure d'ordre sur $\mathbb{Z}^2$ (voir les commentaires précédant l'exercice) peut être définie de deux façons :

- soit on se donne une forme linéaire $f(x, y) = ax + by$ à coefficients réels tels que $\dfrac{b}{a}$ soit irrationnel, et les éléments $\geqslant 0$ de $\mathbb{Z}^2$ sont les éléments (x, y) tels que $f(x, y) \geqslant 0$;

- soit on se donne d'abord une forme linéaire $f(x, y) = ax + by$ à coefficients réels tels que $\dfrac{b}{a}$ soit rationnel ou $a = 0$, et on doit se donner une seconde forme linéaire g indépendante de f (on peut par exemple prendre $g(x, y) = -bx + ay$ ou $g(x, y) = bx - ay$; les autres choix ne donnent rien de différent). Les éléments $\geqslant 0$ de $\mathbb{Z}^2$ sont alors les éléments (x, y) tels que $f(x, y) > 0$ ou $f(x, y) = 0$, $g(x, y) \geqslant 0$. On retrouve l'ordre lexicographique sur $\mathbb{Z}^2$ comme cas particulier.

Chapitre 3

Anneaux et corps

À l'origine de la théorie des anneaux, se trouvent des recherches de théorie des nombres. Vers 1831, Gauss a été amené, dans ses travaux sur les résidus biquadratiques et sur les fonctions elliptiques, à étudier les propriétés de divisibilité dans l'anneau $\mathbb{Z}[i]$. *En 1844, Eisenstein utilisa* $\mathbb{Z}[j]$ *(où* $j^3 = 1$*). Dans les anneaux* $\mathbb{Z}[i]$ *et* $\mathbb{Z}[j]$, *l'arithmétique est semblable à celle de* $\mathbb{Z}$ *; en particulier, ce sont des anneaux factoriels (tout élément de l'anneau se décompose de manière unique – à un élément inversible près – en produits d'irréductibles). Les recherches sur le grand théorème de Fermat, en particulier la résolubilité de l'équation* $x^p + y^p = z^p$, *avec* p *premier impair, amènent à l'étude de l'anneau* $\mathbb{Z}[\zeta]$, *où* ζ *est une racine primitive* p*-ième de l'unité. Celui-ci se révéla avoir pour certaines valeurs de* p, *et contrairement à ce qu'on avait cru initialement (jusqu'en 1840), des propriétés arithmétiques totalement différentes de celles de* $\mathbb{Z}$.

À partir de 1847, Kummer grâce à l'introduction des concepts de nombres idéaux et de classes d'idéaux, élucide le problème de la divisibilité dans $\mathbb{Z}[\zeta]$ *et démontre la conjecture de Fermat dans de nombreux cas. En 1871, Dedekind, généralisant les travaux de Kummer, définit, dans le cadre de la théorie des entiers algébriques, la notions d'*idéal *et démontre qu'un idéal peut s'écrire comme produit d'*idéaux premiers. *Le terme d'*anneau *ne sera introduit que par Hilbert en 1897 : à partir du début du* XX^e *siècle, il a fini par désigner tout ensemble sur lequel sont définies une loi de groupe additif et une multiplication associative et distributive.*

L'idée d'appartenance à un sous-corps de $\mathbb{C}$ *– le corps engendré par les racines d'une équation – est assez claire pour Abel et Galois, mais ils n'ont pas de mot pour désigner l'ensemble de ses éléments. C'est Dedekind qui introduit le terme de* corps *(en 1871), entendant par là des sous-corps de* $\mathbb{C}$, *pour lesquels il étudie en détail les notions d'intersection et d'automorphisme. Kronecker donne en 1882, les premiers exemples de corps définis abstraitement, corps de classes résiduelles de polynômes à coefficients rationnels modulo un polynôme* P. *Il faut attendre 1893 pour que Weber donne à l'expression corps commutatif le sens général actuel. Un peu plus tard vint l'abandon de la notion de commutativité, ce qui conduisit à considérer l'ensemble des quaternions définis par Hamilton en 1843 comme un corps. Weddeburn prouva que tout corps fini est nécessairement commutatif.*

Les exemples de corps vont se multiplier dans les dernières années du XIX^{e} *siècle : corps des nombres p-adiques, introduit par Hensel, premier exemple de la théorie générale des corps valués ; corps des séries formelles. Enfin, vers 1910, Steinitz développe la théorie générale des corps commutatifs et des extensions de corps telle que nous la connaissons aujourd'hui.*

3.1. Inversibilité à droite

Soit A un anneau. Un élément $a \in A$ est dit inversible à droite s'il existe b tel que $ab = 1$. On dit que a est un diviseur de 0 à droite si il existe $c \neq 0$ tel que $ac = 0$. On définit les mêmes notions à gauche.

1. Montrer que si a est inversible à droite il ne peut pas être diviseur de 0 à gauche. Donner un exemple montrant que a peut être inversible à droite et diviseur de 0 à droite.

2. Montrer que si a est inversible à droite et non diviseur de 0 à droite, alors il est inversible.

3. Prouver en fait que si a est inversible à droite et non inversible, alors il admet une infinité d'inverses à droite.

(**École Polytechnique**)

▷ **Solution.**

1. Soit b un inverse à droite de a (donc tel que $ab = 1$) et c tel que $ca = 0$. On multiplie à droite par b. Il vient $0 = (ca)b = c(ab) = c$. Donc a n'est pas diviseur de 0 à gauche. Rappelons que cela équivaut à l'injectivité du morphisme de groupes $x \mapsto xa$.

Donnons maintenant un exemple d'un élément inversible à droite et diviseur de 0 à droite. Prenons pour A l'anneau des endomorphismes d'un K-espace vectoriel E de dimension infinie. Une application linéaire a surjective et non injective va convenir. Par exemple avec $E = \mathbb{R}[X]$ pour espace vectoriel, on peut prendre la dérivation $a : P \mapsto P'$. L'application $b : P \mapsto \int_0^X P(t)dt$ est linéaire et on a $a \circ b = Id_E$. Donc a est bien inversible à droite. Mais si on prend l'application $c : P \mapsto P(0)$, on a $a \circ c = 0$ donc a est aussi un diviseur de 0 à droite. On voit facilement sur cet exemple que a admet une infinité d'inverses à droite.

2. Soit b un inverse à droite de a. On va montrer que $ba = 1$. On a $a = (ab)a = a(ba)$ de sorte que $a(1-ba) = 0$. Comme a n'est pas diviseur de 0 à droite cela implique $ba = 1$. Donc a est inversible d'inverse b.

Cela montre déjà qu'un élément inversible à droite et non inversible admet au moins deux inverses à droite.

3. Notons D l'ensemble des inverses à droite de a (ensemble supposé

non vide) et $N = \{x \in A, ax = 0\}$ (c'est un sous-groupe additif de A et même un idéal à droite). On a une bijection entre D et N puisque si b est un élément de D, alors $D = b + N$. On va utiliser le fait, vu ci-dessus, que $\alpha = ba - 1$ est dans N. C'est même un élément non nul de N puisque a est supposé non inversible. Pour tout x de N, l'élément $xa + \alpha$ est encore dans N et est *non nul* (car sinon $(x + b)a = 1$ et a serait inversible à gauche donc inversible). Ainsi, $\alpha a + \alpha = \alpha(a+1) \in N$, $\alpha(a+1)a + \alpha = \alpha(a^2 + a + 1) \in N$ etc. Une récurrence facile montre que pour tout $n \geqslant 0$, $u_n = \alpha(a^n + a^{n-1} + \cdots + a + 1)$ est dans N. On va montrer que ces éléments sont deux à deux distincts. Supposons $u_n = u_p$ avec $p < n$. On a alors $\alpha(a^n + \cdots + a^{p+1}) = 0$ et donc $u_{n-p-1}a^{p+1} = 0$. Mais comme a est inversible à droite, on a forcément $u_{n-p-1} = 0$. Ce qui n'est pas possible car les u_n sont tous non nuls. Donc N est infini et par suite D aussi. ◁

En particulier, dans un anneau fini, tout élément inversible à droite est inversible (ce qui est facile à établir directement puisque si a est inversible à droite, $x \mapsto xa$ est injective donc bijective par finitude de A).

3.2. Calcul d'inverse

Soit A un anneau et $(a, b) \in A^2$. On suppose $1 - ab$ inversible. Montrer que $1 - ba$ est aussi inversible.

(École normale supérieure)

Solution.

• Supposons dans un premier temps que ab est nilpotent et considérons $n \in \mathbb{N}^*$ tel que $(ab)^n = 0$. On a alors

$$c = (1 - ab)^{-1} = 1 + ab + (ab)^2 + \cdots + (ab)^{n-1}.$$

Mais ba est alors également nilpotent puisque $(ba)^{n+1} = b(ab)^n a = 0$. Donc $1 - ba$ est aussi inversible et on a

$$\begin{aligned}(1 - ba)^{-1} &= 1 + ba + (ba)^2 + \cdots + (ba)^n \\ &= 1 + b[1 + (ab) + \cdots + (ab)^{n-1}]a = 1 + bca.\end{aligned}$$

• Dans le cas général, posons $c = (1 - ab)^{-1}$ et $d = 1 + bca$. Il est naturel d'essayer cette valeur comme inverse de $1 - ba$. On obtient

$$(1-ba)d = (1-ba)(1+bca) = 1-ba+bca-babca = 1+b[c-abc-1]a = 1$$

et de même $d(1 - ba) = 1$. D'où le résultat. ◁

Les deux exercices suivants étudient des conditions suffisantes pour qu'un anneau soit commutatif. Le premier est un exemple d'anneau vérifiant une identité polynomiale.

3.3. Anneaux tels que $x^3 = x$

Soit A un anneau tel que $x^3 = x$ pour tout $x \in A$.

1. Déterminer les éléments nilpotents de A.

2. Soit $e \in A$ tel que $e^2 = e$, $a \in A$ et $b = ea(1-e)$.

Calculer b^2 et en déduire que $ea = ae$. En déduire que pour tout $x \in A$ $x^2 \in Z(A)$, où $Z(A)$ désigne le centre de A.

3. Montrer que A est commutatif.

(École normale supérieure)

▷ **Solution.**

1. Pour tout $n \in \mathbb{N}^*$ on a $x^n \in \{x, x^2\}$ (comme il résulte d'une récurrence immédiate sur n). Donc si x est nilpotent, il vérifie soit $x = 0$, soit $x^2 = 0$. Mais dans le second cas, on a $x = x^3 = xx^2 = 0$. Le seul élément nilpotent de A est donc 0.

2. On a $b^2 = ea(1-e)ea(1-e) = ea(e-e^2)a(1-e) = 0$. Donc b est nilpotent et d'après la question précédente, $b = 0$, c'est-à-dire $ea = eae$. En considérant de même $c = (1-e)ae$, on obtient $c^2 = 0$, puis $c = 0$ et $ae = eae$. Il en résulte que e commute avec a. On a ainsi montré que tout élément idempotent e est central.

Soit $x \in A$. On a $(x^2)^2 = x^4 = x^3x = x^2$: x^2 est idempotent et donc central.

3. Soit $x \in A$. On a $2x = (x+1)^2 - x^2 - 1 \in Z(A)$, d'après la question précédente, puisque $Z(A)$ est un sous-groupe additif de A. On a, d'autre part, $(x+1)^3 = x+1$. Comme $(x+1)^3 = x^3 + 3x^2 + 3x + 1 = 3x^2 + 4x + 1$, on obtient en simplifiant $3x^2 + 3x = 0$. L'élément $3x = -3x^2$ est donc dans $Z(A)$, puisque x^2 y est. Il en résulte que $x = 3x - 2x \in Z(A)$: tout élément de A est central, donc A est commutatif. ◁

Signalons un théorème dû à Jacobson : un anneau A *dans lequel, pour tout élément* x*, il existe un entier* $n(x) > 1$ *tel que* $x^{n(x)} = x$*, est commutatif.*

3.4. Commutativité ou anti-commutativité

Un pseudo-anneau est un triplet $(A, +, \cdot)$ qui vérifie tous les axiomes de la structure d'anneau, sauf celui qui affirme l'existence de l'élément-unité. Soit A un pseudo-anneau tel que, pour tout $(x, y) \in A^2$, $yx \in \{xy, -xy\}$. Montrer que A est commutatif ou anti-commutatif. Que dire si A est un anneau ?

(École normale supérieure)

Solution.

Notons Z le centre de A et posons $Z' = \{x \in A, \forall y \in A,\ yx = -xy\}$. Il est clair que Z et Z′ sont deux sous-groupes additifs de A.

Montrons qu'on a $Z \cup Z' = A$. Supposons par l'absurde qu'il existe $x \notin Z \cup Z'$. Alors, comme x n'est pas dans Z, il existe $y_1 \in A$ tel que $y_1 x \neq xy_1$. Par hypothèse on a nécessairement $y_1 x = -xy_1$. De même, comme $x \notin Z'$, il existe $y_2 \in A$ tel que $y_2 x \neq -xy_2$ et alors $y_2 x = xy_2$. Regardons maintenant l'élément $(y_1 + y_2)x = x(y_2 - y_1)$. Par hypothèse, il doit appartenir à $\{x(y_1+y_2), -x(y_1+y_2)\}$. Or, si $x(y_2-y_1) = x(y_1+y_2)$ on obtient $xy_1 = -xy_1 = y_1 x$, ce qui est contraire au choix de y_1. De même, si $x(y_2 - y_1) = -x(y_1 + y_2)$, il vient $xy_2 = -xy_2 = -y_2 x$, ce qui donne une nouvelle contradiction. On a donc $Z \cup Z' = A$.

Il est bien connu qu'un groupe ne peut pas être réunion de deux sous-groupes stricts. On obtient donc soit $Z = A$ auquel cas A est commutatif, soit $Z' = A$ auquel cas A est anti-commutatif.

Supposons que A soit un anneau. Si $Z' = A$, l'élément unité de A étant dans Z′, on a $1 = -1$ et A est de caractéristique 2. Mais dans ce cas, pour tout $(x, y) \in A^2$, $-xy = xy$ et A est commutatif. Ainsi, un anneau vérifiant l'hypothèse de l'exercice est toujours commutatif. ◁

3.5. Anneaux réguliers

Un anneau A est dit régulier si, pour tout $a \in A$, il existe $u \in A$ tel que $aua = a$.

1. Un corps est-il régulier ? $\mathbb{Z}$ est-il régulier ?

2. Trouver une condition nécessaire et suffisante pour que $\mathbb{Z}/n\mathbb{Z}$ soit régulier.

3. Montrer que si E est un K-espace vectoriel de dimension finie, $\mathcal{L}(E)$ est régulier. Soit $A = (a_{ij}) \in M_n(K)$ définie par $a_{i,i+1} = 1$ pour $1 \leqslant i \leqslant n-1$ et $a_{ij} = 0$ sinon. Trouver une matrice $U \in M_n(K)$ telle que $AUA = A$.

4. Montrer que le centre d'un anneau régulier est régulier.

(**École normale supérieure**)

Solution.

1. Observons que si $a = 0$, tout $u \in A$ vérifie $aua = a = 0$. Par ailleurs, si a est un élément inversible de A, en prenant $u = a^{-1}$ on obtient $aua = a$ et c'est d'ailleurs la seule valeur de u qui convienne. Comme dans un corps, tout élément non nul est inversible, il en résulte qu'un corps est régulier. L'anneau $\mathbb{Z}$ n'est pas régulier, comme on le voit en prenant $a = 2$: l'équation $4u = 2$ n'a pas de solution dans $\mathbb{Z}$. Remarquons d'ailleurs que si A est régulier et intègre, A est nécessairement un

corps. En effet, soit $a \neq 0$ et $u \in A$ tel que $aua = a$. Par intégrité de A, a est régulier à gauche et à droite, d'où en simplifiant $ua = au = 1$. Donc a est inversible et A est un corps.

2. Il est aisé de voir que si A et B sont deux anneaux réguliers, alors l'anneau produit $A \times B$ est régulier. En effet, si $(a,b) \in A \times B$, il existe $u \in A$ tel que $aua = a$ et $v \in B$ tel que $bvb = b$. Alors, $(a,b)(u,v)(a,b) = (aua, bvb) = (a,b)$. D'où le résultat.

Excluons les cas triviaux $n = 0$, où $\mathbb{Z}/n\mathbb{Z}$ est isomorphe à $\mathbb{Z}$ et donc non régulier, et $n = 1$, où l'on obtient l'anneau nul, clairement régulier. Si p est un nombre premier, $\mathbb{Z}/p\mathbb{Z}$ est un corps et est donc régulier. Si $n \geqslant 2$ est un entier quadratfrei, c'est-à-dire de la forme $n = p_1 \dots p_k$ où les p_i sont des nombres premiers distincts, on sait que $\mathbb{Z}/n\mathbb{Z}$ est isomorphe à $\mathbb{Z}/p_1\mathbb{Z} \times \cdots \times \mathbb{Z}/p_k\mathbb{Z}$ (théorème chinois). Et d'après la remarque précédente, $\mathbb{Z}/n\mathbb{Z}$ est régulier.

Montrons que si n n'est pas quadratfrei, alors $\mathbb{Z}/n\mathbb{Z}$ n'est pas régulier. On suppose donc qu'il existe $p \in \mathcal{P}$ tel que $\alpha = \nu_p(n) \geqslant 2$. On écrit $n = p^\alpha m$, où p ne divise pas $m \geqslant 1$ et on pose $k = \dfrac{\alpha}{2}$ si α est pair et $k = \dfrac{\alpha+1}{2}$ sinon. Soit a la classe modulo n de $p^k m$. Il est clair que a est non nulle (car $k < \alpha$), mais pour tout $u \in \mathbb{Z}/n\mathbb{Z}$, $aua = ua^2 = 0$ (car $2k \geqslant \alpha$).

Conclusion. Pour $n \geqslant 2$, $\mathbb{Z}/n\mathbb{Z}$ est régulier si et seulement si n est quadratfrei.

3. Soit $a \in \mathcal{L}(E)$ non inversible (si $a \in GL(E)$, $u = a^{-1}$ convient, cf. 1). On considère un supplémentaire F de $\operatorname{Ker} a$ et un supplémentaire G de $\operatorname{Im} a$. On sait que $a_{|F}$ établit un isomorphisme entre F et $\operatorname{Im} a$. Si $u = a_{|F}^{-1} \circ p$, où p est la projection sur $\operatorname{Im} a$ parallèlement à G, alors on a $aua = a$. L'anneau $\mathcal{L}(E)$ est donc régulier.

On considère

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ 0 & \ddots & 0 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 0 & 0 \end{pmatrix}$$

et l'endomorphisme a de $\mathbb{K}^n$ associé à la matrice A dans la base canonique, notée $(e_1, \dots, e_n)$. On obtient $\operatorname{Im}(a) = \operatorname{Vect}(e_1, \dots, e_{n-1})$, $\operatorname{Ker}(a) = \operatorname{Vect}(e_1)$ et avec les notations précédentes, $F = \operatorname{Vect}(e_2, \dots, e_n)$ et $G = \operatorname{Vect}(e_n)$. On vérifie, toujours avec les notations précédentes, que $(p(e_1), \dots, p(e_n)) = (e_1, e_2, \dots, 0)$ et $(u(e_1), \dots, u(e_n)) = (e_2, \dots, e_n, 0)$. La matrice U de u, dans la base canonique, est définie par $u_{i+1,i} = 1$

pour $1 \leqslant i \leqslant n-1$ et $u_{ij} = 0$ sinon. C'est la transposée de la matrice A. Elle vérifie AUA = A, puisque $aua = a$.

4. Soit Z le centre de l'anneau régulier A, $a \in Z$ et $u \in A$ tel que $aua = a$. On cherche un élément $v \in Z$ tel que $ava = a$. On va montrer que $v = uau$ convient. Il est clair que $ava = auaua = aua = a$. Il reste à montrer que v est central. Pour tout $b \in A$, on a, puisque a est dans Z, $(1-au)bau = (a-aua)bu = 0$ et de même, $aub(1-au) = ub(a-aua) = 0$. Il en résulte que $bau - aubau = aub - aubau = 0$ et donc que $bau = aub$. Ainsi, au est dans Z. Enfin, on obtient, en utilisant le fait que a et au sont éléments de Z, pour tout $b \in A$, $bv = buau = aubu = ubau = uaub = vb$. Ceci montre que v est dans Z. ◁

Dans tous les exercices qui suivent, les anneaux seront commutatifs. Nous rappelons qu'un anneau A *est dit intègre s'il est commutatif et sans diviseur de* 0*, ce qui veut dire qu'un produit ab est nul si et seulement si* a *ou* b *est nul. Un idéal d'un anneau commutatif* A *est un sous-groupe additif* I *tel que pour tout* x *de* I *et tout* a *de* A, $ax \in I$. *L'idéal* I *est dit principal s'il est engendré par un seul élément. On notera* (a) *ou aussi* aA *l'idéal principal engendré par* a.

3.6. Idéaux principaux

Soit A un anneau commutatif. Pour a, b dans A montrer que si l'idéal $(a) + (b)$ est principal, il en est de même de l'idéal $(a) \cap (b)$.
(**École normale supérieure**)

Solution.

Pour avoir des idées, on peut prendre un exemple simple bien connu $A = \mathbb{Z}$. Dans ce cas l'idéal $(a) + (b)$ est engendré par $d = \text{pgcd}(a, b)$ et l'idéal $(a) \cap (b)$ par $m = \text{ppcm}(a, b)$. Si on écrit $a = d\alpha$ et $b = d\beta$, on sait que m est associé à $d\alpha\beta$. Essayons donc ce générateur dans le cas général. On prend les notations ci-dessus : d est un générateur de $(a) + (b)$, $a = d\alpha$, $b = d\beta$ et on pose $m = d\alpha\beta$. Montrons par double inclusion que $(a) \cap (b) = (m)$.

• Soit $x \in (m)$. Il existe $\lambda \in A$ tel que $x = \lambda m = \lambda d\alpha\beta = \lambda\beta a = \lambda\alpha b$, de sorte que $x \in (a) \cap (b)$.

• Réciproquement, soit $x \in (a) \cap (b)$. On écrit $x = ua = vb$. On sait par ailleurs qu'il existe $(\lambda, \mu) \in A^2$ tel que $d = \lambda a + \mu b$. On a alors

$$x = ua = u\alpha d = u\alpha(\lambda a + \mu b) = \alpha\lambda x + u\mu m = \alpha\lambda vb + u\mu m = m(\lambda v + \mu u).$$

D'où la seconde inclusion. ◁

L'énoncé suivant concerne les idéaux maximaux, c'est-à-dire les idéaux distincts de A *qui sont maximaux au sens de l'inclusion.*

3.7. Idéaux maximaux

Soit A un anneau commutatif et I un idéal de A distinct de A.

1. Montrer que I est maximal pour l'inclusion parmi les idéaux de A distincts de A si et seulement si pour tout $a \in A \setminus I$, $I + aA = A$.

On dit qu'un idéal I est premier si $A \setminus I$ est stable pour le produit.

2. Montrer que tout idéal maximal est premier.

3. Donner un exemple d'idéal premier.

4. Soit A l'anneau des fonctions de classe $\mathcal{C}^\infty$ de $\mathbb{R}$ dans $\mathbb{R}$ et I l'idéal des fonctions de A s'annulant en 0. L'idéal I est-il principal ? premier ? maximal ?

5. Soit J l'ensemble des fonctions de A telles que $f^{(k)}(0) = 0$ pour tout $k \in \mathbb{N}$. L'ensemble J est-il un idéal ? Si oui, est-il principal ? premier ? maximal ?

(École Polytechnique)

▷ **Solution.**

1. Soit I un idéal de A, distinct de A. On note que $I + aA$ est encore un idéal. C'est le plus petit idéal de A contenant I et a.

Supposons que I est maximal. Pour tout $a \in A \setminus I$, $I + aA$ est un idéal de A contenant I et distinct de I, car $a = 0 + a \cdot 1$ appartient à $I + aA$. On a donc $I + aA = A$.

Supposons réciproquement que cette propriété est vérifiée. Si J est un idéal de A contenant I et distinct de I, on considère $a \in J \setminus I$. Alors J contient $I + aA = A$ donc $J = A$ et I est maximal.

2. Soit I un idéal maximal, a et b deux éléments qui ne sont pas dans I. Comme $I + aA = A$, il existe $c \in A$ et $i \in I$ tel que $1 = i + ac$. On en déduit $b = bi + abc$. Si $ab \in I$, alors $abc \in I$ et comme $bi \in I$, $b \in I$. C'est faux donc $ab \notin I$ et I est premier.

3. Si $A = \mathbb{Z}$, les idéaux $p\mathbb{Z}$, avec p premier, sont premiers. En effet, si a et b sont deux entiers tels que p divise ab, p divise a ou b. Ce sont les seuls idéaux premiers, car si n est entier naturel non premier, il existe des entiers n_1 et n_2 strictement supérieurs à 1 tels que $n = n_1 n_2$. Alors n_1 et n_2 n'appartiennent pas à $n\mathbb{Z}$ et $n_1 n_2 \in n\mathbb{Z}$, donc $n\mathbb{Z}$ n'est pas premier.

Notons que les idéaux $p\mathbb{Z}$, avec p premier, sont aussi maximaux. En effet si $a \notin p\mathbb{Z}$, a est premier avec p, donc $a\mathbb{Z} + p\mathbb{Z} = \text{pgcd}(a, p)\mathbb{Z} = \mathbb{Z}$.

4. • La fonction $\text{Id}_\mathbb{R}$ appartient à I. Montrons que I est principal, engendré par $\text{Id}_\mathbb{R}$. Si f est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ et s'annule en 0, la fonction

g définie par $g(x) = \dfrac{f(x)}{x}$ pour $x \neq 0$ et $g(0) = f'(0)$ est de classe $\mathcal{C}^\infty$ (théorème de division). Ce théorème a été démontré dans le tome 1 d'analyse (exercice 4.37). Rappelons-en rapidement une démonstration.

Pour $x \neq 0$, le changement de variable $t = ux$ donne

$$f(x) = \int_0^x f'(t)\,\mathrm{d}t = x\int_0^1 f'(ux)\,\mathrm{d}u \text{ et donc } g(x) = \int_0^1 f'(ux)\,\mathrm{d}u,$$

formule qui reste vraie pour $x = 0$. La fonction $h : (x,u) \longmapsto f'(xu)$ est indéfiniment dérivables par rapport à x et les dérivées partielles $\dfrac{\partial^k h}{\partial x^k}$: $(x,u) \longmapsto u^k f^{(k+1)}(xu)$ sont continues sur $\mathbb{R} \times [0,1]$. On en déduit que g est de classe $\mathcal{C}^\infty$.

Enfin, $f = g\,\mathrm{Id}_\mathbb{R}$ et $\mathrm{I} = \mathrm{Id}_\mathbb{R}\,\mathrm{A}$.

• L'idéal I est maximal car si f appartient à $\mathrm{A} \setminus \mathrm{I}$ et $g \in \mathrm{A}$, on a $f(0) \neq 0$, donc on peut écrire $g = \left(g - \dfrac{g(0)}{f(0)}f\right) + \dfrac{g(0)}{f(0)}f$; ainsi g appartient à $\mathrm{I} + f\mathrm{A}$. D'après la question 2, I est *a fortiori* premier.

5. • Il est clair que J est un sous-groupe additif de A. Si $f \in \mathrm{J}$ et $g \in \mathrm{A}$, la formule de Leibniz montre que fg a toutes ses dérivées nulles en 0. Elle est donc dans J : c'est un idéal de A.

• Montrons que J n'est pas principal en raisonnant par l'absurde. Soit f tel que $\mathrm{J} = f\mathrm{A}$. Comme $f(0) = 0$, il existe g dans A telle que $f = \mathrm{Id}_\mathbb{R}\, g$. La démonstration de la question 4 montre que $g(x) = \int_0^1 f'(ux)\,\mathrm{d}u$, pour tout réel x et donc que $g^{(k)}(x) = \int_0^1 u^k f^{(k+1)}(ux)\,\mathrm{d}u$, pour tout (k,x) de $\mathbb{N} \times \mathbb{R}$. On a, en particulier $g^{(k)}(0) = \int_0^1 u^k f^{(k+1)}(0)\,\mathrm{d}u = 0$. Ainsi $g \in \mathrm{J}$ et il existe $h \in \mathrm{A}$ telle que $g = fh = \mathrm{Id}_\mathbb{R}\, gh$. Montrons que g ne s'annule qu'en 0. Il est classique que la fonction φ définie sur $\mathbb{R}$ par $\varphi(x) = \begin{cases} e^{-\frac{1}{x^2}} & \text{si } x \neq 0 \\ 0 & \text{si } x = 0 \end{cases}$ appartient à J. Comme $\varphi \in f\mathrm{A}$ et ne s'annule qu'en 0, il en est de même de f et donc de g. On obtient donc $1 = xh(x)$ pour tout $x \neq 0$. En faisant tendre x vers 0, on obtient une contradiction. Donc J n'est pas principal.

• L'idéal J n'est pas maximal car il est contenu strictement dans I.

• Montrons que J est premier. Soit f et g deux éléments de $\mathrm{A} \setminus \mathrm{J}$. Notons $n = \inf\{k \in \mathbb{N}, f^{(k)}(0) \neq 0\}$ et $p = \inf\{k \in \mathbb{N}, g^{(k)}(0) \neq 0\}$. D'après la formule de Leibniz, on a

$$(fg)^{(n+p)}(0) = \sum_{k=0}^{n+p} \mathrm{C}_{n+p}^k f^{(k)}(0) g^{(n+p-k)}(0) = f^{(n)}(0) g^{(p)}(0) \neq 0,$$

car si $k < n$, $f^{(k)}(0) = 0$ et si $k > n$, $n + p - k < p$ et $g^{(n+p-k)}(0) = 0$. Ainsi fg n'appartient pas à J. ◁

Un anneau principal est un anneau intègre dont tout idéal est principal ; nous savons, en particulier, que les anneaux $\mathbb{Z}$ et $\mathrm{K}[\mathrm{X}]$, où K est un corps, sont principaux. Les exercices suivants sont des études de principalité.

3.8. Anneau des nombres décimaux

Soit D l'anneau des nombres décimaux :

$$\mathrm{D} = \{x \in \mathbb{Q},\ \exists n \in \mathbb{Z},\ x \cdot 10^n \in \mathbb{Z}\}.$$

Montrer que D est principal.

(École polytechnique)

▷ **Solution.**

Il est évident que D est un sous-anneau de $\mathbb{Q}$. Remarquons que les éléments de D sont les nombres rationnels dont le dénominateur possède comme seuls diviseurs premiers 2 et 5. Tout élément non nul x de D s'écrit donc de manière unique $x = 2^\alpha 5^\beta p.$ avec $p \in \mathbb{Z}$ premier avec 10 et α et β dans $\mathbb{Z}$.

La démonstration de la principalité de D sera calquée sur celle de $\mathbb{Z}$. On considère un idéal non nul I de D (car l'idéal nul est évidemment principal). Si, avec les notations précédentes, $x = 2^\alpha 5^\beta p$ est un élément non nul de I, alors $p = 2^{-\alpha}5^{-\beta}x$ est un élément de I, car $2^{-\alpha}5^{-\beta}$ appartient à D. Ainsi $|p|$ appartient à $\mathbb{N}^* \cap \mathrm{I}$. Cet ensemble est une partie non vide de $\mathbb{N}^*$, qui possède un plus petit élément a. On a clairement $a\mathrm{D} \subset \mathrm{I}$. Réciproquement, si $x = 2^\alpha 5^\beta p$ appartient à I, il en est de même de l'entier p. On divise p par a : $p = aq + r$. On obtient que $r = p - aq \in \mathrm{I} \cap \mathbb{N}^*$, avec $0 \leqslant r < a$. De la définition de a, on déduit que $r = 0$, que $p = aq$ et donc que $x = a(q2^\alpha 5^\beta)$. Ceci montre que $\mathrm{I} \subset a\mathrm{D}$ et donc que $\mathrm{I} = a\mathrm{D}$. ◁

Le lecteur pourra montrer de la même manière que tout sous-anneau de $\mathbb{Q}$ est principal.

3.9. Anneau $\mathbb{Z}[\mathrm{X}]$

1. L'anneau $\mathbb{Z}[\mathrm{X}]$ est-il principal ?
2. Soit A un anneau commutatif. À quelle condition $\mathrm{A}[\mathrm{X}]$ est-il principal ?

(École normale supérieure)

Solution.

1. Pour montrer la non-principalité, il suffit d'exhiber un idéal non principal. C'est le cas de l'idéal I engendré par 2 et X. En effet, supposons qu'il existe $P \in \mathbb{Z}[X]$ tel que $I = (P)$. Comme $2 \in I$, P doit être constant et diviser 2. Donc $P = \pm 1$ ou $P = \pm 2$. Mais comme 2 ne divise pas X dans $\mathbb{Z}[X]$, on a $P = \pm 1$. L'idéal $I = (2) + (X)$ contient alors 1, et on peut écrire $1 = 2U + XV$, avec U et V dans $\mathbb{Z}[X]$. On obtient, en considérant les valeurs en 0 des polynômes, $1 = 2U(0)$, ce qui est absurde, puisque $U(0) \in \mathbb{Z}$.

2. Nous savons que si A est un corps, A[X] est principal. Montrons que c'est une condition nécessaire. On raisonne par contraposition, en supposant que A n'est pas un corps. Il possède donc un élément non inversible a. Comme dans la première question, on considère l'idéal engendré par a et X. Supposant qu'il est principal, on trouve de nouveau qu'il est engendré par l'élément unité de A, que celui-ci s'écrit $1 = aU + XV$ et qu'enfin $1 = aU(0)$, ce qui contredit la non-inversibilité de a.

Conclusion. A[X] est principal si et seulement si A est un corps. ◁

L'arithmétique de l'anneau $\mathbb{Z}[X]$ est tout de même assez simple car il est factoriel, c'est-à-dire qu'un élément non nul se décompose de manière essentiellement unique en un produit d'irréductibles (d'après un théorème général de Gauss qui affirme que si A *est un anneau factoriel,* A[X] *l'est aussi). Pour la définition précise d'un anneau factoriel voir l'exercice suivant, qui démontre que tout anneau principal est factoriel.*

3.10. Anneaux factoriels

Soit A un anneau intègre. Un élément $p \in A^*$ est dit irréductible s'il n'est pas inversible et si, pour toute écriture $p = uv$, avec u et v dans A, on a u ou v inversible. Si $(a, b) \in A^2$, a et b sont dits associés si a s'écrit $a = ub$, avec u inversible. «Être associés» est une relation d'équivalence sur A.

A est dit factoriel si tout $a \in A^*$ peut s'écrire $a = up_1^{\alpha_1} \dots p_r^{\alpha_r}$ avec u inversible, les p_i irréductibles deux à deux non associés, et les $\alpha_i \in \mathbb{N}^*$. De plus cette écriture est unique au sens suivant : si $a = vq_1^{\beta_1} \dots q_s^{\beta_s}$ avec v inversible, les q_i irréductibles deux à deux non associés, les $\beta_i \in \mathbb{N}^*$, on a $r = s$ et quitte à renuméroter les q_i, p_i et q_i sont associés et $\alpha_i = \beta_i$ pour tout $1 \leqslant i \leqslant r$.

Montrer qu'un anneau principal est factoriel.

(École normale supérieure)

▷ **Solution.**

• Existence de l'écriture : dans $\mathbb{Z}$, l'existence de la décomposition de n s'obtient par récurrence sur n, dans K[X] où K est un corps commutatif, la décomposition de P s'obtient par récurrence sur deg P. Dans notre cas de figure, une démonstration par récurrence s'avère impossible car on ne voit pas sur quel entier elle se ferait. On va raisonner par l'absurde et supposer qu'il existe $a \in A^*$ qui n'admet pas une telle décomposition. Alors a n'est ni inversible, ni irréductible et par conséquent, il s'écrit $a = uv$ avec u et v dans A^* non inversibles. Si u et v admettaient une décomposition, le produit $a = uv$ en admettrait une aussi, ce qui est exclu par hypothèse. Supposons par exemple que u n'admet pas de décomposition. On pose $a_1 = u$. On a donc trouvé a_1 n'admettant de décomposition, divisant a et non associé à a puisque v est non inversible.

En réitérant ce raisonnement, on construit une suite $(a_n)_{n\in\mathbb{N}}$, avec $a_0 = a$ et pour tout $n \in \mathbb{N}$, a_{n+1} divise a_n, n'est pas associé à a_n et n'admet pas de décomposition.

Considérons les idéaux $I_n = a_nA$ pour tout $n \in \mathbb{N}$. Comme a_{n+1} divise a_n, $I_n \subset I_{n+1}$. L'inégalité est stricte. En effet, si elle ne l'était pas, on aurait $a_n | a_{n+1}$ et il existerait u et v dans A tels que $a_n = ua_{n+1}$ et $a_{n+1} = va_n$. Mais alors $a_n = uva_n$ et comme a_n est non nul et A intègre, $uv = 1$. Il en résulterait que u serait inversible et a_n et a_{n+1} associés, ce qui est exclu.

Considérons $I = \bigcup_{n\in\mathbb{N}} I_n$. Montrons que I est un idéal. Il est non vide. Soit $\alpha \in A$ et $(x, y) \in I^2$. Il existe $(n, p) \in \mathbb{N}^2$ tel que $x \in I_n$ et $y \in I_p$. On pose $m = \max(n, p)$. Alors x et y sont dans I_m puisque la suite $(I_n)_{n\in\mathbb{N}}$ est croissante. On a donc $x + y \in I_m \subset I$ et $\alpha x \in I_m \subset I$.

Comme A est un anneau principal, il existe donc $b \in I$ tel que $I = bA$. Par définition de I, il existe $n_0 \in \mathbb{N}$ tel que $b \in I_{n_0}$. Mais alors si $n \geqslant n_0$, on a

$$bA \subset I_{n_0} \subset I_n \subset I = bA$$

et la suite $(I_n)_{n\in\mathbb{N}}$ est stationnaire à partir du rang n_0, ce qui contredit sa stricte croissante. L'hypothèse absurde était donc l'existence de a indécomposable. On en déduit qu'il existe pour tout $a \in A^*$ une décomposition en produit d'irréductibles[1].

• Prouvons maintenant l'unicité de l'écriture. Soit $a \in A^*$. On suppose que l'on a deux écritures :

$$a = up_1^{\alpha_1} \dots p_r^{\alpha_r} = vq_1^{\beta_1} \dots q_s^{\beta_s}$$

1. Notons que la preuve demeure valide si A a la propriété que toute suite croissante d'idéaux est stationnaire. Un anneau intègre vérifiant cette propriété est dit noethérien.

avec u et v inversibles, les p_i irréductibles deux à deux non associés, les q_j irréductibles deux à deux non associés, les α_i et β_j dans $\mathbb{N}^*$. Montrons par récurrence sur $\alpha_1 + \cdots + \alpha_r$ que $r = s$ et que, quitte à renuméroter les q_j, on a p_i et q_i associés et $\alpha_i = \beta_i$ pour tout $1 \leqslant i \leqslant r$.

⋆ Si $\alpha_1 + \cdots + \alpha_r = 0$, $r = 0$, a est inversible donc nécessairement $s = 0$ (si $s \geqslant 1$, q_1 serait alors inversible).

⋆ Supposons $\alpha_1 + \cdots + \alpha_r \geqslant 1$. Alors p_r divise a. Démontrons le lemme suivant qui peut encore s'appeler lemme d'Euclide :

Lemme. *Soit p irréductible divisant la produit ab avec a et b dans* A. *Alors p divise a ou p divise b.*

Démonstration.

Considérons l'idéal engendré par p et b : $\mathrm{I} = p\mathrm{A} + b\mathrm{A}$. Comme A est principal, il existe $c \in \mathrm{A}$ tel que $\mathrm{I} = c\mathrm{A}$. Puisque p est dans I, c divise p. Donc c est inversible ou c est associé à p puisque p est irréductible. Dans le premier cas, $\mathrm{I} = \mathrm{A}$ et 1 s'écrit $1 = up + vb$ avec $(u, v) \in \mathrm{A}^2$. Donc $a = upa + vab$ et comme p divise ab, il divise a.

Dans le second cas, c associé à p, $\mathrm{I} = c\mathrm{A} = p\mathrm{A}$ et comme $b \in \mathrm{I}$, p divise b. Le lemme est prouvé. ◇

Comme p_r est irréductible, diviseur de $a = vq_1^{\beta_1} \dots q_s^{\beta_s}$, d'après le lemme d'Euclide étendu par récurrence à un nombre quelconque de facteurs, il divise l'un des facteurs. Il ne peut diviser v puisque sinon, p_r serait inversible. Il divise donc l'un des q_i. Comme q_i est irréductible, q_i et p_r sont nécessairement associés. Quitte à les numéroter et à changer v en un autre inversible, on peut supposer $q_s = p_r$. On divise alors a par p_r. Il vient

$$\frac{a}{p_r} = up_1^{\alpha_1} \dots p_{r-1}^{\alpha_{r-1}} p_r^{\alpha_r - 1} = vq_1^{\beta_1} \dots q_{s-1}^{\beta_{s-1}} p_r^{\beta_s - 1}$$

Si $\alpha_r - 1 > 0$, le raisonnement précédent prouve que, puisque p_r divise $\dfrac{a}{p_r}$, il divise l'un des q_i. Comme les q_i sont non associés deux à deux, p_r divise nécessairement le facteur $p_r^{\beta_s - 1} = q_s^{\beta_s - 1}$ et en particulier $\beta_s - 1 > 0$. On est en droit d'appliquer l'hypothèse de récurrence à $\dfrac{a}{p_r}$: on a $r = s$ et quitte à renuméroter les q_i, p_i et q_i sont associés et $\alpha_i = \beta_i$ pour $i < r$, $\alpha_r - 1 = \beta_r - 1$ ce qui donne bien $\alpha_r = \beta_r$.

Si $\alpha_r - 1 = 0$, alors nécessairement $\beta_s - 1 = 0$. En effet, dans le cas contraire $q_s = p_r$ diviserait $up_1^{\alpha_1} \dots p_{r-1}^{\alpha_{r-1}}$ et donc d'après le lemme d'Euclide, il diviserait l'un des p_i avec $i < r$ ce qui est impossible puisque les p_i sont deux à deux non associés. On applique alors l'hypothèse de récurrence à

$$\frac{a}{p_r} = up_1^{\alpha_1} \dots p_{r-1}^{\alpha_{r-1}} = vq_1^{\beta_1} \dots q_{s-1}^{\beta_{s-1}}$$

et on obtient : $r-1=s-1$; quitte à renuméroter les q_i, p_i et q_i sont associés et $\alpha_i=\beta_i$ pour $i<r$. Comme on avait déjà $p_r=q_r$ et $\alpha_r=1=\beta_r$, l'unicité de la décomposition est prouvée.

Conclusion. Un anneau principal est factoriel. ◁

L'exercice suivant rappelle la définition d'un anneau euclidien, montre qu'un anneau euclidien est principal et s'intéresse aux anneaux euclidiens dans lesquels la division euclidienne est unique.

3.11. Anneaux euclidiens

A étant un anneau intègre, on dit que A est euclidien lorsqu'il existe une application $\varphi : A\backslash\{0\} \longrightarrow \mathbb{N}$ telle que, pour $a \in A$ et $b \in A \setminus \{0\}$. il existe $(q,r) \in A^2$ tel que

$$a=bq+r \text{ et } r=0 \text{ ou } \varphi(r)<\varphi(b). \quad (*)$$

1. Donner des exemples d'anneaux euclidiens.

2. Montrer qu'un anneau euclidien est principal.

3. On suppose de plus que A n'est pas un corps et que, pour tout a et b le couple (q,r) défini par $(*)$ est unique. Montrer que A est isomorphe à l'anneau des polynômes sur un certain corps.

(École normale supérieure)

▷ **Solution.**

1. • On peut prendre $A=\mathbb{Z}$ et $\varphi : x \longmapsto |x|$. En effet, soit $a \in \mathbb{Z}$, $b \in \mathbb{Z}\backslash\{0\}$. Si q et r désignent respectivement le quotient et le reste de la division euclidienne de a par $|b|$, on a bien

$$a=b\varepsilon q+r \text{ et } r=0 \text{ ou } 0 \leqslant r<|b|,$$

avec $|b|=\varepsilon b$ et $\varepsilon=\pm 1$.

• L'autre exemple du cours est K[X] où K est un corps commutatif : la division euclidienne donne le couple q et r et pour l'application φ, il suffit de prendre la fonction degré.

• L'exercice 3.12 montre que l'anneau $\mathbb{Z}[i]$ des entiers de Gauss est euclidien.

2. Soit A un anneau euclidien. Par définition, A est intègre. Considérons un idéal I de A et montrons que I est principal. C'est clair si I est nul. Si $I \neq \{0\}$, considérons l'ensemble $F=\{\varphi(a) \in \mathbb{N}, a \in I, a \neq 0\}$. C'est une partie non vide de $\mathbb{N}$, qui admet donc un plus petit élément. Il existe donc un élément a_0 non nul de I tel que $\varphi(a_0)=\min F$. On a, bien entendu, $a_0 A \subset I$.

Réciproquement, si $a \in \mathrm{I}$, il existe $(q, r) \in \mathrm{A}^2$ tel que $a = qa_0 + r$ avec $r = 0$ ou $\varphi(r) < \varphi(a_0)$. On en déduit que $r = a - a_0 q$ est dans I, puisque I est un idéal. Si r n'est pas nul, on obtient $\varphi(r) < \varphi(a_0) = \min \mathrm{F}$, ce qui est impossible. On a donc nécessairement $r = 0$ et $a = a_0 q \in a_0\mathrm{A}$. On conclut que $\mathrm{I} = a_0\mathrm{A}$: l'idéal I est principal.

On note que les exemples d'anneaux principaux les plus importants du cours sont des anneaux euclidiens. Il est assez difficile d'exhiber des anneaux principaux non euclidiens mais cela existe[2].

3. Il s'agit de sortir de notre chapeau le corps K dont A serait l'anneau des polynômes. Or, dans K[X], le corps K est composé de 0 et des éléments de degré minimal (à savoir 0). Cela nous donne une piste pour définir K : il serait constitué, outre 0, des éléments x non nuls de A, pour lesquels $\varphi(x)$ est minimal.

• Avant d'entrer dans le détail, nous allons introduire plusieurs notations et faire quelques remarques simples. A^* désigne comme d'habitude $\mathrm{A} \setminus \{0\}$ et $\mathrm{A}^\times$ l'ensemble des éléments inversibles de A. Deux éléments a et b de A sont dits associés s'il existe u inversible dans A tel que $a = bu$ (être associés est une relation d'équivalence). Enfin, si dans A, $a = bq + r$ avec $b \neq 0$, $r = 0$ ou $\varphi(r) < \varphi(b)$, on dira qu'il s'agit d'une division euclidienne de a par b. Par hypothèse, une division de a par b est unique.

Lemme. *Si a et b appartiennent à A et si $a|b$, alors $\varphi(a) \leqslant \varphi(b)$.*
Si a et b sont associés dans A^*, $\varphi(a) = \varphi(b)$.

Démonstration. En effet, si $a|b$ et si $\varphi(a) > \varphi(b)$, on a deux divisions euclidiennes de b par a : $b = a \times c + 0$ et $b = a \times 0 + b$, où $c = \dfrac{b}{a}$, ce qui est contraire à l'hypothèse. Si a et b sont associés, on a $a|b$ et $b|a$, d'où résultent les relations $\varphi(a) \leqslant \varphi(b)$ et $\varphi(b) \leqslant \varphi(a)$ et l'égalité annoncée. ◇

• Pour tout $a \in \mathrm{A}^*$, 1 divise a ; on a donc $\varphi(1) \leqslant \varphi(a)$. Il s'ensuit que $m_0 = \varphi(1) = \min \varphi$. Puisque 1 est associé à tout inversible de A, on obtient

$$\mathrm{A}^\times \subset \{x \in \mathrm{A}^*,\ \varphi(x) = m_0\}$$

Réciproquement, si $x \in \mathrm{A}^*$ et $\varphi(x) = m_0$, x est inversible. En effet, il existe q et r dans A tels que $1 = xy + r$ avec $r = 0$ ou $\varphi(r) < \varphi(x)$. Comme $\varphi(x) = m_0 = \min \varphi$, cette seconde condition est impossible, et on a $1 = xy$ et $x \in \mathrm{A}^\times$. On conclut :

$$\mathrm{A}^\times = \{x \in \mathrm{A}^*, \varphi(x) = m_0\}$$

2. PERRIN (D.), *Cours d'algèbre*, Ellipses, 1996, p. 53-54.

• Tout nous pousse à poser $K = A^\times \cup \{0\}$. Montrons que K est un corps commutatif.

⋆ Comme $K^* = A^\times$, nous savons déjà que $(K^*, \times)$ est un groupe commutatif.

Montrons que K est un sous-anneau de A.

⋆ K^* et donc K sont stables par multiplication.

⋆ $1 \in K$ et si $a \in K$, alors $-a \in K$.

⋆ Il reste à montrer la stabilité par addition. Soit a et b dans K. Si a ou b est nul, il est clair que $a + b$ est dans K. Supposons $a \neq 0$ et $b \neq 0$. Étant donné que $a + b = a(1 + a^{-1}b)$ et que K est stable par multiplication, il suffit de prouver le résultat suivant : si $u \in K^*$, alors $1 + u \in K$.

Raisonnons par l'absurde et supposons que $1 + u \notin K$. Alors, $1 + u$ est non nul, $\varphi(1 + u) > m_0 = \varphi(1) = \varphi(-u^{-1})$ et on peut écrire

$$1 = (1 + u) \times 0 + 1, \text{ mais aussi } 1 = (1 + u) \times u^{-1} + (-u^{-1}).$$

On a donc deux divisions euclidiennes de 1 par $1+u$, ce qui est contraire à l'hypothèse.

Établissons un bilan provisoire : K est corps commutatif, sous-anneau de A. Comme A n'est par un corps, $K \subsetneq A$.

• Dans K[X], X est un polynôme de degré 1, *i.e.* de degré minimal parmi les polynômes non constants. Il est donc naturel de considérer ici un élément de $A \setminus K$, où φ est minimale. L'entier $m_1 = \min_{x \in A \setminus K} \varphi(x)$ est bien défini (car $A \setminus K \neq \emptyset$) et il existe $X_0 \in A \setminus K$ tel que $\varphi(X_0) = m_1$. Introduisons la fonction

$$\Psi : \begin{array}{ccc} K[X] & \longrightarrow & A \\ P & \longmapsto & P(X_0) \end{array}$$

Il s'agit d'un morphisme de K-algèbres. Montrons que Ψ est un isomorphisme.

• Prouvons l'injectivité de Ψ : montrons par récurrence sur $n \in \mathbb{N}$, que si $\deg P \leqslant n$ et $P(X_0) = 0$, alors $P = 0$.

⋆ C'est bon si $P = 0$ ou $\deg P = 0$ (car alors $P(X_0) = P \in K$).

⋆ Supposons $n \geqslant 1$ et la propriété vraie jusqu'au rang $n - 1$. Soit $P \in K[X]$ tel que $P(X_0) = 0$. En écrivant $P = a_n X^n + \cdots + a_1 X + a_0$, on obtient

$$0 = P(X_0) = (a_n X_0^{n-1} + \cdots + a_1)X_0 + a_0$$

On reconnaît là une division euclidienne de 0 par X_0. En effet soit $a_0 = 0$, soit $a_0 \in A^\times$ et $\varphi(a_0) = m_0 < \varphi(X_0) = m_1$. Puisque $0 = 0 \times X_0 + 0$, on obtient par unicité $a_0 = 0$ et $Q(X_0) = 0$ avec $Q = a_n X^{n-1} + \cdots + a_1$.

Comme $\deg Q \leqslant n-1$, par l'hypothèse de récurrence, on obtient $Q = 0$. Il en résulte que $P = 0$.

• La surjectivité est plus délicate. On veut montrer que les éléments de A sont dans $K[X_0] = \operatorname{Im}\Psi$. C'est vrai pour les éléments $x \in A^*$ tels que $\varphi(x) = m_0$: ils sont dans K. On va tenter une récurrence sur « $\varphi(x)$ » que l'on pourrait diminuer par une division euclidienne par X_0. Plus précisément, montrons par récurrence sur $n \in \mathbb{N}$ que si $x \in A^*$ et $\varphi(x) \leqslant n$, alors $x \in K[X_0]$.

C'est vrai si $n \leqslant m_0$.

Supposons $n > m_0$, et la propriété vérifiée jusqu'à l'entier $n-1$. Prenons $x \in A^*$ avec $\varphi(x) \leqslant n$. On peut supposer $\varphi(x) > m_0$ *i.e.* $x \notin K$. Divisons x par X_0 : il existe q et r dans A tels que $x = X_0 q + r$ avec $r = 0$ ou $\varphi(r) < \varphi(X_0)$. Par conséquent, r appartient à K. Pour montrer que x est dans $K[X]$, il suffit de montrer que q est dans $K[X_0]$. D'après l'hypothèse de récurrence, c'est vrai si $\varphi(q) \leqslant n-1$. Or, on a $\varphi(X_0 q) = \varphi(x-r)$. Puisque $\varphi(x) \leqslant n$, l'inégalité $\varphi(q) \leqslant n-1$ va résulter des deux inégalités suivantes : $\varphi(x-r) = \varphi(x)$ et $\varphi(q) < \varphi(X_0 q)$ (ce qui est le cas lorsque φ est la fonction degré dans $K[X]$). On les déduit de deux lemmes.

Lemme. *Si $x \notin K$ et $a \in K$, alors $\varphi(x+a) = \varphi(x)$.*

Démonstration. C'est évident si $a = 0$. Sinon, a est inversible, $a + x$ est associé à $1 + a^{-1}x$ et x à $a^{-1}x$. On en déduit que

$$\varphi(a+x) = \varphi(1+a^{-1}x) \quad \text{et} \quad \varphi(a^{-1}x) = \varphi(x).$$

On peut donc se ramener à montrer que si $u \notin K$, alors $\varphi(1+u) = \varphi(u)$. On remarque que $1+u$ n'est pas dans K, sinon u y serait (K est un corps et contient 1). On a donc $1+u \neq 0$ et

$$1 = (1+u) \times 1 - u = (1+u) \times 0 + 1.$$

Comme $\varphi(1+u) > \varphi(1)$, on a nécessairement, par unicité de la division, $\varphi(u) \geqslant \varphi(1+u)$. En remplaçant u par $-(1+u)$, qui lui aussi n'est pas dans K, on obtient de même $\varphi(-(1+u)) \geqslant \varphi(1-(1+u))$. On a donc $\varphi(1+u) = \varphi(-(1+u)) \geqslant \varphi(-u) = \varphi(u)$. On conclut que $\varphi(1+u) = \varphi(u)$. Le lemme est prouvé. ◇

Lemme. *Si $x \neq 0$, $\varphi(xX_0) > \varphi(x)$.*

Démonstration. Raisonnons par l'absurde et supposons $\varphi(xX_0) \leqslant \varphi(x)$. Comme x divise xX_0, on a $\varphi(xX_0) \geqslant \varphi(x)$ et donc $\varphi(xX_0) = \varphi(x)$. Divisons x par $X_0 x$: il existe q et r dans A tels que

$$x = xX_0 q + r \quad \text{et} \quad r = 0 \text{ ou } \quad (r) < \quad (xX_0 \quad = \quad x).$$

On remarque que x divise r, ce qui implique, si $r \neq 0$, que $\varphi(x) \leqslant \varphi(r)$. Cette inégalité n'est pas vérifiée. On a donc $r = 0$. On obtient alors $x = x\mathrm{X}_0 q$ et, par intégrité, $1 = \mathrm{X}_0 q$, ce qui montre que X_0 est inversible, donc dans K. C'est faux, par hypothèse. $\diamondsuit$

Résumons : $\varphi(q) < \varphi(\mathrm{X}_0 q) = \varphi(x)$ et donc $\varphi(q) \leqslant n-1$. D'après l'hypothèse de récurrence, $q \in \mathrm{K}[\mathrm{X}_0]$ et finalement $x \in \mathrm{K}[\mathrm{X}_0]$. $\lhd$

Conclusion. L'anneau $\mathrm{A} = \mathrm{K}[\mathrm{X}_0]$ est isomorphe à l'anneau des polynômes $\mathrm{K}[\mathrm{X}]$.

On retiendra des exercices précédents la chaîne d'implications : pour un anneau commutatif et intègre A,

$$\mathrm{A}\ \textit{euclidien} \Longrightarrow \mathrm{A}\ \textit{principal} \Longrightarrow \mathrm{A}\ \textit{factoriel}.$$

Nous allons, dans les deux exercices qui suivent, nous intéresser aux entiers de Gauss, exemple historiquement important, qui donne le coup d'envoi de la théorie algébrique des nombres. Le premier démontre que $\mathbb{Z}[i]$ est euclidien (donc principal et factoriel), ce qui conduit à une arithmétique proche de celle de $\mathbb{Z}$. Le lecteur aura intérêt, avant de l'aborder, à affermir ses connaissances sur les anneaux factoriels en traitant l'exercice 3.10.

3.12. Anneau des entiers de Gauss (1)

On considère l'anneau $\mathbb{Z}[i] = \{u + iv \in \mathbb{C},\ (u,v) \in \mathbb{Z}^2\}$ et l'application $\varphi : a \in \mathbb{Z}[i] \longmapsto a\overline{a} \in \mathbb{N}$.

1. Déterminer le groupe multiplicatif des éléments inversibles de $\mathbb{Z}[i]$.

2. Montrer que $\mathbb{Z}[i]$ est euclidien pour φ, c'est-à-dire que pour tout $(a,b) \in \mathbb{Z}[i] \times \mathbb{Z}[i]^*$, il existe $(q,r) \in \mathbb{Z}[i]^2$ tel que $a = bq + r$ avec $\varphi(r) < \varphi(b)$.

3. En déduire que $\mathbb{Z}[i]$ est factoriel.

4. Montrer que si $\varphi(a)$ est premier, alors a est irréductible dans $\mathbb{Z}[i]$.

5. Soit p un nombre premier. Montrer l'équivalence entre les propriétés suivantes :

(i) p est irréductible dans $\mathbb{Z}[i]$;

(ii) $p \equiv 3 \pmod 4$;

(iii) il n'existe pas $a \in \mathbb{Z}[i]$ tel que $p = \varphi(a)$.

On admettra que -1 est un carré dans $\mathbb{Z}/p\mathbb{Z}$ si et seulement si p n'est pas congru à 3 modulo 4.

6. En déduire tous les irréductibles de $\mathbb{Z}[i]$.

(École normale supérieure)

Solution.

1. Soit a un élément inversible de $\mathbb{Z}[i]$. Il existe b dans $\mathbb{Z}[i]$ tel que $ab = 1$. Comme φ est multiplicative, il vient $\varphi(a)\varphi(b) = 1$. L'entier naturel $\varphi(a)$ est inversible dans $\mathbb{Z}$ et donc égal à 1. Réciproquement si $\varphi(a) = 1$, alors $a\overline{a} = 1$ et $\overline{a}$, qui appartient à $\mathbb{Z}[i]$, est l'inverse de a.

Un élément $a \in \mathbb{Z}[i]$ est donc inversible si et seulement si $\varphi(a) = 1$. En posant $a = u + iv$, avec $(u, v) \in \mathbb{Z}^2$, on obtient $u^2 + v^2 = 1$. Les éléments inversibles de $\mathbb{Z}[i]$ sont donc $1, -1, i$ et $-i$: il s'agit du groupe des racines quatrième de l'unité.

2. Soit $(a, b) \in \mathbb{Z}[i] \times \mathbb{Z}[i]^*$. Considérons le nombre complexe $\dfrac{a}{b}$ qui s'écrit $\dfrac{a}{b} = x + iy$, avec $(x, y) \in \mathbb{R}^2$, u (resp. v) l'entier le plus proche de x (resp. y) et posons $q = u + iv$.

On a alors

$$\left|\frac{a}{b} - q\right|^2 = (u - x)^2 + (v - y)^2 \leqslant \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 < 1.$$

On en déduit $|a - bq|^2 < |b|^2$. Posons $r = a - bq$; a, b, q appartenant à $\mathbb{Z}[i]$, r appartient également à $\mathbb{Z}[i]$ et on a $|r|^2 < |b|^2$, c'est-à-dire $\varphi(r) < \varphi(b)$.

Nous avons déterminé $(q, r) \in \mathbb{Z}[i]^2$ tel que $a = bq + r$ et $\varphi(r) < \varphi(b)$.

Conclusion. L'anneau $\mathbb{Z}[i]$ est un anneau euclidien.

3. D'après l'exercice 3.11, $\mathbb{Z}[i]$ est un anneau principal et d'après l'exercice 3.10. il est alors factoriel.

4. Soit $a \in \mathbb{Z}[i]$ tel que $\varphi(a)$ soit premier. Supposons que a s'écrive $a = bc$. avec b et c dans $\mathbb{Z}[i]$. On a alors $\varphi(a) = \varphi(b)\varphi(c)$. L'entier $\varphi(a)$ étant premier, on en déduit que $\varphi(b) = 1$ ou $\varphi(c) = 1$, c'est-à-dire que b ou c est inversible. Donc a est irréductible.

5. Supposons (i) et montrons (ii) en raisonnant par l'absurde. Si p n'est pas congru à 3 modulo 4, alors -1 est un carré dans $\mathbb{Z}/p\mathbb{Z}$ et il existe $x \in \mathbb{Z}$ tel que $-1 \equiv x^2 \pmod{p}$. On obtient alors que p divise $x^2 + 1 = (x + i)(x - i)$ dans $\mathbb{Z}$ et donc dans $\mathbb{Z}[i]$. Puisque p est irréductible, il divise $x + i$ ou $x - i$. Mais ceci est impossible, car ni $\dfrac{x}{p} + \dfrac{i}{p}$, ni $\dfrac{x}{p} - \dfrac{i}{p}$ ne sont des éléments de $\mathbb{Z}[i]$.

Supposons (ii) et montrons (iii), toujours par l'absurde. Si $p = \varphi(a)$, alors $p = u^2 + v^2$, où u et v sont deux entiers tels que $a = u + iv$. Mais un carré étant toujours congru à 0 ou 1 (mod 4), p ne peut être congru qu'à 0, 1 ou 2 (mod 4), ce qui contredit (ii).

Supposons (iii) et démontrons (i). Soit a et b dans $\mathbb{Z}[i]$ tels que $p = ab$. On a $p^2 = \varphi(p) = \varphi(a)\varphi(b)$. Si ni a, ni b ne sont inversibles, on a $\varphi(a)$ et $\varphi(b)$ différents de 1. On obtient, puisque p est premier, $\varphi(a) = \varphi(b) = p$, ce qui contredit l'hypothèse. On conclut que a ou b est inversible : p est irréductible.

6. D'après les questions précédentes, les entiers premiers congrus à 3 modulo 4 et les éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a)$ soit premier sont irréductibles. Montrons que ce sont les seuls, à multiplication par un inversible près.

Soit a un élément irréductible de $\mathbb{Z}[i]$. On remarque que a divise $\varphi(a) > 1$. L'entier $\varphi(a)$ est un produit d'entiers premiers. Puisque a est irréductible, il divise (dans $\mathbb{Z}[i]$) un de ces entiers premiers, p. Il existe $b \in \mathbb{Z}[i]$ tel que $p = ab$ et donc $p^2 = \varphi(a)\varphi(b)$. Si b est inversible, a est associé à p et nécessairement p est congru à 3 modulo 4, d'après la question 5. Sinon, on a, puisque $\varphi(a) > 1$, $\varphi(a) = p$ et a est un irréductible du type étudié dans la question 4. ◁

On remarque que la condition (iii) de la question 5 de l'exercice équivaut à dire que p n'est pas somme de deux carrés. On obtient donc qu'un entier premier est somme de deux carrés si et seulement s'il n'est pas congru à 3 modulo 4. On trouvera dans le chapitre 4 (arithmétique) deux autres preuves de ce résultat (cf. 4.34 et 4.35).

L'exercice 3.13, qui étudie certaines sommes d'entiers de Gauss, suppose connus les résultats de l'exercice 3.12.

3.13. Anneau des entiers de Gauss (2)

Soit $k \in \mathbb{N}^*$ et $n \in \mathbb{N}$. On pose $s_n = \frac{1}{4} \sum_{\substack{a \in \mathbb{Z}[i] \\ a\overline{a}=n}} a^k$.

1. Montrer que $s_n \in \mathbb{Z}$ et que, si k n'est pas un multiple de 4, on a $s_n = 0$.

2. Calculer s_5, s_{13} et s_{65} pour $k = 4$. Que constate t-on ?

3. Généraliser le résultat de la question précédente.

(École normale supérieure)

▷ **Solution.**

1. Si $n = 0$, alors $s_n = 0$ pour tout k. Dans la suite, nous supposerons donc $n \neq 0$.

Comme dans l'exercice précédent, on note, pour $a \in \mathbb{Z}[i]$, $\varphi(a) = a\overline{a}$. On rappelle que deux éléments a et b de $\mathbb{Z}[i]$ sont associés s'il existe c, élément inversible de $\mathbb{Z}[i]$, tel que $b = ac$ et que la relation «être associés» est une relation d'équivalence. Il a été démontré dans l'exercice précédent, que les inversibles de $\mathbb{Z}[i]$ sont les éléments a tels que $\varphi(a) = 1$, c'est-à-dire $-1, 1, i, -i$. Si a est non nul, la classe de a contient donc quatre éléments distincts $a, -a, ia, -ia$.

On remarque que si a et b sont associés, alors $\varphi(a) = \varphi(b)$. L'ensemble des éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a) = n$ se compose donc

d'une réunion de classes d'équivalence (pour la relation «être associés») $\omega_1, \omega_2, \ldots, \omega_\ell$. On choisit, pour tout $j \in [\![1,\ell]\!]$, $a_j \in \omega_j$. On obtient alors $\sum_{a\in\omega_j} a^k = a_j^k(1^k + (-1)^k + i^k + (-i)^k)$. La valeur de la somme $u_k = 1^k + (-1)^k + i^k + (-i)^k$ ne dépend que du reste de k modulo 4. On trouve $u_0 = 4$, $u_1 = u_2 = u_3 = 0$. On en déduit que si k n'est pas divisible par 4, on a, pour tout $j \in [\![1,\ell]\!]$, $\sum_{a\in\omega_j} a^k = 0$ et donc $s_n = 0$.

Si k est divisible par 4, alors $\sum_{a\in\omega_j} a^k = 4a_j^k$, $s_n = \sum_{j=1}^{\ell} a_j^k$ et donc $s_n \in \mathbb{Z}[i]$. Par ailleurs, de l'égalité $\varphi(\overline{a}) = \varphi(a)$ pour tout $a \in \mathbb{Z}[i]$, on déduit que

$$\overline{s_n} = \frac{1}{4}\sum_{\varphi(a)=n} \overline{a}^k = \frac{1}{4}\sum_{\varphi(\overline{a})=n} \overline{a}^k = s_n.$$

Ceci montre que la partie imaginaire de s_n est nulle et donc que $s_n \in \mathbb{Z}$.

2. Pour calculer s_n, on détermine les solutions dans $\mathbb{Z}$ de $u^2+v^2 = n$. Pour les valeurs données de n, c'est aisé : on considère les entiers u tels que $u^2 \leqslant n$ et on cherche ceux pour lesquels $n - u^2$ est un carré parfait.

Pour $n = 5$, on trouve $\{|u|, |v|\} = \{1, 2\}$. Les éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a) = 5$ sont $\pm 1 \pm 2i$ et $\pm 2 \pm i$. Il y en a huit qui se répartissent en deux classes : $2 + i$ et ses associés, $2 - i$ et les siens. La question précédente montre que

$$s_5 = (2+i)^4 + (2-i)^4 = (3+4i)^2 + (3-4i)^2 = -14.$$

De même, pour $n = 13$, on trouve $\{|u|, |v|\} = \{3, 2\}$. Les éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a) = 5$ sont $\pm 3 \pm 2i$ et $\pm 2 \pm 3i$. On en déduit

$$s_{13} = (3+2i)^4 + (3-2i)^4 = (5+12i)^2 + (5-12i)^2 = -238.$$

Dans le cas $n = 65$, on trouve $\{|u|, |v|\} = \{1, 8\}$ ou $\{4, 7\}$. Il y a seize éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a) = 65$. Ils forment quatre classes, celles de $8 + i$, $8 - i$, $7 + 4i$ et $7 - 4i$. On obtient donc

$$\begin{aligned} s_{65} &= (8+i)^4 + (8-i)^4 + (7+4i)^4 + (7-4i)^4 \\ &= (63+16i)^2 + (63-16i)^2 + (33+56i)^2 + (33-56i)^2 \\ &= 2(63^2 - 16^2 + 33^2 - 56^2) = 3332. \end{aligned}$$

On remarque que $(-14)(-238) = 3332$, c'est-à-dire que

$$\boxed{s_5 s_{13} = s_{65}}.$$

Essayons d'expliquer ce résultat. D'après l'exercice précédent, $\mathbb{Z}[i]$ est un anneau factoriel et a est irréductible dès que $\varphi(a)$ est un entier premier.

Considérons $a \in \mathbb{Z}[i]$ tel que $\varphi(a) = 65$. On peut écrire

$$a\bar{a} = 65 = 13 \times 5 = 13(2+i)(2-i).$$

On en déduit que $2+i$ divise $a\bar{a}$ et donc, $2+i$ étant irréductible puisque 5 est premier, que $2+i$ divise a ou $\bar{a}$. Il existe donc $b \in \mathbb{Z}[i]$ tel que $a = (2+i)b$ ou $\bar{a} = (2+i)b$. Dans le second cas, on a $a = (2-i)\bar{b}$. Remarquons que

$$\varphi(b) = \frac{\varphi(a)}{\varphi(2+i)} = \frac{65}{5} = 13,$$

dans le premier cas et de même $\varphi(\bar{b}) = 13$, dans le second. Ainsi, nous avons démontré que tout $a \in \mathbb{Z}[i]$ tel que $\varphi(a) = 65$ s'écrit $a = bc$ avec $\varphi(b) = 13$ et $\varphi(c) = 5$. De plus, b et c sont irréductibles, car 5 et 13 sont premiers. Réciproquement, tout $a \in \mathbb{Z}[i]$ qui s'écrit $a = bc$ avec $\varphi(b) = 13$ et $\varphi(c) = 5$ vérifie $\varphi(a) = 65$. Une telle écriture de a en produits de facteurs irréductibles est unique, à multiplication par des inversibles près. On peut donc écrire a de quatre manières différentes sous la forme $a = bc$ avec $\varphi(b) = 13$ et $\varphi(c) = 5$. Si (b, c) est un couple solution, les autres sont $(-b, -c)$, $(ib, -ic)$ et $(-ib, ic)$.

De tout cela on déduit que, pour tout $k \in \mathbb{N}$,

$$s_5 s_{13} = \frac{1}{16} \sum_{\varphi(b)=5} b^k \sum_{\varphi(c)=13} c^k = \frac{1}{16} \sum_{\substack{\varphi(b)=5 \\ \varphi(c)=13}} (bc)^k = \frac{1}{16}\, 4 \sum_{\varphi(a)=65} a^k = s_{65}.$$

3. Montrons, plus généralement que, si m et n sont premiers entre eux, on a $s_{mn} = s_m s_n$. Il faut démontrer, comme précédemment, que si $a \in \mathbb{Z}[i]$ vérifie $\varphi(a) = mn$, alors il existe $(b, c) \in (\mathbb{Z}[i])^2$ tel que $a = bc$, $\varphi(b) = m$ et $\varphi(c) = n$ et que de plus, à a correspondent quatre telles décompositions $a = bc$, b pouvant être multiplié par un inversible quelconque. Le résultat s'obtient à partir de la décomposition de a en facteurs irréductibles.

D'après l'exercice précédent les irréductibles de $\mathbb{Z}[i]$ sont les entiers premiers congrus à 3 (mod 4) et les éléments a tels que $\varphi(a)$ soit un entier premier (non congru à 3 (mod 4)). La décomposition de a en facteurs irréductibles peut s'écrire

$$a = \varepsilon \prod_{i=1}^{r} p_i^{\alpha_i} \prod_{j=1}^{s} a_j^{\beta_j},$$

où ε est inversible, r et s sont des entiers naturels, $p_1, \ldots, p_r$ des entiers premiers congrus à 3 (mod 4), $a_1, \ldots, a_s$ sont des éléments de $\mathbb{Z}[i]$ tels que $\varphi(a_1), \ldots, \varphi(a_s)$ soient des entiers premiers, non congrus à 3 (mod 4), $\alpha_1, \ldots, \alpha_r$, $\beta_1, \ldots, \beta_s$ des entiers naturels non nuls. On en déduit que

$$mn = \varphi(a) = \prod_{i=1}^{r} p_i^{2\alpha_i} \prod_{j=1}^{s} \varphi(a_j)^{\beta_j}.$$

On observe que cette écriture est la décomposition de mn en facteurs premiers. Les entiers m et n étant premiers entre eux, ils n'ont aucun diviseur premier commun. Quitte à renuméroter les irréductibles qui divisent a, on peut supposer qu'il existe des entiers naturels r' et s' tels que $r' \leqslant r$, $s' \leqslant s$ et

$$m = \prod_{i=1}^{r'} p_i^{2\alpha_i} \prod_{j=1}^{s'} \varphi(a_j)^{\beta_j} \quad \text{et} \quad n = \prod_{i=r'+1}^{r} p_i^{2\alpha_i} \prod_{j=s'+1}^{s} \varphi(a_j)^{\beta_j}.$$

Si b et c sont deux éléments de $\mathbb{Z}[i]$ tels que $a = bc$, $m = \varphi(b)$ et $n = \varphi(c)$, il résulte du calcul de $\varphi(a)$ à partir de a et de l'expression de m et n, que les irréductibles qui divisent b sont $p_1, \ldots, p_{r'}$ et $a_1, \ldots, a_{s'}$ et que ceux qui divisent c sont $p_{r'+1}, \ldots, p_r$ et $a_{s'+1}, \ldots, a_s$. Puisque $a = bc$, il existe ε' inversible tel que

$$b = \varepsilon' \prod_{i=1}^{r'} p_i^{\alpha_i} \prod_{j=1}^{s'} a_j^{\beta_j} \quad \text{et} \quad c = \varepsilon(\varepsilon')^{-1} \prod_{i=r'+1}^{r} p_i^{\alpha_i} \prod_{j=s'+1}^{s} a_j^{\beta_j}.$$

Il est clair que b et c ainsi définis ont toutes les propriétés voulues.

On obtient donc, pour tout a tel que $\varphi(a) = mn$, quatre décompositions sous la forme $a = bc$, avec $m = \varphi(b)$ et $n = \varphi(c)$, correspondant aux quatre valeurs de ε'. De la même manière que dans la question précédente, on conclut que $s_m s_n = s_{mn}$. ◁

Une écriture d'un entier n sous la forme $n = \varphi(a)$ correspond (bijectivement) à une décomposition de n en somme de deux carrés $n = u^2 + v^2$, où u et v sont deux entiers tels que $a = u + iv$. Pour tout entier $n \geqslant 1$, notons $r(n)$ le nombre de couples $(u, v) \in \mathbb{Z}^2$ tels que $n = u^2 + v^2$, c'est-à-dire le nombre d'éléments a de $\mathbb{Z}[i]$ tels que $\varphi(a) = n$. La dernière question de l'exercice montre que si m et n sont premiers entre eux, on a $r(m)r(n) = 4r(mn)$. Si on pose $\delta(n) = \dfrac{r(n)}{4}$, on obtient $\delta(mn) = \delta(m)\delta(n)$: la fonction δ est donc multiplicative. Pour quelques remarques sur les fonctions arithmétiques multiplicatives, on se reportera aux exercices 4.28 à 4.33 du chapitre 4 (arithmétique).

L'exercice suivant étudie un anneau dont les éléments sont de la forme $\mathrm{P} + \varepsilon\mathrm{Q}$, où ε vérifie $\varepsilon^2 = \mathrm{X}^2 - 1$. On rapprochera cet anneau de l'anneau $\mathbb{Z}[i]$, obtenu de même à partir de $\mathbb{Z}$ en rajoutant un élément de carré -1 (cf. exercice 3.12). En particulier l'application N qu'on y définit joue le rôle de l'application φ de $\mathbb{Z}[i]$ dans la détermination des éléments inversibles.

3.14. Une extension quadratique de $\mathbb{C}[X]$

On considère le sous-ensemble A de $\mathcal{F}(]1,+\infty[,\mathbb{C})$ défini par

$$A = \left\{ x \longmapsto P(x) + Q(x)\sqrt{x^2-1},\ \ (P,Q) \in \mathbb{C}[X]^2 \right\}$$

1. Montrer que A est un anneau. Vérifier que tout $z \in A$ s'écrit de manière unique sous la forme $z = P + \varepsilon Q$ avec P et Q dans $\mathbb{C}[X]$, où ε désigne l'application $x \longmapsto \sqrt{x^2-1}$.

2. Montrer que $z = P + \varepsilon Q$ est une unité de A si, et seulement si, $N(z) = P^2 - (X^2-1)Q^2 \in \mathbb{C}^*$.

3. Montrer qu'il existe $z_0 \in A$ tel que

$$U_0 = \{z \in A,\ N(z) = 1\} = \{\pm z_0^n,\ n \in \mathbb{Z}\}.$$

4. Expliciter les couples de polynômes $(P,Q) \in \mathbb{C}[X]^2$ qui vérifient $P^2 - (X^2-1)Q^2 = 1$.

(École polytechnique)

▷ **Solution.**

1. Il est clair que A contient l'application nulle ainsi que l'application constante égale à 1, et qu'il est stable pour la somme et l'opposé. Soit $(P_1,Q_1,P_2,Q_2) \in \mathbb{C}[X]^4$. On a

$$(P_1 + \varepsilon Q_1)(P_2 + \varepsilon Q_2) = P_1P_2 + \varepsilon^2 Q_1Q_2 + \varepsilon(P_1Q_2 + P_2Q_1) \in A,$$

car ε^2 est la fonction polynôme $x \longmapsto x^2 - 1$. Il en résulte que A est un sous-anneau de l'anneau $\mathcal{F}(]1,+\infty[,\mathbb{C})$.

L'application $\psi : \mathbb{C}[X]^2 \to A$ qui au couple (P,Q) associe $z = P + \varepsilon Q$ est clairement un morphisme de groupes additifs. Il est surjectif par définition de A. Enfin, il est injectif car si $P \in \operatorname{Ker}\psi$ *i.e.* $P + \varepsilon Q = 0$, on obtient $P^2 = (X^2-1)Q^2$, ce qui impose $P = Q = 0$ (car sinon la multiplicité de la racine 1 est paire dans le membre de gauche et impaire dans celui de droite). Il en résulte que tout $z \in A$ s'écrit de manière unique sous la forme $z = P + \varepsilon Q$. En particulier on peut identifier $\mathbb{C}[X]$ à un sous-anneau de A, ce qu'on fera dans la suite.

2. Si $z = P + \varepsilon Q$ est un élément de A, on pose $\overline{z} = P - \varepsilon Q$ et on parle du conjugué de z. Il est clair que l'application $z \longmapsto \overline{z}$ est un automorphisme d'anneau de A.

On a $N(z) = z\overline{z} = P^2 - (X^2-1)Q^2 \in \mathbb{C}[X]$.

• Soit $z \in A$ inversible et $z' \in A$ tel que $zz' = 1$. On a alors

$$1 = N(1) = N(zz') = zz'\overline{zz'} = z\overline{z}z'\overline{z'} = N(z)N(z'),$$

d'où il résulte que $N(z)$ est inversible dans $\mathbb{C}[X]$ et appartient donc à $\mathbb{C}^*$.

• Réciproquement, si $N(z) = \lambda \in \mathbb{C}^*$, alors $\frac{1}{\lambda}\overline{z} \in A$ est l'inverse de z dans A.

3. Notons U le groupe des unités de A. Comme N est un morphisme de groupes de U dans $\mathbb{C}^*$, $U_0 = \operatorname{Ker} N$ est un sous-groupe de U.

Posons $z_0 = X + \varepsilon$. On a $N(z_0) = X^2 - (X^2 - 1) = 1$, donc z_0 est dans U_0. On va montrer que z_0 répond à la question. Il est déjà clair que $\{\pm z_0^n,\ n \in \mathbb{Z}\} \subset U_0$. Soit $z = P + \varepsilon Q \in U_0$. On considère un entier $m \in \mathbb{Z}$ tel que $z' = z_0^{-m} z$ s'écrive $z' = A + \varepsilon B$, avec un polynôme B de degré minimal. Alors z' appartient à U_0, car U_0 est un groupe multiplicatif. On va montrer que $B = 0$. Il s'ensuivra que $A = \pm 1$, puisque $z' \in U_0$, soit encore $z = \pm z_0^m$.

Supposons par l'absurde que $B \neq 0$. On a $A^2 - (X^2 - 1)B^2 = 1$, puisque $z' \in U_0$, ce qui peut s'écrire $(A - XB)(A + XB) = 1 - B^2$. On doit nécessairement avoir $\deg A = 1 + \deg B$. Il en résulte que l'un des deux polynômes $A - XB$ ou $A + XB$ a un degré strictement inférieur à $\deg B$ (car sinon $\deg(A - XB) = \deg(A + XB) = \deg B$, d'où l'on déduit $\deg A = \deg(2A) \leqslant \deg B$, ce qui est faux). Or on a

$$z_0 z' = (X + \varepsilon)(A + \varepsilon B) = XA + (X^2 - 1)B + \varepsilon(XB + A)$$
$$z_0^{-1} z' = (X - \varepsilon)(A + \varepsilon B) = XA - (X^2 - 1)B + \varepsilon(XB - A)$$

et l'une des deux égalités est contredit la minimalité du degré de B.

4. Posons pour $n \in \mathbb{N}$, $z_0^n = (X + \varepsilon)^n = P_n + \varepsilon Q_n$. On obtient alors $z_0^{-n} = \overline{z_0}^n = P_n - \varepsilon Q_n$. D'après la question précédente, les solutions de l'équation $P^2 - (X^2 - 1)Q^2 = 1$ sont donc les couples $(\pm P_n, \pm Q_n)$, où n décrit $\mathbb{N}$. À l'aide de la formule du binôme on obtient

$$P_n = \sum_{0 \leqslant 2k \leqslant n} C_n^{2k} (X^2 - 1)^k X^{n-2k}$$
$$Q_n = \sum_{0 \leqslant 2k+1 \leqslant n} C_n^{2k+1} (X^2 - 1)^k X^{n-2k-1}.$$

On peut reconnaître ici les polynômes de Tchebychev de première et de seconde espèce T_n et U_{n-1} qui sont définis par $\cos n\theta = T_n(\cos\theta)$ et $\sin n\theta = \sin\theta U_{n-1}(\cos\theta)$. Cela s'explique très bien. Soit f l'application de A dans l'anneau des polynômes trigonométriques complexes qui a $z = P + \varepsilon Q$ associe, $P(\cos\theta) + i \sin\theta Q(\cos\theta)$. On vérifie facilement que f est un morphisme d'anneaux et qu'il est injectif. On en déduit que $f(z_0^n) = P_n(\cos\theta) + i \sin\theta Q_n(\cos\theta)$ et, grâce à la formule de Moivre,

$$\begin{aligned} f(z_0^n) = f(z_0)^n &= (\cos\theta + i\sin\theta)^n = \cos n\theta + i \sin n\theta \\ &= T_n(\cos\theta) + i \sin\theta U_{n-1}(\cos\theta). \end{aligned}$$

On obtient donc donc $P_n = T_n$ et $Q_n = U_{n-1}$. ◁

Une autre démonstration, élémentaire, de la dernière question est donnée dans l'exercice 5.7.

Les trois exercices suivants étudient des exemples assez simples de corps. Le premier fait la transition avec les anneaux.

3.15. Anneau sans idéal non premier

Soit A un anneau commutatif dont tout idéal I est premier c'est-à-dire vérifie, pour tout $(x, y) \in A^2$,

$$xy \in I \Longrightarrow x \in I \text{ ou } y \in I.$$

Montrer que A est un corps.

(École normale supérieure)

▷ **Solution.**

Dire que l'idéal nul est premier signifie que A est intègre. Soit x dans A, non nul. On souhaite prouver que x est inversible. Comme l'idéal principal (x^2) est premier, on a $x \in (x^2)$. Il existe donc $a \in A$ tel que $x = ax^2$. Comme A est intègre et x non nul, on a $ax = 1$ et x est inversible. On conclut que A est un corps. ◁

Les corps étudiés maintenant sont des corps de nombres algébriques, dont on peut donner la définition suivante : soit θ un élément de $\mathbb{C}$, racine d'un polynôme non nul P de $\mathbb{Q}[X]$ qui est irréductible sur $\mathbb{Q}$. Alors l'ensemble des nombres complexes de la forme $\Phi(\theta)$, où $\Phi \in \mathbb{Q}[X]$, est un corps noté $\mathbb{Q}(\theta)$, isomorphe au quotient $\mathbb{Q}[X]/P\mathbb{Q}[X]$. De plus $\mathbb{Q}(\theta)$ est un espace vectoriel de dimension $n = \deg P$ sur $\mathbb{Q}$. Pour un exemple plus complexe, le lecteur se reportera à l'exercice 5.17 qui étudie $\mathbb{Q}(e^{i\frac{2\pi}{p}})$, où p est premier.

3.16. Automorphismes de $\mathbb{Q}(\sqrt{2})$

Soit $E = \{a + b\sqrt{2} \in \mathbb{C}^2, (a, b) \in \mathbb{Q}^2\}$. Montrer que E est un sous-corps de $\mathbb{C}$ et en déterminer tous les automorphismes.

(École polytechnique)

▷ **Solution.**

• $\mathbb{C}$ est une $\mathbb{Q}$-algèbre. L'application

$$\varphi : \begin{array}{ccc} \mathbb{Q}[X] & \longrightarrow & \mathbb{C} \\ P & \longmapsto & P(\sqrt{2}) \end{array}$$

est un morphisme d'algèbres. Son image est donc une sous-algèbre de $\mathbb{C}$, qui contient clairement E.

Pour montrer l'inclusion inverse, on observe que, si $n \geqslant 0$, on a $\varphi(X^n) = (\sqrt{2})^n = 2^p(\sqrt{2})^r \in E$, où $n = 2p+r$, avec $r = 0$ ou 1 et $p \in \mathbb{N}$. On en déduit que $\operatorname{Im}\varphi = E$. Par conséquent, E est un sous-anneau de $\mathbb{C}$.

• Soit $x \in E$ non nul, qui s'écrit $x = a + b\sqrt{2}$ avec $(a,b) \in \mathbb{Q}^2$. On a alors $a - b\sqrt{2} \neq 0$ (car $\sqrt{2} \notin \mathbb{Q}$) et donc

$$x^{-1} = \frac{1}{a+b\sqrt{2}} = \frac{a-b\sqrt{2}}{a^2-2b^2} = \frac{a}{a^2-2b^2} + \frac{-b}{a^2-2b^2}\sqrt{2} \in E.$$

On en déduit que E est un sous-corps de $\mathbb{C}$.

• Si f un automorphisme de E, alors f laisse invariants tous les éléments de $\mathbb{Q}$. En effet, on a, pour $x \in \mathbb{Q}$ et $n \in \mathbb{N}$,

$$f(nx) = f(x + \cdots + x) = f(x) + \cdots + f(x) = nf(x) = nf(x),$$

et donc $f(n) = n$, car $f(1) = 1$. On a ensuite $f(-n) + f(n) = f(0) = 0$ et donc $f(-n) = -f(n) = -n$. Enfin si $q \in \mathbb{N}^*$ et $p \in \mathbb{Z}$, on obtient

$$qf\left(\frac{p}{q}\right) = f\left(q\frac{p}{q}\right) = f(p) = p \quad \text{et} \quad f\left(\frac{p}{q}\right) = \frac{p}{q}.$$

On détermine ensuite $f(\sqrt{2})$. De $f(\sqrt{2})^2 = f(\sqrt{2}^2) = f(2) = 2$ résulte $f(\sqrt{2}) = \varepsilon\sqrt{2}$ avec $\varepsilon = \pm 1$. On a donc, pour tout $(a,b) \in \mathbb{Q}^2$,

$$f(a+b\sqrt{2}) = f(a) + f(b)f(\sqrt{2}) = a + \varepsilon b\sqrt{2}.$$

Réciproquement, les deux applications ainsi définies (pour $\varepsilon = -1$ et 1) sont des automorphismes de corps de E. Il n'existe qu'un seul automorphisme de E qui ne soit pas l'identité. ◁

L'exercice suivant est l'occasion de découvrir les propriétés fondamentales de la théorie algébrique des corps et notamment la formule de multiplicativité des degrés.

3.17. Nombres algébriques

On considère un corps commutatif M et $K \subset L$ deux sous-corps de M.

1. Démontrer l'équivalence des deux conditions suivantes :

(*i*) M est un K-espace vectoriel de dimension finie m ;

(*ii*) M est un L-espace vectoriel de dimension finie n et L un K-espace vectoriel de dimension finie p.

Vérifier que dans ces conditions $m = np$.

2. On suppose que $a \in \mathrm{M}$ est *algébrique sur* K : il existe P non nul dans $\mathbb{K}[\mathrm{X}]$ tel que $\mathrm{P}(a) = 0$. Démontrer l'existence d'un unique polynôme unitaire μ de $\mathrm{K}[\mathrm{X}]$ tel que, pour tout polynôme P de $\mathrm{K}[\mathrm{X}]$,

$$\mathrm{P}(a) = 0 \iff \mu | \mathrm{P}.$$

Vérifier que μ est irréductible (μ est appelé le *polynôme minimal de a sur* K) et que l'ensemble $\mathrm{K}(a) = \{\mathrm{Q}(a),\ \mathrm{Q} \in \mathrm{K}[\mathrm{X}]\}$ est à la fois un K-espace vectoriel de dimension $\deg \mu$ et un sous-corps de M.

3. On dit que $a \in \mathbb{C}$ est *algébrique* si a est algébrique sur le corps $\mathbb{Q}$. Démontrer que $a = \sqrt[3]{2} + \sqrt[3]{2}$ est algébrique et déterminer son polynôme minimal.

(**École Polytechnique**)

▷ **Solution.**

1. Supposons (ii). On se donne $(e_1, \ldots, e_p)$ une base du K-espace vectoriel L et $(f_1, \ldots, f_n)$ une base du L-espace vectoriel M. Montrons que la famille $(e_j f_i)_{(i,j) \in [\![1,n]\!] \times [\![1,p]\!]}$ est une base du K-espace vectoriel M.

Soit $x \in \mathrm{M}$. On écrit $x = \lambda_1 f_1 + \cdots + \lambda_n f_n$ avec $\lambda_i \in \mathrm{L}$. Chaque λ_i s'écrit $\lambda_i = \sum_{j=1}^{p} \mu_{i,j} e_j$ avec $\mu_{i,j} \in \mathrm{K}$. Par associativité, on obtient

$$x = \sum_{(i,j) \in [\![1,n]\!] \times [\![1,p]\!]} \mu_{i,j} e_j f_i,$$

ce qui prouve que la famille $(e_j f_i)_{(i,j) \in [\![1,n]\!] \times [\![1,p]\!]}$ est génératrice du K-espace vectoriel M.

Vérifions qu'elle est libre : soit $\mu_{i,j} \in \mathrm{K}$ pour $(i,j) \in [\![1,n]\!] \times [\![1,p]\!]$ tels que

$$0 = \sum_{(i,j) \in [\![1,n]\!] \times [\![1,p]\!]} \mu_{i,j} e_j f_i.$$

Par associativité, on peut écrire

$$0 = \sum_{i=1}^{n} \underbrace{\left(\sum_{j=1}^{p} \mu_{i,j} e_j \right)}_{\in \mathrm{L}} f_i$$

ce qui impose que pour tout i, $\sum_{j=1}^{p} \mu_{i,j} e_j = 0$ puisque $(f_1, \ldots, f_n)$ est libre dans le L-espace M. Comme $(e_1, \ldots, e_p)$ est libre dans le K-espace L, il vient que tous les $\mu_{i,j}$ sont nuls. On conclut que la famille $(e_j f_i)_{(i,j) \in [\![1,n[\![\times [\![1,p]\!]}$ est une base du K-espace vectoriel M et que $m = np$.

Supposons (i). Comme L est un sous-espace du K-espace vectoriel M, L est de dimension finie sur K. D'autre part, si $(\varepsilon_1, \ldots, \varepsilon_m)$ constitue une base du K-espace M, c'est clairement un système générateur de M en tant que L-espace vectoriel. Donc M est un L-espace vectoriel de dimension finie.

On dit dans ces conditions que L *est une extension de* K *et* M *est une extension de* L. *La dimension de* L *comme* K*-espace vectoriel est notée* $[\mathrm{L} : \mathrm{K}]$ *: c'est le degré de l'extension* L *sur le corps* K. *Nous venons de démontrer que*

$$[\mathrm{K} : \mathrm{M}] = [\mathrm{K} : \mathrm{L}][\mathrm{L} : \mathrm{M}].$$

Cette identité est la relation de multiplicativité des degrés.

2. Le corps M est une K-algèbre. Considérons le morphisme de K-algèbres

$$\begin{array}{rccc} & \mathrm{K}[\mathrm{X}] & \longrightarrow & \mathrm{M} \\ \Phi : & \mathrm{P} & \longmapsto & \mathrm{P}(a) \end{array}$$

L'ensemble $\mathrm{K}(a)$, qui est l'image de Φ, est une sous-algèbre de M. Le noyau $\operatorname{Ker}\Phi = \{\mathrm{P} \in \mathbb{Q}[\mathrm{X}],\ \mathrm{P}(a) = 0\}$ forme un idéal de K[X]. Par hypothèse, a est algébrique donc $\mathrm{I} \neq \{0\}$. Dans ces conditions, comme I est un idéal principal, on sait qu'il existe un unique polynôme μ unitaire de degré supérieur ou égal à 1 dans K[X] tel que $\mathrm{I} = \mu\mathrm{K}[\mathrm{X}]$. On a l'équivalence

$$\forall \mathrm{P} \in \mathrm{K}[\mathrm{X}],\ \ \mu | \mathrm{P} \iff \mathrm{P}(a) = 0.$$

Si le polynôme μ' unitaire de K[X] vérifie la même équivalence, alors $\mu' | \mu$ et $\mu | \mu'$ donc $\mu = \mu'$: μ est unique.

Montrons que le polynôme μ est irréductible. Si $\mu = \mathrm{AB}$ avec A et B dans K[X], on a

$$0 = \mu(a) = \mathrm{A}(a)\mathrm{B}(a)$$

et donc $\mathrm{A}(a) = 0$ ou $\mathrm{B}(a) = 0$, ce qui implique $\mu | \mathrm{A}$ ou $\mu | \mathrm{B}$ et donc A ou B constant : le polynôme μ est irréductible. Remarquons que si P est irréductible unitaire avec $\mathrm{P}(a) = 0$, alors $\mathrm{P} = \mu$ puisque $\mu | \mathrm{P}$.

Montrons ensuite que $\mathrm{K}(a)$ est de dimension $r = \deg \mu$ comme K-espace vectoriel. Pour cela, vérifions que $(1, a, \ldots, a^{r-1})$ est une base de $\mathrm{K}(a)$. Le fait que μ soit un polynôme non nul de degré minimal annulant a garantit la liberté de la famille $(1, a, \ldots, a^{r-1})$. D'autre part, si $x \in \mathrm{K}(a)$, il s'écrit $x = \mathrm{A}(a)$ avec $\mathrm{A} \in \mathrm{K}[\mathrm{X}]$ et en notant R le reste de la division euclidienne de A par μ, on obtient $x = \mathrm{R}(a) \in \operatorname{Vect}(1, a, \ldots, a^{r-1})$.

Montrons enfin que $\mathrm{K}(a)$ est un sous-corps de M : c'est déjà un sous-anneau de M. Si x est non nul dans $\mathrm{K}(a)$, il s'écrit $x = \mathrm{A}(a)$ avec A non divisible par μ. Dans ces conditions, μ étant irréductible, il est premier avec A. D'après le théorème de Bezout, il existe U et V dans

$K[X]$ tels que $1 = UA + V\mu$ et en évaluant en a, il vient $1 = U(a)x$ et donc $x^{-1} = U(a) \in K(a)$.

3. Il s'agit de trouver un polynôme non nul de $\mathbb{Q}[X]$ qui admet a comme racine. On a $(a - \sqrt{2})^3 = 2$ ce qui donne en développant

$$a^3 + 6a - 2 = \sqrt{2}(3a^2 + 2),$$

et en élevant au carré

$$a^6 - 6a^4 - 4a^3 + 12a^2 - 24a - 4 = 0.$$

Le polynôme $P = X^6 - 6X^4 - 4X^3 + 12X^2 - 24X - 4 \in \mathbb{Q}[X]$ annule donc a, qui est un nombre algébrique.

Notons μ le polynôme minimal de a ; c'est un diviseur de P. On va prouver que $P = \mu$ en considérant la dimension du $\mathbb{Q}$-espace vectoriel $\mathbb{Q}(a)$. En fait, la relation $a^3 + 6a - 2 = \sqrt{2}(3a^2 + 2)$ prouve que $\sqrt{2}$ est dans $\mathbb{Q}(a)$ puisque $3a^2 + 2 > 0$, $\sqrt{2} = (a^3 + 6a - 2)(3a^2 + 2)^{-1}$ et $\mathbb{Q}(a)$ est un corps. On a également $\sqrt[3]{2} = a - \sqrt{2} \in \mathbb{Q}(a)$. Comme $\mathbb{Q}(a)$ est une $\mathbb{Q}$-algèbre, pour tout $Q \in K[X]$, $Q(\sqrt{2})$ et $Q(\sqrt[3]{2})$ sont dans $Q(a)$. Il s'ensuit les inclusions

$$\mathbb{Q} \subset \mathbb{Q}(\sqrt{2}) \subset \mathbb{Q}(a) \quad \text{et} \quad \mathbb{Q} \subset \mathbb{Q}(\sqrt[3]{2}) \subset \mathbb{Q}(a).$$

Or le polynôme minimal de $\sqrt{2}$ sur $\mathbb{Q}$ est $X^2 - 2$ et celui de $\sqrt[3]{2}$ est $X^3 - 2$ (ils sont bien irréductibles car de degré 2 ou 3 sans racine rationnelle). On en déduit que la dimension de $\mathbb{Q}(\sqrt{2})$ (resp. $\mathbb{Q}(\sqrt[3]{2})$) sur $\mathbb{Q}$ est 2 (resp. 3). D'après la première question et la multiplicativité des degrés des extensions, 2 et 3 divisent la dimension de $\mathbb{Q}(a)$ sur $\mathbb{Q}$. Donc, cette dimension est supérieure ou égale à 6. On a donc $\deg \mu \geqslant 6$. Comme μ divise P, ces polynômes unitaires sont égaux :

$$\boxed{\mu = X^6 - 6X^4 - 4X^3 + 12X^2 - 24X - 4}. \lhd$$

L'exercice suivant détermine toutes les valuations sur $\mathbb{Q}$.

3.18. Valuations sur $\mathbb{Q}$

On appelle valuation sur un anneau A toute application ν de A dans $\mathbb{R} \cup \{+\infty\}$ telle que, pour tout $(x, y) \in A^2$:

(i) $\nu(xy) = \nu(x) + \nu(y)$;
(ii) $\nu(x+y) \geqslant \min(\nu(x), \nu(y))$;
(iii) $\nu(x) = +\infty \iff x = 0$.

1. Donner des exemples de valuation sur $\mathbb{Z}$, sur $\mathbb{Q}$.

2. Déterminer toutes les valuations sur $\mathbb{Q}$.

(**École normale supérieure**)

Solution.

1. Le cours fournit l'exemple des valuations p-adiques sur $\mathbb{Z}$: pour tout nombre premier p et tout entier non nul n, $\nu_p(n)$ est le plus grand entier naturel α tel que p^α divise n. On pose $\nu_p(0) = +\infty$ et il est aisé de vérifier les propriétés (i) et (ii). Ces valuations se prolongent à $\mathbb{Q}$: si x est un nombre rationnel non nul admettant comme représentant $\dfrac{a}{b}$, on pose $\nu_p(x) = \nu_p(a) - \nu_p(b)$; $\nu_p(x)$ ne dépend que de x et pas du représentant choisi et ν_p est encore une valuation sur $\mathbb{Q}$.

On va montrer dans la seconde question que les seules valuations sur $\mathbb{Q}$ sont la valuation triviale (*i.e.* nulle sur $\mathbb{Q}^*$) et les applications $\lambda\nu_p$, où λ est un réel strictement positif arbitraire et p un nombre premier.

2. Soit ν une valuation sur $\mathbb{Q}$, que l'on suppose non triviale.

• On a $\nu(1) = \nu(1^2) = \nu(1) + \nu(1)$, de sorte que $\nu(1) = 0$, mais aussi $2\nu(-1) = \nu(1) = 0$ et donc $\nu(-1) = 0$. On en déduit que, pour tout $x \in \mathbb{Q}^*$, $\nu(-x) = \nu(x) + \nu(-1) = \nu(x)$: ν est paire. On a par la propriété (ii), $\nu(2) \geqslant 0$ et par une récurrence évidente, $\nu(n) \geqslant 0$ pour tout entier $n \in \mathbb{N}$. Observons enfin que $\nu(x^n) = n\nu(x)$ pour tout $n \in \mathbb{N}$ et tout $x \in \mathbb{Q}$.

• L'ensemble $\mathrm{E} = \{n \in \mathbb{N}^*,\ \nu(n) > 0\}$ est non vide, car sinon ν serait triviale. Soit p le plus petit élément de E. Nécessairement, p est premier : en effet, si $p = ab$ avec $1 < a, b < p$, $\nu(p) = \nu(a) + \nu(b) = 0 + 0 = 0$, d'après la propriété (i), ce qui est absurde.

Soit q un nombre premier distinct de p. Comme $\gcd(p, q) = 1$, on peut trouver $(u, v) \in \mathbb{Z}^2$ tel que $up + vq = 1$ par le théorème de Bezout. Par (ii) on obtient

$$0 \geqslant \min(\nu(up), \nu(vq)) = \min(\nu(u)+\nu(p), \nu(v)+\nu(q)) \geqslant \min(\nu(p), \nu(q)).$$

On ne peut donc pas avoir $\nu(q) > 0$ et donc $\nu(q) = 0$.

Il résulte alors du théorème de décomposition en facteurs premiers et du point (i) que, pour tout $n \in \mathbb{Z}^*$, $\nu(n) = \lambda\nu_p(n)$ où $\lambda = \nu(p) > 0$. Et toujours en vertu de (i) (qui implique que $\nu\left(\dfrac{x}{y}\right) = \nu(x) - \nu(y)$, pour tout $(x, y) \in \mathbb{Q} \times \mathbb{Q}^*$), cette égalité s'étend à $\mathbb{Q}$ tout entier. ◁

Les valuations sur $\mathbb{Q}$ résultent toutes d'un prolongement de celles de $\mathbb{Z}$. C'est vrai pour le corps des fractions d'un anneau intègre quelconque.

Chaque valuation de $\mathbb{Z}$ ou $\mathbb{Q}$, ν_p est associée à un nombre premier p. Il y a correspondance entre irréductibles et valuations prenant des valeurs positives ou nulles sur l'anneau A, *dans tous les anneaux* A *principaux et même dans des anneaux plus généraux, les anneaux de Dedekind (anneaux intègres dans lesquels tout idéal est produit d'idéaux premiers), à condition de remplacer les nombres premiers par les idéaux premiers.*

L'exercice suivant détermine les valeurs absolues non-archimédiennes sur $\mathbb{C}(X)$. Une valeur absolue sur un corps K *de caractéristique nulle est non-archimédienne si $\mathbb{N}^*$, considéré comme sous-anneau de* K, *est borné par* 1. *Elles s'opposent donc à la valeur absolue usuelle sur $\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$, que nous savons être archimédienne.*

3.19. Valeurs absolues non-archimédiennes sur $\mathbb{C}(X)$

Soit K un corps commutatif de caractéristique nulle. On appelle valeur absolue de K toute application $|\ |\ : K \longrightarrow \mathbb{R}_+$ telle que, pour tout $(x, y) \in K^2$:

(*i*) $|x| = 0 \Longleftrightarrow x = 0$;

(*ii*) $|xy| = |x|\ |y|$;

(*iii*) $|x + y| \leqslant |x| + |y|$.

1. Soit $|\ |$ une valeur absolue de K telle que $|n| \leqslant 1$ pour $n \in \mathbb{N}^*$. Montrer que $|\ |$ vérifie l'inégalité *ultra-métrique* : pour tout couple $(x, y) \in K^2$,

$$|x + y| \leqslant \max(|x|, |y|).$$

2. Soit ν une valeur absolue de $\mathbb{C}(X)$ telle que $\nu(x) = 1$ pour tout $x \in \mathbb{C}^*$. Montrer que ν est de l'un des types suivants :

(*i*) $\nu(F) = 1$ pour toute fraction F non nulle ;

(*ii*) $\nu(F) = a^{\deg F}$, avec $a > 1$, pour F non nulle ;

(*iii*) $\nu(F) = a^{\mathrm{val}(F)}$, avec $0 < a < 1$, pour F non nulle, où l'on précisera ce qu'est val(F).

(**École normale supérieure**)

▷ **Solution.**

1. Ici, n désigne $n \cdot 1$, somme de n termes égaux à l'élément neutre pour la multiplication de K. Soit $(x, y) \in K^2$. Supposons, par exemple, que $|y| \leqslant |x|$. Pour tout entier naturel n, on a le développement du binôme $(x + y)^n = \sum_{k=0}^{n} C_n^k x^k y^{n-k}$. En prenant la valeur absolue, il vient

$$|x+y|^n \leqslant \sum_{k=0}^{n} |C_n^k x^k y^{n-k}| \leqslant \sum_{k=0}^{n} |C_n^k \cdot 1||x^k y^{n-k}| \leqslant \sum_{k=0}^{n} |x|^k |y|^{n-k}$$
$$\leqslant \sum_{k=0}^{n} |x|^n = (n+1)|x|^n.$$

Il en résulte que pour tout entier n, $|x+y| \leqslant (n+1)^{1/n}|x|$. Le résultat suit, en faisant tendre n vers l'infini.

Une valeur absolue telle que $\mathbb{N}^*$ soit bornée par 1 est dite non-archimédienne. *Elle vérifie l'inégalité ultra-métrique, qui est plus forte que l'inégalité triangulaire (iii). Observons que la réciproque est vraie : (i) et (ii) impliquent $|1| = 1$, puis $|-1| = 1$; on en déduit que, pour toute valeur absolue qui vérifie l'inégalité ultra-métrique, $\mathbb{Z}$ est borné par 1.*

2. La valeur absolue ν est non-archimédienne et donc vérifie l'inégalité ultra-métrique.

• Observons pour commencer que, pour tout $(F, G) \in (\mathbb{C}(X))^2$ tel que $\nu(F) < \nu(G)$, on a $\nu(F+G) = \nu(G)$. En effet, on sait, d'après l'inégalité ultra-métrique, que $\nu(F+G) \leqslant \nu(G)$, mais comme $G = (F+G)+(-F)$, on a aussi $\nu(G) \leqslant \max(\nu(F), \nu(F+G)) = \nu(F+G)$, d'où l'égalité.

Posons $a = \nu(X)$.

• Supposons que $a > 1$. Tout polynôme non nul P de $\mathbb{C}[X]$ s'écrit $P = \sum_{k=0}^{n} a_k X^k$, avec $n = \deg(P)$. Comme pour tout k, $\nu(a_k X^k) = a^k$, on a

$$\nu(a_0) = 1 < \nu(a_1 X) < \nu(a_2 X^2) < \cdots < \nu(a_n X^n) = a^n.$$

Ceci permet d'affirmer que $\nu(P) = \nu(a_n X^n) = a^n = a^{\deg(P)}$. Grâce à la propriété (ii) de la valeur absolue, on conclut que, pour toute fraction rationnelle non nulle $F = \dfrac{P}{Q}$, on a

$$\nu(F) = \nu(P) - \nu(Q) = \frac{a^{\deg(P)}}{a^{\deg(G)}} = a^{\deg(F)}.$$

• Le cas $a < 1$ se traite de la même façon. On a $\nu(a_k X^k) = a^k$ si $a_k \neq 0$ et on trouve cette fois, $\nu(P) = a^{\mathrm{val}(P)}$ pour tout polynôme non nul P, où val(P) est la valuation en X de P, c'est-à-dire le plus petit indice k tel que $a_k \neq 0$. Si $F = \dfrac{P}{Q}$ est une fraction non nulle, la quantité $\mathrm{val}(P) - \mathrm{val}(Q)$ ne dépend que de F et pas du représentant choisi. Cette quantité étant notée val(F), on a $\nu(F) = a^{\mathrm{val}\,F}$.

• Supposons enfin $a = 1$. Dans ce cas, on a $\nu(X^k) = 1$, pour tout k et $\nu(P) \leqslant 1$, pour tout polynôme P. Distinguons deux cas.

⋆ Si $\nu(P) = 1$ pour tout polynôme non nul P, alors ν est triviale : elle vaut 1 sur toutes les fractions non nulles.

⋆ Dans le cas contraire, notons Q un polynôme de degré minimal tel que $\nu(Q) < 1$. Ce polynôme Q est irréductible car, si $Q = AB$ avec $\deg A < \deg Q$ et $\deg B < \deg Q$, on a $\nu(A) = \nu(B) = 1$ et donc $\nu(Q) = 1$, ce qui est absurde. On peut bien entendu choisir Q unitaire, donc de la forme $Q = X - \alpha$, $\alpha \in \mathbb{C}$. Posons alors $b = \nu(X - \alpha) < 1$. On fait le même travail que dans le cas $a < 1$, en remplaçant X par $X - \alpha$ (tout polynôme P s'écrit $P = \sum_{k=0}^{n} b_k (X - \alpha)^k$, d'après la formule de Taylor). On a alors, pour tout polynôme non nul P, $\nu(P) = b^{\nu_{X-\alpha}(P)}$, où $\nu_{X-\alpha}(P)$ est la valuation de P en $X - \alpha$ c'est-à-dire, si l'on préfère, l'ordre de multiplicité de la racine α dans P. Ici aussi, cette valuation s'étend au corps des fractions $\mathbb{C}(X)$ et $\nu(F) = b^{\nu_{X-\alpha}(F)}$ pour toute fraction non nulle F. ◁

Si $|\ |$ est une valeur absolue non-archimédienne sur le corps K, *l'application* $\nu : K \longmapsto \mathbb{R}$, *définie par* $\nu(x) = -\lambda \ln(|x|)$ *si* $x \neq 0$ *et* $\nu(0) = +\infty$, *où* λ *appartient à* $\mathbb{R}_+^*$, *est une valuation. Réciproquement, toute valuation* ν *sur* K *définit une valeur absolue non-archimédienne par la formule* $|x| = e^{-\mu\nu(x)}$ *si* $x \neq 0$, *où* $\mu \in \mathbb{R}_+^*$.

Il résulte de l'exercice précédent que les seules valeurs absolues non-archimédiennes sur $\mathbb{Q}$ *sont, outre la valeur absolue triviale (*$|x| = 1$ *si* $x \neq 0$*), les applications définies pour* $x \neq 0$ *par* $|x| = e^{-\lambda\nu_p(x)}$, *avec* $\lambda > 0$ *et p premier. Le théorème d'Ostrowski affirme d'autre part que les seules valeurs absolues archimédiennes sur* $\mathbb{Q}$ *sont les applications* $x \longmapsto |x|^s$, *où* $|\ |$ *est la valeur absolue usuelle et* $0 < s \leqslant 1$. *Elle définissent la même topologie que la valeur absolue usuelle.*

L'exercice suivant montre que n boules quelconques, de rayon $\varepsilon > 0$ *dans* $\mathbb{Q}$, *correspondant à n valeurs absolues deux à deux non équivalentes ont toujours une intersection non vide.*

3.20. Indépendance des valeurs absolues sur $\mathbb{Q}$

Soit $p_1, \ldots, p_n$ des nombres premiers deux à deux distincts, $q_0, q_1, \ldots, q_n$ dans $\mathbb{Q}$ et $\varepsilon > 0$. Prouver qu'il existe $q \in \mathbb{Q}$ tel que $|q - q_0| < \varepsilon$ et $|q - q_i|_{p_i} < \varepsilon$ pour tout $i \in [\![1, n]\!]$ (pour p premier, $|x|_p = p^{-\nu_p(x)}$ désigne la valeur absolue p-adique du rationnel x).

(École normale supérieure)

▷ **Solution.**

• Nous chercherons q sous la forme $\sum_{i=0}^{n} \alpha_i q_i$, avec $\alpha_0, \ldots, \alpha_n$ nombres rationnels dépendant d'un entier k à déterminer.

$\star$ On aura alors, $q - q_0 = \sum_{i=1}^{n} \alpha_i q_i + (\alpha_0 - 1)q_0$. Cette quantité devra être rendue assez petite. Pour cela, nous imposerons que les α_i, pour $i \geqslant 1$ et $\alpha_0 - 1$ tendent vers 0 avec k.

$\star$ Si $i \in [\![1, n]\!]$, $q - q_i = \sum_{\substack{0 \leqslant j \leqslant n \\ j \neq i}} \alpha_j q_j + (\alpha_i - 1)q_i$. Il faudra rendre la valeur absolue p_i-adique de cette quantité assez petite, c'est-à-dire sa valuation p_i-adique assez grande. Nous choisirons les α_i pour que, pour $j \in [\![0, n]\!]$, $j \neq i$, on ait $\nu_{p_i}(\alpha_j) = k$ et $\nu_{p_i}(\alpha_i - 1) = k$.

• Soit $k \in \mathbb{N}$ et p un nombre premier distinct de $p_1, \ldots, p_n$. Posons, pour tout $j \in [\![1, n]\!]$,

$$\alpha_j = \frac{(p_1 \ldots \widehat{p_j} \ldots p_n)^k}{p^k p_j^k + (p_1 \ldots \widehat{p_j} \ldots p_n)^k} \text{ et } \alpha_0 = \frac{(p_1 \ldots p_n)^k}{(p_1 \ldots p_n)^k + 1}$$

et montrons qu'ils conviennent. Soit $\varepsilon > 0$.

$\star$ Choisissons p tel que $p > p_1 p_2 \ldots p_n$. On a alors $\lim_{k \to +\infty} \alpha_i = 0$ pour tout i de $[\![1, n]\!]$ et, d'autre part, $\lim_{k \to +\infty} \alpha_0 = 1$. On en déduit que $\lim_{k \to +\infty} q = q_0$, et donc, pour k assez grand, $|q - q_0| < \varepsilon$.

$\star$ Soit $i \in [\![1, n]\!]$. Pour $j \in [\![0, n]\!]$, $j \neq i$, le numérateur de α_j est divisible par p_i^k et son dénominateur est premier avec p_i. On en déduit que $\nu_{p_i}(\alpha_j) = k$. On obtient également $\nu_{p_i}(\alpha_i - 1) = k$.

En posant $m = \min_{\substack{0 \leqslant j \leqslant n \\ 1 \leqslant i \leqslant n}} \nu_{p_i}(q_j)$, on a alors

$$\nu_{p_i}(\alpha_j q_j) = \nu_{p_i}(\alpha_j) + \nu_{p_i}(q_j) \geqslant k + m, \text{ pour } j \neq i,$$
$$\nu_{p_i}((\alpha_i - 1)q_i) \geqslant \nu_{p_i}(\alpha_i - 1) + \nu_{p_i}(q_i) \geqslant k + m.$$

Sachant que, pour tout $(x, y) \in \mathbb{Q}^2$, $\nu_{p_i}(x + y) \geqslant \min(\nu_{p_i}(x), \nu_{p_i}(y))$, on en déduit que, pour $i \geqslant 1$,

$$\nu_{p_i}(q - q_i) \geqslant k + m \text{ et donc } |q - q_i|_{p_i} \leqslant {p_i}^{-(k+m)}.$$

On a alors, pour k assez grand,

$$|q - q_i|_{p_i} \leqslant \varepsilon, \text{ pour tout } i \geqslant 1.$$

Pour k assez grand, q a donc toutes les propriétés voulues. $\triangleleft$

Chapitre 4

Arithmétique

Les nombres entiers ont toujours exercé sur les esprits une sorte de fascination. Il n'est pas étonnant que ce soit au sein de l'école pythagoricienne, éprise de mysticisme que débute l'étude de leurs propriétés : avec les catégories de pair et d'impair commencent les premières réflexions sur la divisibilité. Celles-ci aboutissent deux siècles plus tard, dans les Éléments *d'Euclide dont le chapitre VII s'ouvre par une série de définitions : nombre premier et composé, nombre parfait, diviseur et multiple, même si le langage est très différent du langage moderne. On trouve exposés entre autres l'algorithme de détermination du* pgcd *de deux nombres (algorithme d'Euclide) et la démonstration de l'infinité des nombres premiers (lemme d'Euclide).*

Aux pythagoriciens remontent les premiers exemples d'équations diophantiennes, notamment la résolution de l'équation $x^2 + y^2 = z^2$ en nombres entiers. Diophante, au III^e^ *siècle de notre ère, va s'intéresser à un grand nombre d'équations analogues. Sa grande œuvre,* les Arithmétiques *rompt avec la Mathématique grecque, centrée jusqu'alors sur les problèmes géométriques, et se rapproche des méthodes algébriques hautement élaborées des Babyloniens, mais contrairement à ceux-ci, il s'intéresse uniquement aux solutions exactes. Il s'évertue à donner des règles simples de calcul algébrique comme les règles sur les produits de puissance ou une première idée de la règle des signes dans un produit. Il est le premier à utiliser un symbole littéral pour désigner les inconnues et peut être vu comme le père de l'algèbre. Il résout des équations à coefficients entiers du premier degré ($ax + by = c$) et du second degré (cas particuliers d'équations de Pell telles que $x^2 = 1 + 30y^2$). Il s'intéresse à l'écriture d'un nombre comme somme de deux carrés (il sait qu'un entier de la forme $8n + 7$ ne peut pas s'écrire comme somme de deux carrés). Il énonce l'identité de Lagrange.*

Aux VII^e^ *et* VIII^e^ *siècles, les savants arabes, notamment à Bagdad, assimilent l'héritage hellénique et oriental à travers la traduction des ouvrages de l'Antiquité (Euclide, Archimède, Diophante...). Mais leurs théories mathématiques propres sont plus orientées vers l'Algèbre que vers l'Arithmétique proprement dite.*

En Occident, le réveil de la théorie des nombres doit attendre Pierre de Fermat (1601-1665) dont les premiers travaux s'appuient sur la récente traduction des textes grecs de Diophante par Bachet de Méziriac (1621). Il s'intéresse particulièrement à la divisibilité et aux nombres

premiers. Il aborde de nombreux problèmes et formule des conjectures qui pour la plupart seront prouvées au siècle suivant : citons le petit théorème de Fermat, l'équation de Pell-Fermat ($x^2 - dy^2 = 1$ a, pour tout entier positif non carré parfait d, une infinité de solutions entières), les nombres de Fermat et le grand théorème de Fermat dont nous parlerons un peu plus loin. Il développe la première méthode générale d'attaque des équations diophantiennes : la descente infinie. Au XVIII^e *siècle, Euler et Lagrange s'inspirent de ses méthodes et prouvent la plupart de ses résultats.*

En 1801, Gauss publie Disquisitiones arithmeticae *qui contribue à unifier de nombreux résultats et à développer des méthodes et des techniques nouvelles qui marquent le début de l'Arithmétique moderne : il fixe la terminologie et les notations de la théorie des congruences, réfléchit à la notion de divisibilité et de décomposition en facteurs premiers et l'étend au cas de nombres algébriques (comme les entiers de Gauss $\mathbb{Z}+i\mathbb{Z}$), élabore la théorie des formes quadratiques à coefficients entiers, énonce le théorème des nombres premiers.*

Les premiers exercices de ce chapitre s'appuient sur la notion élémentaire de divisibilité et les théorèmes fondamentaux que constituent le lemme d'Euclide et le théorème de Gauss.

4.1. Étude de l'irréductibilité d'une fraction

1. Montrer que pour tout $n \in \mathbb{N}$, la fraction $\dfrac{5^{n+1}+6^{n+1}}{5^n+6^n}$ est irréductible.

2. Trouver une condition nécessaire et suffisante sur $(\lambda, \mu, \alpha, \beta) \in \mathbb{N}^4$ pour que la fraction $\dfrac{\lambda\alpha^{n+1}+\mu\beta^{n+1}}{\lambda\alpha^n+\mu\beta^n}$ soit irréductible pour tout $n \in \mathbb{N}^*$.

(École polytechnique)

▷ **Solution.**

1. Soit $n \in \mathbb{N}$. Raisonnons par l'absurde et supposons qu'il existe p premier divisant à la fois $5^n + 6^n$ et $5^{n+1} + 6^{n+1}$. Alors modulo p, $5^n \equiv -6^n$ et $6.5^n \equiv -6^{n+1} \equiv 5^{n+1}$, d'où p divise $5^n(6-5) = 5^n$, ce qui entraîne $p = 5$. Or clairement, le numérateur et dénominateur ne sont pas congrus à 0 modulo 5 mais à 1. La fraction considérée est donc irréductible.

2. • Pour que la fraction considérée soit irréductible, il est nécessaire que λ soit premier avec μ et β et que α soit premier avec μ et β. Supposons ces conditions réalisées dans la suite.

• Supposons qu'il existe $n \geqslant 1$ tel que $\mathrm{F}_n = \dfrac{\lambda\alpha^{n+1} + \mu\beta^{n+1}}{\lambda\alpha^n + \mu\beta^n}$ ne soit pas irréductible. Il existe p nombre premier divisant à la fois le numérateur et le dénominateur de F_n. Modulo p, on obtient

$$\lambda\alpha^n \equiv -\mu\beta^n \quad \text{et} \quad \lambda\alpha^{n+1} \equiv -\mu\beta^{n+1}$$

Si p divise α, p divise $\mu\beta^n$ et d'après le lemme d'Euclide, il divise μ ou β. On aboutit une contradiction puisque α est premier avec μ et β. On en déduit que p ne divise pas α. On montre de même que p ne divise pas λ, μ et β. Les classes modulo p de ces entiers ne sont pas nulles dans le corps $\mathbb{Z}/p\mathbb{Z}$. On en déduit

$$\overline{\lambda} = -\overline{\mu}\left(\frac{\overline{\beta}}{\overline{\alpha}}\right)^n = -\overline{\mu}\left(\frac{\overline{\beta}}{\overline{\alpha}}\right)^{n+1}$$

et finalement $\dfrac{\overline{\beta}}{\overline{\alpha}} = 1$ après simplification (justifiée puisque les classes sont non nulles). Donc modulo p, $\alpha \equiv \beta$ et $\lambda \equiv -\mu$. En particulier, p divise $\alpha - \beta$ et $\lambda + \mu$. Nous avons donc

$$\operatorname{pgcd}(\alpha - \beta, \lambda + \mu) \neq 1.$$

• Réciproquement, supposons $\operatorname{pgcd}(\alpha - \beta, \lambda + \mu) \neq 1$. Il existe un nombre premier p divisant à la fois $\alpha - \beta$ et $\lambda + \mu$. Le système de congruences :

$$\lambda\alpha^n \equiv -\mu\beta^n \quad \text{et} \quad \lambda\alpha^{n+1} \equiv -\mu\beta^{n+1}$$

est vérifié pour tout $n \in \mathbb{N}^*$ et F_n n'est donc pas irréductible.

Conclusion. Pour que F_n soit irréductible pour tout $n \in \mathbb{N}^*$, il faut et il suffit que

$$\begin{aligned} \operatorname{pgcd}(\lambda, \mu) &= \operatorname{pgcd}(\lambda, \beta) = \operatorname{pgcd}(\alpha, \mu) = \operatorname{pgcd}(\alpha, \beta) \\ &= \operatorname{pgcd}(\alpha - \beta, \lambda + \mu) = 1. \triangleleft \end{aligned}$$

4.2. Équation $a^b = b^a$ dans $\mathbb{N}$

Trouver les couples (a, b) d'entiers strictement positifs tels que $a \neq b$ et $a^b = b^a$.

(École polytechnique)

▷ **Solution.**

• L'examinateur attendait sans doute une preuve arithmétique mais on constate qu'une simple étude de fonctions permet de répondre à la question. Un couple $(a,b) \in (\mathbb{N}^*)^2$, tel que $a < b$, est solution de l'équation si et seulement si $f(a) = f(b)$, où f est la fonction

$$f : x \in \mathbb{R}_+^* \longmapsto \frac{\ln x}{x}.$$

Cette fonction est strictement croissante sur $]0,e]$, allant de $-\infty$ à $\frac{1}{e}$, puis strictement décroissante sur $[e,+\infty[$, allant de $\frac{1}{e}$ à 0. On a donc nécessairement $a \leqslant e$. Comme $a = 1$ est visiblement exclu, on obtient $a = 2$. On remarque que $f(2) = f(4)$; le couple $(2,4)$ est donc solution de l'équation et c'est la seule (si $a < b$), étant données les variations de la fonction f.

• Donnons maintenant une preuve arithmétique. Si $(a,b) \in \mathbb{N}^{*2}$ est un couple solution avec $a < b$, on pose $d = \text{pgcd}(a,b)$, $a = da'$ et $b = db'$. Les entiers a' et b' sont premiers entre eux et $a' < b'$. L'équation devient alors $d^{b'-a'}a'^{b'} = b'^{a'}$. L'entier a' divise $b'^{a'}$, avec qui il est premier. On a donc $a' = 1$, $b' > 1$ et $d^{b'-1} = b'$; on en déduit que $d > 1$. Si $b' \geqslant 3$, on a $d^{b'-1} > b'$; en effet, on peut écrire

$$d^{b'-1} \geqslant (1+1)^{b'-1} > 1 + \mathrm{C}_{b'-1}^1 = b'.$$

On conclut que $b' = 2$ et $d = 2$, d'où l'on tire $a = 2$ et $b = 4$. Réciproquement, on constate que le couple $(2,4)$ est solution.

Vus les rôles symétriques joués par a et b, on conclut : les couples $(2,4)$ et $(4,2)$ sont les seuls couples solutions de l'équation $a^b = b^a$ avec $a \neq b$ dans $\mathbb{N}$. ◁

4.3. Divisibilité

Soit a, b dans $\mathbb{N}$ avec $b \geqslant a \geqslant 2$. On suppose que $a^n - 1$ divise $b^n - 1$ pour tout $n \in \mathbb{N}^*$. Montrer qu'il existe $p \in \mathbb{N}^*$ tel que $b = a^p$.
(**École normale supérieure**)

▷ **Solution.**

Notons tout d'abord que la réciproque est vraie et facile à démontrer. En effet, si $b = a^p$ pour un certain entier $p \geqslant 1$, alors, modulo $a^n - 1$, on a $a^n \equiv 1$ et donc $(a^n)^p \equiv 1^p \equiv 1$ c'est-à-dire $b^n \equiv 1$. Donc $a^n - 1$ divise $b^n - 1$ et cela pour tout $n \geqslant 1$. On va prouver par l'absurde que les entiers qui ne sont pas des puissances de a ne conviennent pas.

On fait donc l'hypothèse que $u_n = \dfrac{b^n - 1}{a^n - 1}$ est entier pour tout $n \geqslant 1$ et que b n'est pas une puissance de a. La suite u_n tend vers $+\infty$ et de plus $u_n \sim \left(\dfrac{b}{a}\right)^n$. L'idée est d'aboutir à une contradiction de nature asymptotique en considérant une combinaison linéaire à coefficients entiers des termes de la suite u. Comme $u_{n+1} \sim \left(\dfrac{b}{a}\right)^{n+1}$ il est naturel de poser

$$v_n = au_{n+1} - bu_n.$$

Cette suite (v_n) est toujours une suite d'entiers et elle va être plus petite que la suite (u_n). En effet, un petit calcul donne

$$v_n = \frac{(1-a)b^{n+1} + (b-1)a^{n+1} - b + a}{(a^n - 1)(a^{n+1} - 1)} \sim \frac{(1-a)b^{n+1}}{a^{2n+1}} = \frac{(1-a)b}{a}\left(\frac{b}{a^2}\right)^n.$$

Supposons alors pour commencer que $a < b < a^2$. L'équivalent obtenu montre que (v_n) tend vers 0 en étant non nulle à partir d'un certain rang : mais c'est absurde puisqu'il s'agit d'une suite d'entiers !

Avant de traiter le cas général, regardons encore ce qui se passe si on suppose que $a^2 < b < a^3$. Notre suite (v_n) diverge vers $+\infty$. Étant donné l'équivalent trouvé plus haut, il est naturel de recommencer et de poser

$$w_n = a^2 v_{n+1} - bv_n.$$

Le calcul exact de w_n est un peu long et on peut le confier à un logiciel de calcul formel. On obtient comme équivalent

$$w_n \sim \frac{(1-a^2)(1-a)b^2}{a^3}\left(\frac{b}{a^3}\right)^n$$

et cela conduit à la même contradiction que précédemment : w_n est une suite d'entiers non nuls à partir d'un certain rang qui converge vers 0.

Traitons maintenant le cas général et notons p l'unique entier naturel tel que $a^p < b < a^{p+1}$. L'étude des cas précédents nous invite à considérer les opérateurs $\mathrm{T}_k : (x_n)_{n\geqslant 0} \mapsto (a^k x_{n+1} - bx_n)_{n\geqslant 0}$ sur l'espace des suites réelles. Il est clair que l'image par T_k d'une suite à coefficients dans $\mathbb{Z}$ est encore une suite à coefficients dans $\mathbb{Z}$. On considère alors la suite $(\mathrm{T}_p \circ \cdots \circ \mathrm{T}_2 \circ \mathrm{T}_1)(u)$ et on va montrer qu'elle tend vers 0 en étant non nulle à partir d'un certain rang, ce qui fournira la contradiction cherchée. Le calcul exact de cette suite étant inextricable on va se contenter de travailler sur des développements asymptotiques. On a

$$u_n = \frac{b^n}{a^n} + \frac{b^n}{a^{2n}} + \cdots + \frac{b^n}{a^{pn}} + \frac{b^n}{a^{(p+1)n}} + o\left(\frac{b^n}{a^{(p+1)n}}\right)$$

car $\dfrac{1}{a^n}$ est négligeable devant $\dfrac{b^n}{a^{(p+1)n}}$ étant donné que $\dfrac{1}{a} < \dfrac{b}{a^{p+1}}$ par hypothèse. On a bien entendu poussé le développement de u_n jusqu'à

un terme qui tend vers 0. On a alors le développement de u_{n+1} dans la même échelle :

$$u_{n+1} = \frac{b}{a}\frac{b^n}{a^n} + \frac{b}{a^2}\frac{b^n}{a^{2n}} + \cdots + \frac{b}{a^p}\frac{b^n}{a^{pn}} + \frac{b}{a^{p+1}}\frac{b^n}{a^{(p+1)n}} + o\left(\frac{b^n}{a^{(p+1)n}}\right).$$

On en déduit que

$$v_n = \mathrm{T}_1(u)_n = b\left(\frac{1}{a} - 1\right)\frac{b^n}{a^{2n}} + \cdots + b\left(\frac{1}{a^p} - 1\right)\frac{b^n}{a^{(p+1)n}} + o\left(\frac{b^n}{a^{(p+1)n}}\right).$$

L'opérateur T_1 a annulé le premier coefficient du développement asymptotique (on retrouve notamment l'équivalent ci-dessus). En itérant de même on voit que $w_n = (\mathrm{T}_2 \circ \mathrm{T}_1)(u)_n$ a pour développement

$$b^2\left(\frac{1}{a^2} - 1\right)\left(\frac{1}{a} - 1\right)\frac{b^n}{a^{3n}} + \cdots + b^2\left(\frac{1}{a^p} - 1\right)\left(\frac{1}{a^{p-1}} - 1\right)\frac{b^n}{a^{(p+1)n}} + o\left(\frac{b^n}{a^{(p+1)n}}\right).$$

Une récurrence finie facile permet de voir que

$$(\mathrm{T}_p \circ \cdots \circ \mathrm{T}_2 \circ \mathrm{T}_1)(u)_n \sim b^p\left(\frac{1}{a^p} - 1\right)\cdots\left(\frac{1}{a} - 1\right)\frac{b^n}{a^{(p+1)n}}$$

ce qui permet de conclure car le coefficient est bien non nul. ◁

4.4. Points du réseau $\mathbb{Z}^n$ visibles de l'origine

Soit $\mathrm{P} = (a_1, \ldots, a_n) \in \mathbb{Z}^n$ et $\mathrm{Q} = (b_1, \ldots, b_n) \in \mathbb{Z}^n$.

1. Montrer l'équivalence des deux conditions suivantes :

(*i*) les $b_i - a_i$ sont premiers entre eux dans leur ensemble ;

(*ii*) il n'existe pas de points de $\mathbb{Z}^n$ sur le segment $]\mathrm{P}, \mathrm{Q}[\subset \mathbb{R}^n$ (si cette condition est réalisée, on dit que Q est visible depuis P).

2. On prend $n = 2$. Montrer que pour tout $r \in \mathbb{N}^*$, il existe un carré de $\mathbb{Z}^2$ de côté r (*i.e.* une partie du type $[\![a+1, a+r]\!] \times [\![b+1, b+r]\!]$ avec $(a, b) \in \mathbb{Z}^2$), dont tous les points sont invisibles de l'origine.

(École polytechnique)

▷ **Solution.**

1. Quitte à faire une translation de vecteur $(a_1, \ldots, a_n) \in \mathbb{Z}^n$, on peut se ramener au cas où $\mathrm{P} = (0, \ldots, 0)$.

• Supposons que $]\mathrm{P}, \mathrm{Q}[$ admette un point $(\ell_1 \ldots, \ell_n) \in \mathbb{Z}^n$. Alors il existe $\lambda \in]0, 1[$ tel que $\ell_i = \lambda b_i$ pour tout $1 \leqslant i \leqslant n$. Nécessairement, λ est un rationnel qui s'écrit $\frac{p}{q}$, avec $(p, q) \in \mathbb{N}^{*2}$ et $p < q$. Ainsi, pour

tout $1 \leqslant i \leqslant n$, on a $pb_i = q\ell_i$ et donc

$$p\,\mathrm{pgcd}(b_1, \ldots, b_n) = q\,\mathrm{pgcd}(\ell_1, \ldots, \ell_n),$$
$$\mathrm{pgcd}(b_1, \ldots, b_n) = \frac{q}{p}\,\mathrm{pgcd}(\ell_1, \ldots, \ell_n) > 1.$$

Les b_i ne sont donc pas premiers entre eux dans leur ensemble.

• Réciproquement, supposons que les b_i ne sont pas premiers entre eux. Soit donc $k \in \mathbb{N}$, $k > 1$, un diviseur commun à tous les b_i. Alors le point $\left(\frac{b_1}{k}, \frac{b_2}{k}, \ldots, \frac{b_n}{k}\right)$ est dans $\mathbb{Z}^n$ et il se situe sur le segment $]\mathrm{P}, \mathrm{Q}[$ puisque $0 < \frac{1}{k} < 1$. L'équivalence est prouvée.

2. D'après ce qui précède, il suffit de trouver deux entiers a et b tels que pour tout couple $(i, j) \in [\![1, r]\!]^2$, $a+i$ et $b+j$ ne soient pas premiers entre eux. En particulier l'entier $a+i$ doit avoir un facteur premier en commun avec $b+j$. On se donne donc une famille $(p_{ij})_{1 \leqslant i,j \leqslant r}$ de nombres premiers, deux à deux distincts. Le théorème chinois, assure l'existence d'un entier a vérifiant pour tout $(i, j) \in [\![1, r]\!]^2$, $a \equiv -i\ [p_{ij}]$. De même, il existe un entier b tel que $b \equiv -j\ [p_{ij}]$ pour tout couple (i, j). On a alors la propriété désirée, puisque $a+i$ et $b+j$ ont le nombre premier p_{ij} comme diviseur commun. ◁

4.5. Produits d'entiers consécutifs

Soit $\mathrm{N}_1, \ldots, \mathrm{N}_q$ des entiers non nuls, deux à deux distincts. On pose $\mathrm{P}_k = \prod_{i=1}^{q} (\mathrm{N}_i + k)$, pour $k \in \mathbb{Z}$ et on suppose que pour tout $k \in \mathbb{Z}$, P_0 divise P_k.

1. Montrer qu'il existe $i \in [\![1, q]\!]$ tel que $|\mathrm{N}_i| = 1$.

2. On suppose de plus que, pour tout $1 \leqslant i \leqslant q$, on a $\mathrm{N}_i \geqslant 1$. Montrer que $\mathrm{N}_1, \ldots, \mathrm{N}_q$ sont les q premiers entiers naturels non nuls.

(École polytechnique)

Solution.

Rappelons que pour N *et* d *dans* $\mathbb{N}$, *si* $d|\mathrm{N}$ *et* $d > \mathrm{N}$ *alors* $\mathrm{N} = 0$.

1. On a

$$\mathrm{P}_0 = \mathrm{N}_1\mathrm{N}_2 \ldots \mathrm{N}_q,\ \mathrm{P}_1 = (\mathrm{N}_1+1)\ldots(\mathrm{N}_q+1),\ \mathrm{P}_{-1} = (\mathrm{N}_1-1)\ldots(\mathrm{N}_q-1).$$

Comme P_0 divise P_1 et P_{-1}, P_0^2 divise $\mathrm{P}_1\mathrm{P}_{-1}$. Puisque

$$0 \leqslant \mathrm{P}_1\mathrm{P}_{-1} = (\mathrm{N}_1^2 - 1)\ldots(\mathrm{N}_q^2 - 1) < \mathrm{N}_1^2 \ldots \mathrm{N}_q^2 = \mathrm{P}_0^2,$$

on a nécessairement $\mathrm{P}_1\mathrm{P}_{-1} = 0$ et il existe un indice i tel que $\mathrm{N}_i^2 = 1$, c'est-à-dire $|\mathrm{N}_i| = 1$.

2. On peut supposer $1 \leqslant N_1 < N_2 < \cdots < N_q$, quitte à renuméroter les N_i. Montrons, par récurrence sur k compris entre 1 et q, que $N_k = k$.

- Le cas $k = 1$ résulte de la question précédente.
- Supposons $k \geqslant 2$. On a, d'après l'hypothèse de récurrence,

$$\begin{aligned} P_0 &= 1 \cdot 2 \ldots (k-1) N_k \ldots N_q = (k-1)! N_k \ldots N_q \text{ et} \\ P_{-k} &= (-k+1)(-k+2) \ldots (-2)(-1)(N_k - k) \ldots (N_q - k) \\ &= (-1)^{k-1}(k-1)!(N_k - k) \ldots (N_q - k). \end{aligned}$$

De P_0 divise P_{-k}, on déduit que $N_k \ldots N_q$ divise $(N_k - k) \ldots (N_q - k)$. Or si $\ell \in [\![k, q]\!]$, on a $N_\ell > N_{k-1} = k - 1$, c'est-à-dire $N_\ell \geqslant k$. On obtient donc

$$0 \leqslant (N_k - k) \ldots (N_q - k) < N_k \ldots N_q.$$

Il en résulte que $(N_k - k) \ldots (N_q - k) = 0$. Il existe $\ell \in [\![k, q]\!]$ tel que $N_\ell = k$. Comme $k \leqslant N_k \leqslant N_\ell$, on a, puisque les N_i sont deux à deux distincts, $\ell = k$ et $N_k = k$. ◁

L'exercice suivant exploite le théorème de Bezout : des entiers $a_1, \ldots, a_n$ sont premiers entre eux, si et seulement si il existe une combinaison linéaire des a_k à coefficients entiers égale à 1. Ce résultat fut démontré par Bachet de Méziriac (1621), puis généralisé par Bezout (1730-1752) pour des polynômes à coefficients réels.

4.6. Parties de $\mathbb{N}$ additivement stables

Soit P une partie de $\mathbb{N}$ stable par addition. Montrer qu'il existe $(n, k) \in \mathbb{N}^2$ tel que $P \cap [\![n, +\infty[\![= k\mathbb{N} \cap [\![n, +\infty[\![$.

(**École normale supérieure**)

▷ **Solution.**

Supposons $P \neq \{0\}$, le cas $P = \{0\}$ étant trivial. Soit d le pgcd de tous les éléments de P ; d appartient à $\mathbb{N}^*$.

- Traitons, pour commencer, le cas où que $d = 1$ et montrons qu'alors $k = 1$ convient. On fait l'hypothèse supplémentaire que P contient 0, l'y rajouter ne changeant rien à la propriété à démontrer. La remarque fondamentale est que si P est additivement stable et contient 0, alors toute combinaison linéaire d'éléments de P à coefficients entiers naturels appartient à P.

⋆ Montrons qu'il existe $m \in \mathbb{N}$ tel que m et $m + 1$ soient dans P.

L'ensemble P est infini car si $x \in P \setminus \{0\}$, alors $x\mathbb{N} \subset P$. Soit $(a_i)_{i \in \mathbb{N}}$ la suite des éléments de P, rangés par ordre croissant. Pour tout entier naturel i, posons $d_i = \text{pgcd}(a_0, \ldots, a_i)$. Alors (d_i) est une suite décroissante

d'entiers naturels, donc stationnaire : il existe des entiers naturels δ et i_0 tels que $d_i = \delta$ pour $i \geqslant i_0$. Comme δ divise tous les éléments de P, on a $\delta = 1$ et en particulier, $\operatorname{pgcd}(a_0, \ldots, a_{i_0}) = 1$.

D'après le théorème de Bezout, il existe $(u_0, \ldots, u_{i_0}) \in \mathbb{Z}^{i_0+1}$ tel que $\sum_{i=0}^{i_0} u_i a_i = 1$. Soit I (resp. J) l'ensemble des indices i tels que $u_i > 0$ (resp. $u_i \leqslant 0$). On peut écrire

$$1 = \sum_{i \in \mathrm{I}} u_i a_i - \sum_{i \in \mathrm{J}} |u_i| a_i.$$

D'après la remarque initiale, on a

$$m = \sum_{i \in \mathrm{J}} |u_i| a_i \in \mathrm{P} \quad \text{et} \quad m + 1 = \sum_{i \in \mathrm{I}} u_i a_i \in \mathrm{P}.$$

$\star$ La même remarque justifie le fait que P contient tous les entiers naturels de la forme $a(m+1) + bm$, avec $(a, b) \in \mathbb{N}^2$. Montrons que tout entier naturel supérieur ou égal à $m(m-1)$ peut s'écrire ainsi. Si $\mathrm{N} \geqslant m(m-1)$, on effectue la division euclidienne de N par m. On obtient $\mathrm{N} = qm + r$, avec $0 \leqslant r \leqslant m-1$ et $q \geqslant m-1$ car $\mathrm{N} \geqslant m(m-1)$ (en effet, si $q \leqslant m-2$, $qm + r \leqslant m(m-2) + m - 1 = m(m-1) - 1$). On peut alors écrire

$$\mathrm{N} = qm + r(m + 1 - m) = (q - r)m + r(m+1),$$

avec r et $q - r$ positifs. C'est la décomposition cherchée, qui montre que P contient tous les entiers supérieurs ou égaux à $m(m-1)$. On a donc bien $1\mathbb{N} \cap [\![m(m-1), +\infty[\![= \mathrm{P} \cap [\![m(m-1), +\infty[\![$.

• Passons au cas général, $d \geqslant 2$. En posant $\mathrm{P} = d\mathrm{P}'$, on obtient une partie P′ additivement stable de $\mathbb{N}$, telle que le pgcd de tous ses éléments soit 1. Il existe donc, d'après ce qui précède, $n \in \mathbb{N}$ tel que $\mathrm{P}' \cap [\![n, +\infty[\![= [\![n, +\infty[\![$ et donc $\mathrm{P} \cap [\![dn, +\infty[\![= d\mathbb{N} \cap [\![dn, +\infty[\![$, ce qui montre la propriété pour P. $\lhd$

L'exercice suivant montre l'intérêt qu'on peut avoir à réduire certains problèmes modulo n pour un entier n bien choisi. Il commence une série d'exercices consacrés aux congruences et aux anneaux $\mathbb{Z}/n\mathbb{Z}$.

4.7. Un exercice pour les années impaires

Montrer qu'il n'existe pas d'application $f : \mathbb{N} \to \mathrm{N}$ telle que pour tout $n \in \mathbb{N}$, $f(f(n)) = n + 2007$.

(École normale supérieure)

▷ **Solution.**

Supposons l'existence d'une telle application f et soit $n \in \mathbb{N}$. On a

$$f(n+2007) = f(f(f(n))) = (f \circ f)(f(n)) = f(n) + 2007.$$

Par une récurrence immédiate, on en déduit que, pour tout $k \in \mathbb{N}$,

$$f(n+2007k) = f(n) + 2007k.$$

Il en résulte que le résidu de $f(n)$ modulo 2007 ne dépend que du résidu de n modulo 2007. Ainsi, f induit une application $\overline{f}$ de $\mathbb{Z}/2007\mathbb{Z}$ dans lui-même, définie par $\overline{f}(\overline{n}) = \overline{f(n)}$, où $\overline{k}$ désigne le projeté de l'entier k dans $\mathbb{Z}/2007\mathbb{Z}$. La propriété vérifiée par f implique que $\overline{f} \circ \overline{f} = \mathrm{Id}$. Or, le cardinal de $\mathbb{Z}/2007\mathbb{Z}$ est impair et toute involution d'un ensemble fini de cardinal impair admet au moins un point fixe (en effet, la permutation f est d'ordre 2 et se décompose donc en produit de transpositions à supports disjoints ; il y a donc une orbite réduite à un singleton). Il existe donc un entier $a \in \mathbb{N}$ (que l'on peut même prendre dans $[\![0, 2006]\!]$) et $k \in \mathbb{N}$ tels que $f(a) = a + 2007k$. On obtient alors

$$f(f(a)) = a + 2007 = f(a + 2007k) = f(a) + 2007k = a + 2 \cdot 2007k.$$

D'où l'on tire $2k = 1$, ce qui est impossible dans $\mathbb{Z}$ et fournit la contradiction souhaitée. ◁

4.8. Équation du second degré dans $\mathbb{Z}/p\mathbb{Z}$

Montrer que pour tout nombre premier p, il existe un entier naturel n tel que $6n^2 + 5n + 1 \equiv 0\ [p]$.

(École polytechnique)

▷ **Solution.**

Pour $p = 2$ ou 3, $n = 1$ est solution. Supposons $p \geqslant 5$ premier. Alors $\mathbb{Z}/p\mathbb{Z}$ est alors un corps commutatif de caractéristique distincte de 2. Comme $\overline{6}$ est non nul dans $\mathbb{Z}/p\mathbb{Z}$, l'équation $6x^2+5x+1$ est une équation du second degré. On calcule son discriminant $\Delta = 25 - 24 = 1 = 1^2$. C'est un carré. Donc l'équation admet deux racines distinctes dans $\mathbb{Z}/p\mathbb{Z}$, ce qui assure l'existence de $n \in \mathbb{N}$ tel que $6n^2 + 5n + 1 \equiv 0\ [p]$. ◁

Lorsque p est premier, $\mathbb{Z}/p\mathbb{Z}$ est un corps et réciproquement. Dans ces conditions, $(\mathbb{Z}/p\mathbb{Z})^*$ est un groupe multiplicatif de cardinal $p-1$ et le théorème de Lagrange impose alors que si p ne divise pas a, on ait

$a^{p-1} \equiv 1 \pmod{p}$: *c'est le petit théorème de Fermat qui est souvent utilisé dans les calculs de congruences.*

4.9. Un problème de congruence

Montrer que pour tout $n \in \mathbb{N}^*$, $10^{10^n} \equiv 4 \pmod{7}$.

(École polytechnique)

Solution.

Posons $A = 10^{10^n}$. On a $A \equiv 3^{10^n} \pmod{7}$ et, puisque 7 est un nombre premier ne divisant pas 3, d'après le petit théorème de Fermat, $3^6 \equiv 1 \pmod{7}$. Recherchons donc le reste de 10^n modulo 6, c'est-à-dire le reste de 4^n modulo 6. On obtient $4^2 \equiv 4 \pmod{6}$, puis, pour tout entier $n \geqslant 2$,

$$4^n \equiv 4^2 4^{n-2} \equiv 4 \cdot 4^{n-2} = 4^{n-1} \pmod{6}.$$

On a donc, pour $n \geqslant 1$, $4^n \equiv 4 \pmod{6}$ et $A \equiv 3^4 = 81 \equiv 4 \pmod{7}$. ◁

4.10. Un multiple de 1996 qui ne s'écrit qu'avec des 4

Montrer qu'il existe un multiple de 1996 dont l'écriture décimale ne comporte que le chiffre 4.

(École normale supérieure)

Solution.

Notons $u_n = 444\ldots4 = \sum_{k=0}^{n-1} 4.10^k = \frac{4}{9}(10^n - 1)$ l'entier dont l'écriture décimale est composée du chiffre 4 répété n fois. On a $1996 = 4 \times 499$ et 499 est un nombre premier. On en déduit, d'après le petit théorème de Fermat, $10^{498} \equiv 1 \ [499]$. Comme 9 est premier avec 499, 499 divise aussi l'entier $\dfrac{10^{498} - 1}{9}$. En multipliant par 4 on en déduit donc que 1996 divise u_{498}. ◁

4.11. Somme des puissances k-ièmes dans $\mathbb{Z}/p\mathbb{Z}$

Soit p un nombre premier et $k \in \mathbb{N}^*$. Montrer que $\sum_{x \in \mathbb{Z}/p\mathbb{Z}} x^k$ est égal à 0 ou -1. Préciser à quelle condition on obtient 0 ou -1.

(École polytechnique)

▷ **Solution.**

Posons $S_k = \sum_{x \in \mathbb{Z}/p\mathbb{Z}} x^k$. Soit y un élément non nul de $\mathbb{Z}/p\mathbb{Z}$. L'application $x \longmapsto yx$ est une bijection de $\mathbb{Z}/p\mathbb{Z}$. Il en résulte que :

$$y^k S_k = \sum_{x \in \mathbb{Z}/p\mathbb{Z}} (yx)^k = \sum_{z \in \mathbb{Z}/p\mathbb{Z}} z^k = S_k.$$

De deux choses l'une :

- soit il existe y non nul tel que $y^k \neq 1$ et alors $S_k = 0$;
- soit $y^k = 1$ pour tout y non nul et dans ce cas, $S_k = \overline{p-1} = -1$.

On sait déjà par le petit théorème de Fermat que, si k est un multiple de $p-1$, alors c'est la deuxième éventualité qui se produit. Montrons réciproquement que si $p-1$ ne divise pas k, alors $S_k = 0$. Pour cela, posons $d = \operatorname{pgcd}(p-1, k)$. On a par hypothèse $d < p-1$. Supposons alors par l'absurde que $y^k = 1$ pour tout $y \in (\mathbb{Z}/p\mathbb{Z})^*$. Dans ce cas l'ordre de tout y non nul de $(\mathbb{Z}/p\mathbb{Z})^*$ divise k (par hypothèse) et $p-1$ (par le théorème de Lagrange) donc divise d. Comme le polynôme $X^d - 1$ ne peut pas avoir $p-1$ racines distinctes, c'est absurde.

Ainsi, $S_k = 0$ si et seulement si $p-1$ divise k. ◁

4.12. Théorème de Wilson (1759)

Soit $p \in \mathbb{N}^*$, $p \geqslant 2$. Montrer l'équivalence :

$$(p-1)! \equiv -1 \pmod{p} \iff p \text{ premier.}$$

(École polytechnique)

▷ **Solution.**

- Supposons p premier. Nous pouvons écrire $\overline{(p-1)!} = \prod_{x \in (\mathbb{Z}/p\mathbb{Z})^*} x$.

Comme p est premier, $\mathbb{Z}/p\mathbb{Z}$ est un corps et dans ce produit, on peut associer par paires chaque élément x et son inverse x^{-1}, lorsque $x \neq x^{-1}$, *i.e.* $x^2 \neq 1$ ou encore $x \neq \pm\overline{1}$. Le produit de x par x^{-1} fait 1 et il reste donc

$$\overline{(p-1)!} = \overline{1} \times (-\overline{1}) = -\overline{1}.$$

- Réciproquement, supposons que p divise $1 + (p-1)!$. On a donc $\overline{(p-1)!} = -\overline{1}$ dans $\mathbb{Z}/p\mathbb{Z}$. On en déduit que si $1 \leqslant k \leqslant p-1$, $\overline{k}$ est inversible (d'inverse $-\prod_{\substack{1 \leqslant \ell \leqslant p-1 \\ \ell \neq k}} \overline{\ell}$). Donc $\mathbb{Z}/p\mathbb{Z}$ est un corps et p est premier. ◁

Bien entendu ce critère de primalité est inutilisable en pratique. Il semblerait que le résultat était déjà connu de Leibniz. On peut en déduire que si p est un nombre premier congru à 1 modulo 4, alors -1 est un

carré modulo p (regrouper chaque facteur de $(p-1)!$ avec son opposé pour obtenir $-1 \equiv \left(\frac{p-1}{2}!\right)^2$; on pourra aussi se reporter à l'exercice 4.34).

4.13. Cyclicité du groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$

Soit p un nombre premier.

1. Soit q un nombre premier qui divise $p-1$. Établir l'existence d'un élément de $((\mathbb{Z}/p\mathbb{Z})^*, \times)$ d'ordre multiplicatif q.

2. Soit q un nombre premier et $\alpha \in \mathbb{N}^*$ tels que q^α divise $p-1$. Montrer l'existence d'un élément de $((\mathbb{Z}/p\mathbb{Z})^*, \times)$ d'ordre q^α.

3. En déduire que $((\mathbb{Z}/p\mathbb{Z})^*, \times)$ est cyclique.

(École normale supérieure)

▷ **Solution.**

1. L'entier p étant premier, $\mathbb{Z}/p\mathbb{Z}$ est un corps et $(\mathbb{Z}/p\mathbb{Z})^*$ est son groupe multiplicatif. Pour tout x dans $(\mathbb{Z}/p\mathbb{Z})^*$, notons $y_x = x^{\frac{p-1}{q}}$. On a alors $(y_x)^q = x^{p-1} = 1$, d'après le petit théorème de Fermat. L'ordre de y_x divise donc q et puisque q est premier, il est donc égal à q ou à 1. Dire que l'ordre de y_x est 1, c'est dire que $y_x = 1$. Imaginons que cela soit le cas pour tout x dans $(\mathbb{Z}/p\mathbb{Z})^*$. Le polynôme $X^{\frac{p-1}{q}} - 1$ aurait au moins $p-1$ racines distinctes. Comme son degré $\frac{p-1}{q}$ est strictement inférieur à $p-1$, c'est absurde. Il existe donc $x \in (\mathbb{Z}/p\mathbb{Z})^*$ pour lequel $y_x \neq 1$. Cet élément y_x est d'ordre q.

Le résultat de cette question est un cas particulier du lemme de Cauchy démontré dans l'exercice 2.10.

2. Inspirons-nous de ce qui précède : pour tout x dans $(\mathbb{Z}/p\mathbb{Z})^*$, on pose $y_x = x^{\frac{p-1}{q^\alpha}}$. On a alors $(y_x)^{q^\alpha} = x^{p-1} = 1$. L'ordre de y_x divise donc q^α ; il est de la forme q^{r_x} avec $r_x \leqslant \alpha$. On considère le plus grand des entiers r_x pour x décrivant l'ensemble fini $(\mathbb{Z}/p\mathbb{Z})^*$; on le note r. On a $r \leqslant \alpha$. On obtient $x^{\frac{p-1}{q^\alpha}q^r} = (y_x)^{q^r} = \left((y_x)^{q^{r_x}}\right)^{q^{r-r_x}} = 1$, pour tout $x \in (\mathbb{Z}/p\mathbb{Z})^*$. Le polynôme $X^{\frac{p-1}{q^{\alpha-r}}} - 1$ a au moins $p-1$ racines distinctes. Ce polynôme n'étant manifestement pas le polynôme nul, son degré doit être supérieur ou égal à $p-1$. On a donc $\frac{p-1}{q^{\alpha-r}} = p-1$ et $\alpha = r$. Étant donnée la définition de r, il existe donc dans $(\mathbb{Z}/p\mathbb{Z})^*$ un élément d'ordre q^α.

3. Il a été démontré dans le chapitre 2 sur les groupes (exercice 2.8) que si G est un groupe abélien et x et y sont deux éléments de G d'ordre

p et q respectivement, p et q étant premiers entre eux, alors l'ordre de xy dans G est pq.

Bien évidemment, on établit par récurrence que si $x_1, \ldots, x_r$ sont d'ordres respectifs $p_1, \ldots, p_r$, les p_i étant deux à deux premiers entre eux, l'ordre de leur produit $x_1 \ldots x_r$ est $p_1 \ldots p_r$.

Décomposons donc $p-1$ en produit de facteurs premier $q_1^{\alpha_1} \ldots q_r^{\alpha_r}$, les q_i étant des entiers premiers et distincts deux à deux et les α_i des entiers naturels non nuls. D'après la question 2, il existe, pour tout $1 \leqslant i \leqslant r$, un élément x_i de $(\mathbb{Z}/p\mathbb{Z})^*$ d'ordre $q_i^{\alpha_i}$. Ainsi, le produit $x_1 \ldots x_r$ est d'ordre $q_1^{\alpha_1} \ldots q_r^{\alpha_r} = p-1$. Le groupe $(\mathbb{Z}/p\mathbb{Z})^*$ a donc un élément d'ordre $p-1$. Il est cyclique. ◁

On pourra également consulter l'exercice 2.8 qui établit un résultat plus général, à savoir : tout sous-groupe fini du groupe multiplicatif d'un corps commutatif est cyclique.

Dans le cas où on ne suppose plus n premier, on peut considérer le groupe des inversibles de $\mathbb{Z}/n\mathbb{Z}$, habituellement noté $(\mathbb{Z}/n\mathbb{Z})^\times$. Son cardinal est le nombre d'entiers k compris entre 1 et n, premiers avec n. Il est noté $\varphi(n)$, où φ est appelée indicatrice d'Euler. *Si $a \in \mathbb{Z}$ est premier avec n, il vérifie, d'après le théorème de Lagrange, $a^{\varphi(n)} \equiv 1 \pmod{n}$: c'est un théorème d'Euler qui généralise le petit théorème de Fermat.*

4.14. Critères de primalité

Soit a et p des entiers de $\mathbb{N}^*$ tels que $a^{p-1} \equiv 1 \pmod{p}$.

1. On suppose que pour tout diviseur d de $p-1$ autre que $p-1$, l'entier $a^d - 1$ est premier avec p. Montrer que p est premier.

2. On suppose que $p-1$ se décompose en $p-1 = rs$ avec $r \geqslant s$ et que, pour tout diviseur t de r autre que 1, l'entier $a^{\frac{p-1}{t}} - 1$ est premier avec p. Montrer que p est premier.

(École normale supérieure)

▷ **Solution.**

1. Considérons le groupe multiplicatif des inversibles de $\mathbb{Z}/p\mathbb{Z}$ que nous noterons $(\mathbb{Z}/p\mathbb{Z})^\times$. L'entier p est premier si et seulement si $\mathbb{Z}/p\mathbb{Z}$ est un corps, c'est-à-dire si et seulement si tout élément de $(\mathbb{Z}/p\mathbb{Z})^*$ est inversible, soit encore si et seulement si $\varphi(p) = \mathrm{Card}(\mathbb{Z}/p\mathbb{Z})^\times = p-1$. On a, dans $\mathbb{Z}/p\mathbb{Z}$, $\overline{a}^{p-1} = 1$; $\overline{a}$ est donc inversible et son ordre dans $(\mathbb{Z}/p\mathbb{Z})^\times$ divise $p-1$. On sait par hypothèse que, pour tout diviseur d de $p-1$, $a^d - 1$ est premier avec p et en particulier $\overline{a}^d \neq 1$. L'ordre de $\overline{a}$ est donc $p-1$. Il en résulte que $\varphi(p) \geqslant p-1$ et donc $\varphi(p) = p-1$. Cela montre que p est premier.

2. Raisonnons par l'absurde et supposons p non premier. Il admet alors un diviseur premier $q \leqslant \sqrt{p}$. On a *a fortiori* $a^{p-1} \equiv 1 \pmod{q}$. Dans l'anneau $\mathbb{Z}/q\mathbb{Z}$, on obtient donc $\overline{a}^{p-1} = 1$, soit $(\overline{a}^s)^r = 1$; $\overline{a}^s$ est donc un inversible de $\mathbb{Z}/q\mathbb{Z}$ dont l'ordre divise r. Traduisons l'hypothèse : pour tout diviseur $t \neq 1$ de r, $a^{\frac{p-1}{t}} - 1$ n'est pas divisible par q, c'est-à-dire $\overline{a}^{\frac{p-1}{t}} \neq 1$. Or nous avons

$$\overline{a}^{\frac{p-1}{t}} = \overline{a}^{\frac{sr}{t}} = (\overline{a}^s)^{\frac{r}{t}}.$$

Puisque $\dfrac{r}{t}$ décrit les diviseurs stricts de r, lorsque t décrit les diviseurs de r autre que 1, l'ordre de $\overline{a}^s$ est r. On en déduit que $r \leqslant q-1$. D'autre part, comme r et s ne peuvent être tous deux inférieurs strictement à $\sqrt{p-1}$ et comme $r \geqslant s$, on a $r \geqslant \sqrt{p-1}$. Il en résulte $\sqrt{p-1} \leqslant q-1 \leqslant \sqrt{p}-1$. En élevant au carré, il vient $p-1 \leqslant p+1-2\sqrt{p}$ et $\sqrt{p} \leqslant 1$, ce qui est exclu.

Conclusion. p est un nombre premier. ◁

Les critères de primalité de ce type sont dus à Lehmer. Le lecteur intéressé par ces questions pourra consulter l'excellent cours d'algèbre de Michel Demazure[1].

4.15. Diviseurs premiers communs aux termes d'une suite arithmétique

1. Soit a et r deux entiers relatifs premiers entre eux. Montrer qu'il existe $k \in \mathbb{N}^*$ tel que $a^k \equiv 1 \pmod{r}$.

2. Soit a et r deux entiers avec $a > r \geqslant 2$. Montrer que la progression arithmétique de premier terme a et de raison r contient une infinité de termes ayant tous les mêmes diviseurs premiers.

(École polytechnique)

▷ **Solution.**

1. Comme a est premier avec r, sa classe $\overline{a}$ est inversible dans $\mathbb{Z}/r\mathbb{Z}$. L'ensemble des éléments inversibles de $\mathbb{Z}/r\mathbb{Z}$ est un groupe fini. Si $k \geqslant 1$ est l'ordre de $\overline{a}$ dans ce groupe, on a $\overline{a}^k = \overline{1}$, c'est-à-dire $a^k \equiv 1 \pmod{r}$.

2. Si $\operatorname{pgcd}(a,r) = d$, on pose $a = \alpha d$ et $r = \rho d$, avec $\operatorname{pgcd}(\alpha,\rho) = 1$ et $\alpha > \rho \geqslant 1$. Pour tout $n \in \mathbb{N}$, on a $a + nr = d(\alpha + n\rho)$. L'ensemble des diviseurs premiers de $a + nr$ étant la réunion de ceux de d et de $\alpha + n\rho$, il suffit donc de démontrer le résultat pour α et ρ. Autrement dit, on peut supposer a et r premiers entre eux et $a \geqslant 2$.

Reprenons les notations de la première question. Pour tout $n \in \mathbb{N}$, on a $a^{nk} \equiv 1 \pmod{r}$ et donc $a^{nk+1} \equiv a \pmod{r}$. Autrement dit, il

1. DEMAZURE (M.), *Cours d'algèbre*, Cassini, 1997, p 71-83.

existe $\ell_n \in \mathbb{Z}$ tel que $a^{nk+1} = \ell_n r + a$. Alors la suite (ℓ_n) est strictement croissante (puisque $k > 0$) et les entiers $\ell_n r + a$ ont bien les mêmes facteurs premiers que a. $\lhd$

Pierre de Fermat est issu d'une famille de commerçants et exerce le métier de conseiller au Parlement de Toulouse. Pour lui, les mathématiques sont un hobby, ce qui n'a pas empêché ses travaux, dont peu furent publiés de son vivant, d'être à l'origine de développements féconds. Il s'intéresse particulièrement aux nombres premiers.

4.16. Nombres de Fermat

1. Déterminer une condition nécessaire sur $m \in \mathbb{N}$ pour que le nombre $2^m + 1$ soit premier.

On pose pour tout $n \in \mathbb{N}$, $\mathrm{F}_n = 2^{2^n} + 1$: F_n est le n-ième nombre de Fermat.

2. Vérifier que F_0, F_1, F_2, F_3 et F_4 sont des nombres premiers. Montrer que F_5 est divisible par 641 (pour cela, on observera que $641 = 5^4 + 2^4 = 1 + 5 \times 2^7$).

3. Prouver que si $n \neq m$, alors F_n et F_m sont premiers entre eux. Retrouver ainsi le théorème d'Euclide : il existe une infinité de nombres premiers.

4. Si p est un diviseur premier de F_n, établir que $p \equiv 1\ [2^{n+1}]$. Expliquer pourquoi on en vient naturellement à essayer 641 comme diviseur de F_5.

(École normale supérieure)

▷ **Solution.**

1. Il est nécessaire que m soit une puissance de 2. En effet, si ce n'est pas le cas on peut écrire $m = 2^\alpha k$ avec $k \geqslant 3$ impair et on a la factorisation suivante :

$$2^m + 1 = 2^{2^\alpha k} + 1 = (2^{2^\alpha})^k + 1 = (2^{2^\alpha} + 1)\left(\sum_{i=1}^{k-1} (2^{2^\alpha})^{k-i}(-1)^{i-1}\right)$$

de sorte que $2^n + 1 = (2^{2^\alpha})^m + 1$ est divisible par $2^{2^\alpha} + 1$ et n'est donc pas premier.

2. On observe que $\mathrm{F}_0 = 3$, $\mathrm{F}_1 = 5$, $\mathrm{F}_2 = 17$, $\mathrm{F}_3 = 257$ et $\mathrm{F}_4 = 65537$ sont bien des nombres premiers. On va montrer que F_5 est divisible par 641, donc non premier. Le très joli calcul qui suit est dû à Euler. On a $641 = 2^4 + 5^4 = 1 + 5 \times 2^7$ de sorte que $5 \times 2^7 \equiv -1\ [641]$. Élevons cette congruence à la puissance 4 : $5^4 \times 2^{28} \equiv 1\ [641]$. Comme, $641 = 5^4 + 2^4$ on a $5^4 \equiv -2^4\ [641]$ et donc $2^{32} \equiv -1\ [641]$. C'est le résultat voulu.

On a, plus précisément, $F_5 = 641 \times 6700417$.

3. Sans perte de généralité, on peut poser $m = n + k$ avec $k \geqslant 1$. On remarque alors que F_n divise $F_m - 2 = (2^{2^n})^{2^k} - 1$. Si p est un diviseur premier commun à F_n et F_m il en résulte que p doit diviser 2, *i.e.* être égal à 2. Or les nombres de Fermat sont impairs. Ils sont donc premiers entre eux deux à deux. Comme chaque F_n a au moins un diviseur premier, on en déduit qu'il existe une infinité de nombres premiers.

4. Soit p un diviseur premier de F_n. On se place dans $G = (\mathbb{Z}/p\mathbb{Z})^*$, groupe multiplicatif de cardinal $p-1$. Comme p est impair, $\overline{2} \in G$. On a : $\overline{2}^{2^n} = -\overline{1}$. En élevant cette égalité au carré, il vient : $\overline{2}^{2^{n+1}} = \overline{1}$. L'ordre de $\overline{2}$ dans G est donc un diviseur de 2^{n+1}, *i.e.* une puissance de 2. Cependant, si cet ordre valait 2^k avec $k < n+1$ on aurait : $\overline{2}^{2^n} = \overline{1}$. Ce qui n'est pas! Donc l'ordre de $\overline{2}$ est exactement 2^{n+1}. Cet ordre est un diviseur du cardinal de G, c'est-à-dire de $p-1$. Le résultat en découle. Pour $n = 5$ on doit avoir $p \equiv 1$ [64]. On teste donc les nombres premiers parmi 65, 129, 193, 257, 321, 385, 449, 513, 577, 641,... Après 193, 257, 449 et 577, on en vient très vite à essayer 641. ◁

Le résultat de la question 4 est du à Lucas. Fermat avait conjecturé la primalité de F_n *pour tout* n*, conjecture invalidée par Euler. En fait, on ne connaît pas d'autres* F_n *premiers que ceux vus ci-dessus. A l'heure actuelle (2007) on sait que* F_n *est composé pour* $5 \leqslant n \leqslant 30$ *et on connaît sa factorisation pour* $5 \leqslant n \leqslant 11$*. En revanche on ne dispose d'aucun facteur premier pour* F_{14}, F_{20}, F_{22} *et* F_{24}*. On conjecture actuellement qu'il n'y a qu'un nombre fini de* F_n *premiers.*

À la suite de Pierre de Fermat, la communauté mathématique montre un fort intérêt pour les nombres premiers. Euclide savait déjà qu'il y en a une infinité (sa preuve est celle proposée dans l'exercice suivant). Le XIX$^{\text{e}}$ *siècle aura été marqué par une concentration des efforts des mathématiciens en vue de démontrer le théorème des nombres premiers, conjecturé par Gauss : si* $\pi(n)$ *désigne le nombre d'entiers premiers inférieurs ou égaux à* n*, alors*

$$\pi(n) \underset{n\to+\infty}{\sim} \frac{n}{\ln n}.$$

On notera la raréfaction des nombres premiers à l'infini puisque $\lim_{n\to\infty} \dfrac{\pi(n)}{n} = 0$*. Si les travaux de Tchebychev et Riemann furent déterminants, c'est seulement en 1896 que de la Vallée-Poussin et Hadamard de manière séparée viennent à bout de la conjecture.*

Trouver des nombres premiers nouveaux est une activité qui n'intéresse pas seulement les mathématiciens. Les grands nombres premiers

sont utilisés dans certains algorithmes de cryptage des données. Beaucoup de grands nombres premiers connus sont des nombres de Mersenne $M_p = 2^p - 1$ *où p est premier (si $a^n - 1$ est premier, $a = 2$ et n est premier). En 1644, Mersenne affirma que M_p était premier pour $p = 2, 3, 5, 7, 13, 17, 19, 31, 67, 127, 257$ et composé pour les autres valeurs de p premier inférieures à 257. Mais en 1806, Pervasin et Seelhoff démontrèrent que M_{61} était premier. En 1876, Lucas établit un méthode efficace pour tester la primalité des M_p (et prouva ainsi que M_{127} était premier). En septembre 2006 le 44-ième nombre de Mersenne premier a été découvert : il s'agit de $M_{32582657}$ et il possède* 9808358 *chiffres.*

4.17. Infinité des nombres premiers congrus à 3 modulo 4

Montrer que l'ensemble $\mathcal{P}$ des nombres premiers est infini. Montrer qu'il en est de même de l'ensemble des nombres premiers congrus à 3 modulo 4.

(École polytechnique)

▷ **Solution.**

• Raisonnons par l'absurde et supposons qu'il n'existe qu'un nombre fini de nombres premiers. Notons-les $p_1, p_2, \ldots, p_n$. Considérons l'entier $N = p_1 p_2 \ldots p_n + 1$. Comme N est strictement supérieur à 1, il possède un diviseur premier p_k. Dans ces conditions, p_k divise $N - p_1 p_2 \ldots p_n = 1$, ce qui est impossible. Donc $\mathcal{P}$ est infini.

• Pour montrer qu'il existe une infinité de nombres premiers congrus à 3 modulo 4, raisonnons de nouveau par l'absurde et supposons qu'il n'en existe qu'un nombre fini n. Notons-les $p_1, p_2, \ldots, p_n$. Considérons cette fois l'entier $N = 4p_1 p_2 \ldots p_n - 1 \geqslant 2$.

Aucun des p_k ne divise N, sinon il diviserait 1. Comme N est impair, tout diviseur premier de N ($N > 1$) est impair et n'est pas congru à 3 modulo 4 d'après ce qui précède ; il est donc congru à 1 modulo 4 et N est donc lui-même congru à 1 modulo 4. Or, manifestement, N est congru à 3 modulo 4. C'est la contradiction cherchée.

Conclusion. L'ensemble des nombres premiers congrus à 3 modulo 4 est infini. ◁

Ce dernier résultat est en fait un cas particulier du théorème de la progression arithmétique de Dirichlet qui affirme l'existence d'une infinité de nombres premiers de la forme $an + b$ lorsque a et b sont premiers entre eux. La preuve de ce théorème a fait l'objet du problème de six heures posé en 1993 aux ENS. L'exercice suivant traite le cas où $b = 1$.

4.18. Version faible du théorème de la progression arithmétique de Dirichlet (1837)

On note $\Phi_1 = X - 1$ et pour $n \geqslant 2$, $\Phi_n = \prod_{\substack{1 \leqslant k \leqslant n \\ \text{pgcd}(k,n)=1}} \left(X - e^{\frac{2ik\pi}{n}}\right)$, le n-ième polynôme cyclotomique.

1. Montrer que Φ_n est à coefficients entiers pour tout $n \in \mathbb{N}^*$.

2. Que peut-on dire d'un nombre premier p qui divise $\Phi_n(a)$, où $a \in \mathbb{Z}$, mais aucun des $\Phi_d(a)$, où d décrit l'ensemble des diviseurs stricts de n ?

3. En déduire que pour $n \geqslant 1$ fixé, il existe une infinité de nombres premiers de la forme $\lambda n + 1$ avec λ entier.

(École normale supérieure)

Solution.

1. • Montrons que $X^n - 1 = \prod_{d|n} \Phi_d$. On sait que $X^n - 1 = \prod_{\ell=1}^{n} \left(X - e^{\frac{2i\ell\pi}{n}}\right)$. Notons pour $d \geqslant 1$, P_d l'ensemble des racines primitives d-ièmes de l'unité et U_d l'ensemble des racines d-ièmes de l'unité. On a, par définition, $\Phi_n = \prod_{\xi \in P_n} (X - \xi)$. Si $\xi \in U_n$, l'ordre de ξ est un diviseur d de n et alors $\xi \in P_d$. Par conséquent, U_n est réunion disjointe des P_d pour d divisant n. D'où il résulte

$$X^n - 1 = \prod_{\xi \in U_n} (X - \xi) = \prod_{d|n} \left(\prod_{\xi \in P_d} (X - \xi) \right) = \prod_{d|n} \Phi_d.$$

On obtient le résultat voulu

$$\boxed{X^n - 1 = \prod_{d|n} \Phi_d}.$$

On retrouve ainsi un résultat classique sur l'indicatrice d'Euler φ : pour tout entier $n \geqslant 1$, $\varphi(n)$ est le degré de Φ_n i.e. *le nombre d'entiers k compris entre 1 et n, premiers avec n. En considérant les degrés, l'identité précédente donne*

$$\boxed{n = \sum_{d|n} \varphi(d)}.$$

• Nous allons établir que Φ_n est à coefficients entiers par récurrence sur $n \geqslant 1$ en utilisant le résultat suivant :

Lemme. *Soit* A *et* B *deux polynômes à coefficients entiers,* B *étant non nul unitaire. Alors* Q *et* R, *le quotient et le reste de la division euclidienne de* A *par* B *dans* $\mathbb{C}[X]$ *sont aussi à coefficients entiers.*

Démonstration. Le détail de la démonstration est laissé au lecteur ; on peut l'obtenir par une récurrence sur le degré de A (comme pour la division dans K[X] avec K corps) ou bien en constatant que dans les opérations de l'algorithme de division euclidienne, seuls des entiers interviennent.

Établissons maintenant par récurrence sur n que Φ_n est à coefficients entiers.

$*$ C'est vrai, par définition, si $n = 1$.

$*$ Si $n \geqslant 2$, Φ_n est le quotient dans $\mathbb{C}[X]$ de $X^n - 1$ par B, où B est égal au produit des Φ_d, où d est un diviseur strict de n. Si on suppose la propriété vraie pour les entiers $\leqslant n-1$, chacun de ces Φ_d est à coefficients entiers et unitaires par définition, donc B est aussi à coefficients entiers et unitaire. En vertu du lemme, Φ_n est à coefficients entiers.

2. Soit p premier vérifiant l'hypothèse. Comme p divise $\Phi_n(a)$, il divise aussi $a^n - 1$. Ainsi, l'ordre de $\overline{a}$ dans le groupe multiplicatif $(\mathbb{Z}/p\mathbb{Z})^*$ divise n. Montrons que cet ordre est exactement n. Si d divise n, $d < n$, on a dans $\mathbb{Z}/p\mathbb{Z}$

$$\overline{a}^d - 1 = \prod_{d'|d} \overline{\Phi_{d'}(a)}.$$

Or si d' divise d, d' divise aussi n et par hypothèse, $\overline{\Phi_{d'}(a)} \neq 0$. Comme $\mathbb{Z}/p\mathbb{Z}$ est un corps, le produit de ces éléments non nuls est également non nul, si bien que $\overline{a}^d \neq 1$. L'ordre de $\overline{a}$ est donc n. Comme cet ordre divise $p - 1$ d'après le théorème de Lagrange, p est de la forme $\lambda n + 1$ avec λ entier.

3. Raisonnons par l'absurde et supposons qu'il n'existe qu'un nombre fini d'entiers premiers congrus à 1 modulo n, soit $p_1, \ldots, p_q$. La question précédente, si on arrive à trouver a et p vérifiant les hypothèses, assure que p est congru à 1 modulo n. Ce sera insuffisant pour aboutir à une contradiction, p pouvant être alors un des p_i. Pour éviter cela, on va changer n en $N = np_1p_2 \ldots p_q$. Si p est congru à 1 (mod N), p ne peut être un des p_i et pourtant, il est congru à 1 (mod n).

Il faut donc trouver $a \in \mathbb{Z}$ et p premier, tels que p divise $\Phi_N(a)$, mais aucun des $\Phi_d(a)$ pour $d|N$, $d < N$. On note $B = \prod_{d|,d<N} \Phi_d$. Le problème est donc de trouver $a \in \mathbb{Z}$ et p premier tels que p divise $\Phi_N(a)$ et ne divise pas $B(a)$.

Le polynôme B est premier avec Φ_N dans $\mathbb{C}[X]$ (en effet, ils sont scindés sur $\mathbb{C}$ et n'ont aucune racine commune), donc dans $\mathbb{Q}[X]$, puisque

ces polynômes sont à coefficients rationnels et que le pgcd est invariant par extension de corps (l'algorithme d'Euclide s'écrit de la même manière dans $\mathbb{C}[X]$ et dans $\mathbb{Q}[X]$).

D'après le théorème de Bezout, il existe donc $(U, V) \in \mathbb{Q}[X]^2$ tel que $1 = U\Phi_N + VB$. Il existe $a \in \mathbb{Z}$ tel que $U' = aU$ et $V' = aV$ appartiennent à $\mathbb{Z}[X]$ (il suffit de prendre un multiple du ppcm des dénominateurs des coefficients qui apparaissent dans U et V). Comme $\Phi_N \neq 0$ et $\Phi_N \neq \pm 1$, on peut même choisir a tel que $\Phi_N(a) \neq 0$ et $\Phi_N(a) \neq \pm 1$, étant donnée l'infinité de $a \in \mathbb{Z}$ vérifiant $aU \in \mathbb{Z}[X]$ et $aV \in \mathbb{Z}[X]$ (ceci en vue d'avoir des nombres premiers qui divisent $\Phi_N(a)$). On a donc

$$a = U'\Phi_N + V'B \quad \text{et en particulier} \quad a = U'(a)\Phi_N(a) + V'(a)B(a). \quad (*)$$

Soit p un nombre premier divisant $\Phi_N(a)$. Alors p divise $a^N - 1$, car Φ_N divise $X^N - 1$ dans $\mathbb{Z}[X]$. Dans $\mathbb{Z}/p\mathbb{Z}$, $\overline{a}^N = 1$ et donc $\overline{a}$ est inversible, ce qui signifie que a est premier avec p. Si p divisait $B(a)$, il diviserait a, d'après $(*)$, ce qui est exclu.

On est donc dans les hypothèses de la question précédente : p est congru à 1 modulo N, et donc modulo n, avec p forcément distinct des p_i, pour $1 \leqslant i \leqslant q$. C'est la contradiction voulue. ◁

4.19. Plus petit nombre premier ne divisant pas n

1. Montrer que tout entier $n > 6$ s'écrit comme somme de deux entiers premiers entre eux, strictement supérieurs à 1.

2. Soit $(p_n)_{n\geqslant 1}$ la suite strictement croissante des nombres premiers. Montrer que pour tout $k > 2$, on a $p_{k+1} + p_{k+2} \leqslant p_1 p_2 \dots p_k$.

3. Pour $n \in \mathbb{N}^*$, on note q_n le plus petit nombre premier ne divisant pas n. Montrer que la suite $\dfrac{q_n}{n}$ tend vers 0.

(École normale supérieure)

▷ **Solution.**

1. Si n est impair, on peut écrire $n = 2 + (n-2)$ et $\operatorname{pgcd}(2, n-2) = 1$.

Si $n \equiv 0 \pmod 4$, il s'écrit $n = 4k$ avec $k \geqslant 2$. Alors la décomposition $n = (2k-1) + (2k+1)$ convient.

Enfin, si $n \equiv 2 \pmod 4$, il s'écrit $n = 4k + 2$ avec $k \geqslant 2$. Alors la décomposition $n = (2k+3) + (2k-1)$ est aussi du type souhaité (si p premier divise $2k+3$ et $2k-1$ il divise la différence c'est-à-dire 4, ce qui est impossible, car les entiers sont impairs).

Par contre, $6 = 2 + 4 = 3 + 3$ n'a pas de décomposition de ce type.

2. Pour $k = 3$ on a $p_1 p_2 p_3 = 2 \times 3 \times 5 = 30$ et $p_4 + p_5 = 7 + 11 = 18$.

Posons, pour $k \geqslant 3$, $n = p_1 p_2 \dots p_k$. Comme $n > 6$, on peut l'écrire sous la forme $n = a + b$, $a > 1$, $b > 1$, $a \wedge b = 1$. Aucun des nombres p_i,

où $1 \leqslant i \leqslant k$, ne peut diviser a (car il diviserait aussi b). Les facteurs premiers de a sont donc supérieurs ou égaux à p_{k+1}. De même pour b. Or, a et b sont premiers entre eux et n'ont donc aucun diviseur premier commun. Il en résulte que $n = a + b \geqslant p_{k+1} + p_{k+2}$.

3. Pour $n \geqslant 2$, on note k_n l'unique entier tel que

$$p_1 p_2 \dots p_{k_n} \leqslant n < p_1 p_2 \dots p_{k_n+1}.$$

Il est alors évident que $q_n \leqslant p_{k_n+1}$. Ainsi, on a pour $n \geqslant 2 \cdot 3 \cdot 5 \cdot 7 = 210$,

$$\frac{q_n}{n} \leqslant \frac{p_{k_n+1}}{p_1 p_2 \dots p_{k_n}} \leqslant \frac{1}{p_{k_n}},$$

la dernière inégalité résultant de la question 2 :

$$p_{k+1} < p_{k_n+1} + p_{k_n} \leqslant p_1 \dots p_{k_n-1}$$

pour $k_n \geqslant 4$, c'est-à-dire pour $n \geqslant 210$. Comme k_n tend vers l'infini lorsque n tend vers l'infini, le résultat est prouvé. ◁

Les exercices suivants sont centrés sur la factorialité de $\mathbb{Z}$: tout entier $n \geqslant 2$ s'écrit de manière unique (à l'ordre des facteurs près) sous la forme $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ où $p_1, \dots, p_k$ sont des nombres premiers deux à deux distincts et $\alpha_1, \dots, \alpha_k$ des entiers naturels non nuls.

4.20. Théorème de Kurschak (1918)

Pour quelles valeurs entières $n \geqslant m$ a-t-on $\displaystyle\sum_{i=m}^{n} \frac{1}{i} \in \mathbb{N}$?

(École normale supérieure)

▷ **Solution.**

On obtient un entier pour $n = m = 1$ et on va voir que c'est le seul cas. Supposons $n \geqslant 2$. L'idée consiste à regarder la valuation 2-adique des entiers entre m et n. On peut supposer $m < n$ car $\dfrac{1}{n}$ n'est pas entier pour $n \geqslant 2$. Soit $\alpha = \max\{\nu_2(k),\ m \leqslant k \leqslant n\}$. On a $\alpha \geqslant 1$, car il y a au moins un entier pair entre m et n. En fait, le point essentiel est que la valuation 2-adique maximale α n'est atteinte qu'une et une seule fois. En effet, supposons qu'il existe deux entiers k, k' avec $m \leqslant k < k' \leqslant n$ et $k = 2^\alpha(2r+1)$, $k' = 2^\alpha(2s+1)$. Alors $2^\alpha(2r+2) = 2^{\alpha+1}(r+1)$ appartient à $[\![m, n]\!]$ et est de valuation 2-adique supérieure ou égale à $\alpha + 1$ ce qui contredit la définition de α. Il en résulte que le représentant irréductible de la somme $\displaystyle\sum_{i=m}^{n} \frac{1}{i}$ est de la forme $\dfrac{\mathrm{A}}{2^\alpha \mathrm{B}}$ où A, B sont impairs. Ce qui prouve que la somme en question n'est pas un entier. ◁

Le résultat a d'abord été prouvé en 1915 par Taeisinger dans le cas $n = 1$ et généralisé en 1918 par Kurschak. En 1932, le célèbre Paul Erdös a prouvé que le résultat se généralise pour des entiers formant une progression arithmétique quelconque.

4.21. Théorème de Legendre (1808)

Soit n un entier supérieur à 2 et p un nombre premier. Montrer que la valuation p-adique de $n!$ est égale à $\sum_{k=1}^{+\infty} \mathrm{E}\left(\frac{n}{p^k}\right)$.

(École polytechnique)

▷ **Solution.**

Notons qu'en fait la somme considérée est finie. Parmi les entiers entre 1 à n, le nombre de multiples de p qui ne sont pas multiples de p^2 est $\mathrm{E}\left(\frac{n}{p}\right) - \mathrm{E}\left(\frac{n}{p^2}\right)$. Chacun de ces entiers amène une contribution de 1 dans la valuation p-adique de $n!$. Le nombre de multiples de p^2 non multiples de p^3 est $\mathrm{E}\left(\frac{n}{p^2}\right) - \mathrm{E}\left(\frac{n}{p^3}\right)$. Chacun amène une contribution de 2. On continue ... Si on note q le plus grand entier tel que $p^q \leqslant n$, la valuation p-adique de $n!$ est donc

$$\nu_p(n!) = \sum_{k=1}^{q} k\left(\mathrm{E}\left(\frac{n}{p^k}\right) - \mathrm{E}\left(\frac{n}{p^{k+1}}\right)\right) = \sum_{k=1}^{q} \mathrm{E}\left(\frac{n}{p^k}\right) = \sum_{k=1}^{+\infty} \mathrm{E}\left(\frac{n}{p^k}\right).$$

C'est le résultat demandé. ◁

Ce calcul permet par exemple de déterminer le nombre de zéros placés à droite de l'écriture décimale de $\mathrm{N} = 2007!$. *Il s'agit de trouver la plus grande puissance de* 10 *qui divise cet entier. Comme* $\nu_5(\mathrm{N}) \leqslant \nu_2(\mathrm{N})$ *le nombre cherché est égal à la valuation* 5*-adique de* $2007!$. *Elle vaut*

$$\mathrm{E}\left(\frac{2007}{5}\right) + \mathrm{E}\left(\frac{2007}{25}\right) + \mathrm{E}\left(\frac{2007}{125}\right) + \mathrm{E}\left(\frac{2007}{625}\right) = 500.$$

4.22. Un produit de trois entiers consécutifs n'est jamais une puissance k-ième

Soit $k \geqslant 2$. Montrer que le produit de trois entiers naturels non nuls consécutifs ne peut pas être une puissance k-ième.

(École normale supérieure)

▷ **Solution.**

On cherche à montrer que l'équation $n(n+1)(n+2) = x^k$ n'a pas de solution non nulle. Mieux vaut écrire cette équation sous la forme plus symétrique $(n-1)n(n+1) = x^k$. Supposons par l'absurde qu'il existe une solution (x, n) avec $n \geqslant 2$. Les entiers n et n^2-1 sont premiers entre eux. Or on a le résultat essentiel suivant, conséquence de la factorialité de $\mathbb{Z}$:

Lemme. *Soit a et b deux entiers naturels non nuls premiers entre eux, $k \geqslant 2$. On suppose qu'il existe $c \in \mathbb{N}$ tel que $ab = c^k$. Alors a et b sont eux aussi des puissances k-ièmes.*

Démonstration.

Écrivons la décomposition de a, b et c en produit de facteurs premiers :

$$a = \prod_{p \text{ premier}} p^{\alpha_p}, \quad b = \prod_{p \text{ premier}} p^{\beta_p}, \quad c = \prod_{p \text{ premier}} p^{\gamma_p},$$

où les α_p, β_p et γ_p sont des familles d'entiers à support fini. Comme $ab = c^k$, on obtient, par unicité de la décomposition, $\alpha_p + \beta_p = k\gamma_p$ pour tout p premier. Puisque a et b sont premiers entre eux, on a $\alpha_p \beta_p = 0$ pour tout p. Il en résulte que pour tout p premier α_p et β_p sont divisibles par k. Ainsi a et b sont des puissances k-ièmes. ◇

Il existe d'après le lemme y et z dans $\mathbb{N}^*$ tels que $n = y^k$ et $n^2-1 = z^k$. D'où $y^{2k} - z^k = 1$. Or deux puissances k-ièmes non nulles consécutives diffèrent au moins de k puisque, pour $u \geqslant 1$, $(u+1)^k - u^k \geqslant ku \geqslant k \geqslant 2$. La relation $(y^2)^k - z^k = 1$ implique donc $z = 0$ et $y = 1$, d'où $n = 1$. Il n'y a donc pas de solution avec $n \geqslant 2$. ◁

Erdös et Selfridge ont démontré en 1975 une conjecture vieille de plus de 150 ans : pour tout $m \geqslant 2$, un produit de m entiers consécutifs n'est jamais une puissance k-ième[2].

4.23. Théorème de Palfy-Erdös

Pour tout $x \in \mathbb{Z}$ et tout premier p, on note x_p le résidu de x modulo p. Soit a et b deux entiers positifs tels que pour tout p premier, $a_p \leqslant b_p$.

1. Montrer que $a \leqslant b$.

2. ERDÖS (P.) & SELFRIDGE (J.L.), *The product of consecutive integers is never a power*, Illinois J. of Math. 19, 1975, p. 292-301.

On désire montrer que $a = b$. Raisonnons par l'absurde et supposons $a < b$. On note $A = 1 \cdot 2 \dots (a-1)a$ et $B = (b-a+1) \dots (b-1)b$. Pour p premier et $k \geqslant 1$, on note $r(p^k)$ (resp. $s(p^k)$) le nombre de facteurs de A (resp. B) divisibles par p^k.

2. Montrer que $s(p^k)$ est égal à $r(p^k)$ ou $r(p^k)+1$.

3. Montrer en utilisant l'hypothèse que $r(p) = s(p)$. En déduire que si $p > a$, alors $r(p^k) = s(p^k) = 0$.

4. On note $t(p)$ le plus grand entier k tel que $s(p^k) > 0$. Prouver que $C_b^a = \dfrac{B}{A}$ divise $\prod\limits_{p \leqslant a} p^{t(p)-1}$ et en déduire que

$$\frac{(b-a+1)\dots(b-1)b}{\prod\limits_{p\leqslant a} p^{t(p)}} \text{ divise } \frac{1.2\dots(a-1)a}{\prod\limits_{p\leqslant a} p}.$$

5. On suppose de plus $a \leqslant \dfrac{b}{2}$. Aboutir à une contradiction.

6. Prouver, en utilisant 5, que l'on aboutit également à une contradiction si $\dfrac{b}{2} < a < b$.

(École normale supérieure)

Solution.

1. Soit p premier, $p > a$ et $p > b$. Alors on a $a = a_p \leqslant b_p = b$.

2. L'entier $r(p^k)$ est le nombre d'entiers d vérifiant $1 \leqslant dp^k \leqslant a$. Il en résulte que $r(p^k) = E\left(\dfrac{a}{p^k}\right)$ et $\dfrac{a}{p^k} - 1 < r(p^k) \leqslant \dfrac{a}{p^k}$. On peut écrire, d'autre part,

$$s(p^k) = \text{Card}\{d \in \mathbb{N}, b-a < dp^k \leqslant b\} = \text{Card}\left\{d \in \mathbb{N}, \frac{b-a}{p^k} < d \leqslant \frac{b}{p^k}\right\}$$
$$= E\left(\frac{b}{p^k}\right) - E\left(\frac{b-a}{p^k}\right).$$

Encadrons $s(p^k)$:

$$\begin{aligned} \frac{b}{p^k} - 1 - \frac{b-a}{p^k} < \quad & s(p^k) \quad < \frac{b}{p^k} - \left(\frac{b-a}{p^k} - 1\right), \\ \frac{a}{p^k} - 1 < \quad & s(p^k) \quad < \frac{a}{p^k} + 1, \\ r(p^k) - 1 < \quad & s(p^k) \quad < r(p^k) + 2. \end{aligned}$$

Comme dans cette dernière inégalité, il ne figure que des entiers, $s(p^k)$ vaut $r(p^k)$ ou $r(p^k)+1$.

3. Considérons maintenant les deux suites finies $S'_1 = (a, a-1, \dots, 1)$ et $S'_2 = (b, b-1, \dots, b-a+1)$. Le premier multiple de p dans S'_1 est $a - a_p$ et le premier multiple de p dans S'_2 est $b - b_p$. Donc, comme $a_p \leqslant b_p$, celui de S'_1, qui est le (a_p+1)-ième terme de la suite S'_1 arrive avant ou en même temps que celui de S'_2, qui est le (b_p+1)-ième terme de la suite S'_2. Les deux suites étant de même longueur, cela entraîne $r(p) \geqslant s(p)$ et finalement $r(p) = s(p)$, en utilisant la question précédente.

Ainsi, si $p > a$, $r(p)$ est nul et donc $s(p)$ aussi. Donc aucun facteur de A et de B n'est divisible par p. Il s'ensuit que $r(p^k) = s(p^k) = 0$ pour tout $k \geqslant 1$.

4. On peut écrire

$$\mathrm{A} = \prod_{p \text{ premier}} p^{\sum_{k=1}^{\infty} r(p^k)} \text{ et } \mathrm{B} = \prod_{p \text{ premier}} p^{\sum_{k=1}^{\infty} s(p^k)}$$

et même, compte tenu de la question précédente,

$$\mathrm{A} = \prod_{p \leqslant a} p^{\sum_{k=1}^{\infty} r(p^k)} \text{ et } \mathrm{B} = \prod_{p \leqslant a} p^{\sum_{k=1}^{\infty} s(p^k)},$$

p en indice désignant toujours un nombre premier. On en déduit que

$$C_b^a = \frac{\mathrm{B}}{\mathrm{A}} = \prod_{p \leqslant a} p^{\sum_{k=1}^{\infty} s(p^k) - r(p^k)}.$$

De la question 3, on déduit l'égalité

$$\sum_{k=1}^{\infty} s(p^k) - r(p^k) = \sum_{k=2}^{\infty} s(p^k) - r(p^k).$$

Si $k > t(p)$, on a $0 = s(p^k) \geqslant r(p^k)$ et $r(p^k) = 0$. La somme devient donc

$$\sum_{k=1}^{\infty} s(p^k) - r(p^k) = \sum_{k=2}^{t(p)} s(p^k) - r(p^k) \leqslant \sum_{k=2}^{t(p)} 1 = t(p) - 1,$$

la dernière inégalité résultant de la question 2. Par conséquent, l'entier $\dfrac{\mathrm{B}}{\mathrm{A}}$ divise $\prod\limits_{p \leqslant a} p^{t(p)-1}$. Il existe donc $\lambda \in \mathbb{N}$ tel que

$$\lambda \frac{\mathrm{B}}{\mathrm{A}} = \prod_{p \leqslant a} p^{t(p)-1} \quad \text{c'est-à-dire} \quad \lambda \left(\frac{\mathrm{B}}{\prod\limits_{p \leqslant a} p^{t(p)}} \right) = \frac{\mathrm{A}}{\prod\limits_{p \leqslant a} p}.$$

Mais $\dfrac{B}{\prod\limits_{p\leqslant a} p^{t(p)}}$ et $\dfrac{A}{\prod\limits_{p\leqslant a} p}$ sont des entiers (car $p^{t(p)}$ divise l'un des facteurs de B, pour tout p). On est donc en droit d'écrire que

$$\frac{(b-a+1)\dots(b-1)b}{\prod\limits_{p\leqslant a} p^{t(p)}} \text{ divise } \frac{1.2\dots(a-1)a}{\prod\limits_{p\leqslant a} p}.$$

5. On note $\pi(a)$ le nombre d'entiers premiers inférieurs ou égaux à a. Considérons le terme de droite dans la relation ci-dessus $\dfrac{1.2\dots(a-1)a}{\prod\limits_{p\leqslant a} p}$. Après simplification, on obtient un produit de $a-\pi(a)$ facteurs tous inférieurs ou égaux à a. De même, après simplification, le terme de gauche de la relation $\dfrac{(b-a+1)\dots(b-1)b}{\prod\limits_{p\leqslant a} p^{t(p)}}$ «contient» $a-\pi(a)$ facteurs supérieurs ou égaux à $b-a+1$ (tous les termes du numérateur qui ne sont pas divisible par un $p^{t(p)}$) avec, par hypothèse, $b-a+1 \geqslant 2a-a+1 > a$. On en déduit que

$$\frac{(b-a+1)\dots(b-1)b}{\prod\limits_{p\leqslant a} p^{t(p)}} > a^{a-\pi(a)} \geqslant \frac{1.2\dots(a-1)a}{\prod\limits_{p\leqslant a} p},$$

ce qui est manifestement contradictoire avec le résultat de la question 4.

6. Supposons $\dfrac{b}{2} < a < b$ et posons $c = b-a$. On a $0 < c < \dfrac{b}{2}$. On obtient, pour tout p premier, $c \equiv b_p - a_p\ [p]$ et $0 \leqslant b_p - a_p < p$ et donc $c_p = b_p - a_p \leqslant b_p$. Ceci est impossible d'après 5. On conclut donc que nécessairement $a = b$. ◁

Les exercices suivants sont consacrés aux propriétés arithmétiques des coefficients binomiaux.

4.24. Valuation p-adique de $C_{p^n}^k$

Soit p un nombre premier, $n \in \mathbb{N}^*$ et $k \in [\![1, p^n-1]\!]$. Quelle est la plus grande puissance de p qui divise $C_{p^n}^k$?

(École normale supérieure)

Solution.

La question consiste à chercher la valuation p-adique de $C_{p^n}^k$. On écrit

$$k!C_{p^n}^k = p^n(p^n-1)(p^n-2)\dots(p^n-(k-1)).$$

On note ν_p la valuation p-adique. Pour tout entier $\lambda \in [\![1, k-1]\!]$, on a

$$\nu_p(p^n - \lambda) = \nu_p(\lambda),$$

car si $\lambda = p^\alpha u$ avec u non divisible par p et $\alpha < n$ (puisque $\lambda < p^n$), on a $p^n - \lambda = p^\alpha(p^{n-\alpha} - u)$ et $p^{n-\alpha} - u$ n'est pas divisible par p. En passant à la valuation dans l'égalité ci-dessus on obtient, puisque la valuation d'un produit est la somme des valuations des différents termes,

$$\nu_p(k!) + \nu_p(\mathrm{C}_{p^n}^k) = n + \nu_p((k-1)!)$$

et finalement

$$\boxed{\nu_p(\mathrm{C}_{p^n}^k) = n - \nu_p(k)}. \quad \lhd$$

L'énoncé suivant regroupe trois exercices posés indépendamment.

4.25. Congruences de Lucas (1878)

Soit p un nombre premier et n, k des entiers naturels.

1. On suppose $n \geqslant 2$. Montrer l'équivalence entre les assertions suivantes :

(i) n est premier ;

(ii) $\forall i \in [\![1, n-1]\!]$, n divise C_n^i.

2. On écrit n et k en base p : $n = n_0 + n_1 p + \cdots + n_j p^j$ et $k = k_0 + k_1 p + \cdots + k_j p^j$ (avec le même indice j quitte à compléter avec des zéros). Établir le théorème de Lucas :

$$\mathrm{C}_n^k \equiv \mathrm{C}_{n_0}^{k_0} \mathrm{C}_{n_1}^{k_1} \ldots \mathrm{C}_{n_j}^{k_j} \pmod{p}.$$

(avec la convention habituelle que $\mathrm{C}_n^k = 0$ si $k > n$).

3. Montrer que le nombre d_n de coefficients binomiaux impairs sur la n-ième ligne du triangle de Pascal (c'est-à-dire parmi les C_n^k, $0 \leqslant k \leqslant n$) est une puissance de 2.

(École normale supérieure)

▷ **Solution.**

1. Supposons n premier. On a pour tout $i \in [\![1, n-1]\!]$, $i\mathrm{C}_n^i = n\mathrm{C}_{n-1}^{i-1}$. Il en résulte que n divise $i\mathrm{C}_n^i$. Par le lemme de Gauss, n divise i ou n divise C_n^i. Le premier cas étant visiblement exclu, l'assertion (ii) est prouvée.

Réciproquement, supposons que (ii) est vérifiée. Soit d un diviseur positif de n. On a comme ci-dessus, $\mathrm{C}_n^d = \dfrac{n}{d}\mathrm{C}_{n-1}^{d-1} \equiv 0 \pmod{n}$. Cela invite à regarder le résidu des C_{n-1}^r modulo n. On obtient, pour $i = 1$,

$C^1_{n-1} = n-1 \equiv -1 \pmod{n}$ et la formule de Pascal, $C^i_{n-1} + C^{i+1}_{n-1} = C^{i+1}_n$ pour $1 \leqslant i \leqslant n-2$, donne $C^{i+1}_{n-1} \equiv -C^i_{n-1} \pmod{n}$, compte-tenu de (ii). Il en résulte que pour tout $i \in [\![1, n-1]\!]$, on a $C^i_{n-1} \equiv (-1)^i \pmod{n}$ et donc $\dfrac{n}{d} C^{d-1}_{n-1} \equiv \dfrac{n}{d}(-1)^d \equiv 0 \pmod{n}$, ce qui n'est possible que si $d = 1$. L'entier n est donc premier.

2. Il résulte de la question 1 et de la formule du binôme de Newton que $(1+X)^p = 1 + X^p$ dans $(\mathbb{Z}/p\mathbb{Z})[X]$. Pour $a \in \mathbb{N}$, $a < p$, on a donc dans $(\mathbb{Z}/p\mathbb{Z})[X]$,

$$(1+X)^{np+a} = (1+X^p)^n (1+X)^a.$$

Si $b \in \mathbb{N}$, $b < p$, on obtient, en identifiant les coefficients de X^{kp+b} dans les deux expressions de ce polynôme, la congruence

$$C^{kp+b}_{np+a} \equiv C^k_n C^b_a \pmod{p},$$

ceci étant valable même si $k > n$ ou $b > a$, avec la convention indiquée dans l'énoncé. En opérant la division euclidienne de n (resp. k) par p, on peut écrire $n = n_0 + pq$ et $k = k_0 + pq'$ avec $(q, q') \in \mathbb{N}^*$. La formule précédente donne

$$C^k_n \equiv C^{k_0}_{n_0} C^{q'}_q \pmod{p}.$$

On réitère le même procédé en divisant q et q' par p, ce qui fait apparaître comme restes n_1 et k_1 et, de proche en proche, après j itérations, on obtient

$$C^k_n \equiv C^{k_0}_{n_0} C^{k_1}_{n_1} \dots C^{k_j}_{n_j} \pmod{p}.$$

3. Soit $k \in [\![0, n]\!]$. Écrivons n et k en base 2 : $n = n_0 + 2n_1 + \dots + 2^j n_j$ et $k = k_0 + 2k_1 + \dots + 2^j k_j$, où $n_i, k_j \in \{0, 1\}$ pour tout $0 \leqslant i \leqslant j$. Puisque, d'après la question 2, on a $C^k_n \equiv C^{k_0}_{n_0} C^{k_1}_{n_1} \dots C^{k_j}_{n_j} \pmod{2}$, pour que C^k_n soit impair, *i.e.* $C^k_n \equiv 1 \pmod{2}$, il faut et il suffit que $C^{k_i}_{n_i} \equiv 1 \pmod{2}$ pour tout i. C'est le cas si, et seulement si, $0 \leqslant k_i \leqslant n_i$ pour tout i.

On obtient donc $d_n = (n_0+1)(n_1+1)\dots(n_j+1) = 2^{c_n}$, où c_n est le nombre de chiffres 1 dans l'écriture binaire de n. Par exemple, si n est une puissance de 2, il n'y a que 2 coefficients binomiaux impairs sur la n-ième ligne du triangle de Pascal (les extrémités, évidemment). ◁

4.26. Un problème de congruence

Soit p premier. Montrer que $\displaystyle\sum_{k=0}^{p} C^k_p C^k_{p+k} \equiv 2^p + 1 \pmod{p^2}$.

(École polytechnique)

▷ **Solution.**

Posons $S_p = \sum_{k=0}^{p} C_p^k C_{p+k}^k$. Il est classique que $C_p^k \equiv 0 \pmod{p}$ pour tout $k \in [\![1, p-1]\!]$ (le lecteur trouvera ce résultat dans la question 1 de l'exercice précédent). Étudions le résidu modulo p de C_{p+k}^k. On a, pour tout $k \in [\![1, p-1]\!]$,

$$k! C_{p+k}^k = (p+k)(p+k-1)\dots(p+1) \equiv k! \pmod{p}.$$

Il en résulte que p divise $k!(C_{p+k}^k - 1)$ et comme $k!$ et p sont premiers entre eux, p divise $C_{p+k}^k - 1$. Ainsi p^2 divise $C_p^k(C_{p+k}^k - 1)$, ce qui équivaut à $C_p^k C_{p+k}^k \equiv C_p^k \pmod{p^2}$. On a donc

$$S_p \equiv 1 + C_{2p}^p + \sum_{k=1}^{p-1} C_p^k \equiv 1 + C_{2p}^p + 2^p - 2 \pmod{p^2},$$

la dernière congruence résultant de la formule du binôme de Newton.

Pour terminer, on est ramené à prouver que $C_{2p}^p \equiv 2 \pmod{p^2}$. Cela est vrai pour $p = 2$ puisque $C_4^2 = 6$. Dans la suite, on supposera que p est un nombre premier impair. On écrit

$$C_{2p}^p - 2 = \frac{(2p)(2p-1)\dots(p+1)}{p!} - 2 = \frac{2}{(p-1)!}\left(\prod_{k=1}^{p-1}(p+k) - (p-1)!\right)$$

Comme p est différent de 2 et est premier avec $(p-1)!$, il nous faut prouver que $\prod_{k=1}^{p-1}(p+k) - (p-1)! = Q(p) - (p-1)!$ (où $Q(X)$ est le polynôme $\prod_{k=1}^{p-1}(X+k) \in \mathbb{Z}[X]$) est divisible par p^2.

Si on écrit $Q(X) = X^{p-1} + a_{p-2}X^{p-2} + \dots + a_1 X + a_0$, on a clairement $Q(p) \equiv a_0 + a_1 p \pmod{p^2}$. Comme $a_0 = (p-1)!$ il suffit donc de prouver que a_1 est nul modulo p. En notant $\overline{Q} \in \mathbb{Z}/p\mathbb{Z}[X]$ la réduction modulo p de Q, on a

$$\overline{Q} = \prod_{k=1}^{p-1}(X + \overline{k}) = \prod_{k=1}^{p-1}(X - \overline{p-k}) = \prod_{k=1}^{p-1}(X - \overline{k}).$$

Les racines de $\overline{Q}$ sont donc exactement les éléments non nuls de $\mathbb{Z}/p\mathbb{Z}$ et ces racines sont simples. Le polynôme $X^{p-1} - \overline{1}$ a le même degré que $\overline{Q}$ et s'annule aussi en tout $\overline{k}$, pour $1 \leqslant k \leqslant p-1$, d'après le petit théorème de Fermat. Il est donc égal à $\overline{Q}$.

De l'égalité $\overline{Q} = X^{p-1} + \overline{a}_{p-2}X^{p-2} + \dots + \overline{a}_1 X + \overline{a_0} = X^{p-1} - \overline{1}$, on déduit $\overline{a}_1 = 0$, *i.e.* p divise a_1. ◁

4.27. Le problème de Ducci

Pour tout $n \in \mathbb{N}$, on écrit $n = \sum_{q=0}^{+\infty} \varepsilon_q(n) 2^q$ où pour tout $q \geqslant 0$, $\varepsilon_q(n) \in \{0, 1\}$ et $\varepsilon_q(n)$ nul pour q assez grand : il s'agit de l'écriture de n en numération binaire.

1. Pour $n \geqslant 1$, on note $s(n) = \sum_{q=0}^{+\infty} \varepsilon_q(n)$ et $\nu(n)$ le plus petit entier q tel que $\varepsilon_q(n) = 1$. Montrer que $\nu(n) = 1 + s(n-1) - s(n)$, puis que $\nu(n!) = n - s(n)$.

2. Soit $r \geqslant 1$ et $\mathrm{D} : \mathbb{N}^r \to \mathbb{N}^r$ définie pour $(a_1, \ldots, a_r) \in \mathbb{N}^r$ par

$$\mathrm{D}(a_1, \ldots, a_r) = (|a_1 - a_2|, |a_2 - a_3|, \ldots, |a_{r-1} - a_r|, |a_r - a_1|).$$

Montrer qu'il y a équivalence entre :

(*i*) pour tout $a \in \mathbb{N}^r$, la suite $(\mathrm{D}^n(a))_{n \geqslant 0}$ stationne à 0 ;

(*ii*) r est une puissance de 2.

On pourra utiliser le $\mathbb{Z}/2\mathbb{Z}$-espace vectoriel $(\mathbb{Z}/2\mathbb{Z})^r$.

(École normale supérieure)

Solution.

1. Si $n \in \mathbb{N}$, $s(n)$ désigne le nombre de 1 figurant dans l'écriture de n en base 2. Si $n \neq 0$, $\nu(n)$ désigne la valuation 2-adique de n, *i.e.* le plus grand entier k tel que 2^k divise n, ou encore l'exposant de 2 dans la décomposition de n en produit de facteurs premiers.

• Soit $n \geqslant 1$. Supposons que le chiffre des unités de $n-1$ soit 0. Alors celui de n est 1, les autres chiffres restant inchangés. On a dans ce cas $\nu(n) = 0$, $s(n) = s(n-1) + 1$; la relation $\nu(n) = 1 + s(n-1) - s(n)$ est vérifiée.

Supposons maintenant qu'il y ait r chiffres 1 à droite de l'écriture de $n-1$: $\varepsilon_q(n-1) = 1$ si $0 \leqslant q \leqslant r-1$ et $\varepsilon_r(n-1) = 0$. Alors, par le jeu des retenues, n va s'écrire avec r zéros à droite et un 1 ensuite : $\varepsilon_q(n) = 0$ si $0 \leqslant q \leqslant r-1$ et $\varepsilon_r(n) = 1$, les autres chiffres restant inchangés. Par exemple, en binaire $1011100111 + 1 = 1011101000$. Par conséquent, on a

$$s(n) = s(n-1) - r + 1 \quad \text{et} \quad \nu(n) = r = 1 + s(n-1) - s(n).$$

On conclut que pour tout $n \geqslant 1$, $\boxed{\nu(n) = r = 1 + s(n-1) - s(n)}$.

• Comme, pour $n \geqslant 1$, $\nu(n)$ est l'exposant de 2 dans la décomposition de n en facteurs premiers, on a, si $(a, b) \in \mathbb{N}^{*2}$, $\nu(ab) = \nu(a) + \nu(b)$. On en déduit que

$$\nu(n!) = \sum_{k=1}^{n} \nu(k) = \sum_{k=1}^{n} (1 + s(k-1) - s(k)) = n + s(0) - s(n).$$

$$\boxed{\nu(n!) = n - s(n)}.$$

2. • L'idée est de «plonger» le problème dans $\mathbb{Z}/2\mathbb{Z}$ pour le rendre linéaire. On introduit donc l'application linéaire $\delta : (\mathbb{Z}/2\mathbb{Z})^r \to (\mathbb{Z}/2\mathbb{Z})^r$ qui à $a = (a_1, \ldots, a_r)$ associe $\delta(a) = (a_1 + a_2, a_2 + a_3, \ldots, a_r + a_1)$. Si on note $\overline{a}$ le projeté dans $(\mathbb{Z}/2\mathbb{Z})^r$ d'un r-uplet $a = (a_1, \ldots, a_r) \in \mathbb{N}^r$ on a $\delta(\overline{a}) = \overline{\mathrm{D}(a)}$ puisque, quels que soient les entiers x et y, $|x-y|$ et $x+y$ ont même parité.

Montrons que la propriété (i) équivaut à dire que δ nilpotent.

⋆ Supposons (i). Pour tout $a \in \mathbb{N}^r$, si n est un entier naturel tel que $\mathrm{D}^n(a) = 0$, on obtient $\delta^n(\overline{a}) = \overline{\mathrm{D}^n(a)} = 0$. On applique ce résultat aux vecteurs $e_1, \ldots, e_r$ de la base canonique de $(\mathbb{Z}/2\mathbb{Z})^r$. Pour $1 \leqslant i \leqslant r$, il existe $n_i \in \mathbb{N}$ tel que $\delta^{n_i}(e_i) = 0$. Si p désigne le plus grand des n_i, on obtient $\delta^p = 0$, puisque cet endomorphisme est nul sur une base. L'endomorphisme δ est nilpotent.

⋆ Réciproquement, supposons δ nilpotent, d'indice de nilpotence k. Pour $\mathrm{X} = (a_1, \ldots, a_r) \in \mathbb{N}^r$, on pose $m(\mathrm{X}) = \max_{1 \leqslant i \leqslant r} a_i$. Il est facile de voir que $m(\mathrm{D}(\mathrm{X})) \leqslant m(\mathrm{X})$. Si on fixe X, la suite $(m(\mathrm{D}^n(\mathrm{X})))$ est donc décroissante et le problème est de montrer qu'elle est stationnaire à 0. On a $\overline{\mathrm{D}^k(\mathrm{X})} = \delta^k(\mathrm{X}) = 0$, puisque $\delta^k = 0$. Tous les coefficients de $\mathrm{D}^k(\mathrm{X})$ sont donc pairs et on peut poser $\mathrm{D}^k(\mathrm{X}) = 2\mathrm{X}_1$. On a alors $m(\mathrm{X}_1) \leqslant \frac{1}{2} m(\mathrm{X})$. On reprend le raisonnement avec X_1 : $\mathrm{D}^k(\mathrm{X}_1)$ a tous ses coefficients pairs. On écrit $\mathrm{D}^k(\mathrm{X}_1) = 2\mathrm{X}_2$. Mais comme $\mathrm{D}(2\mathrm{X}) = 2\mathrm{D}(\mathrm{X})$, on a

$$\mathrm{D}^{2k}(\mathrm{X}) = 2\mathrm{D}^k(\mathrm{X}_1) = 4\mathrm{X}_2 \quad \text{et} \quad m(\mathrm{X}_2) \leqslant \frac{1}{2} m(\mathrm{X}_1) \leqslant \frac{1}{4} m(\mathrm{X}).$$

On itère ce raisonnement jusqu'à un indice p tel que

$$m(\mathrm{X}_p) \leqslant \frac{1}{2^p} m(\mathrm{X}) < 1.$$

On a alors $\mathrm{X}_p = 0$ et $\mathrm{D}^{pk}(\mathrm{X}) = 0$ La propriété (i) est vérifiée.

• Il faut maintenant démontrer que la propriété (ii) équivaut à δ nilpotent, c'est-à-dire à $\delta^r = 0$. En effet, l'espace vectoriel $(\mathbb{Z}/2\mathbb{Z})^r$ étant de dimension r, on sait que si δ est nilpotent, alors $\delta^r = 0$ (on se reportera à l'exercice 6.8 pour une démonstration de ce résultat). La matrice de l'application linéaire δ dans la base canonique est

$$\mathrm{A} = \begin{pmatrix} 1 & 1 & 0 & \cdots & 0 \\ 0 & 1 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \\ 0 & & & 1 & 1 \\ 1 & 0 & \cdots & 0 & 1 \end{pmatrix} = \mathrm{I} + \mathrm{P}.$$

où P est la matrice de permutation d'ordre r

$$\mathrm{P} = \begin{pmatrix} 0 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & 0 & 1 & & & \vdots \\ 0 & 0 & 0 & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 & 0 \end{pmatrix}.$$

On a alors

$$\mathrm{P}^2 = \begin{pmatrix} 0 & 0 & 1 & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & & & \ddots & \ddots & 1 \\ 1 & 0 & & & \ddots & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \end{pmatrix}, \ldots, \mathrm{P}^{r-1} = \begin{pmatrix} 0 & 0 & 0 & \cdots & \cdots & 1 \\ 1 & 0 & 0 & & & \vdots \\ 0 & 1 & 0 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & 1 & 0 & 0 \\ 0 & \cdots & \cdots & 0 & 1 & 0 \end{pmatrix}, \mathrm{P}^r = \mathrm{I}.$$

On en déduit que

$$\mathrm{A}^r = (\mathrm{I}+\mathrm{P})^r = \sum_{k=0}^{r} \mathrm{C}_r^k \mathrm{P}^k = 2\mathrm{I} + \sum_{k=1}^{r-1} \mathrm{C}_r^k \mathrm{P}^k = \sum_{k=1}^{r-1} \mathrm{C}_r^k \mathrm{P}^k.$$

L'endomorphisme δ est nilpotent, autrement dit $\mathrm{A}^r = 0$, si et seulement si les C_r^k sont tous pairs pour $1 \leqslant k \leqslant r-1$, car $(\mathrm{I}, \mathrm{P}, \mathrm{P}^2, \ldots, \mathrm{P}^{r-1})$ est une famille libre de matrices. Montrons que c'est le cas si et seulement si r est une puissance de 2.

$\star$ Supposons que les C_r^k soient pairs, pour tout $1 \leqslant k \leqslant r-1$. Pour prouver que r est une puissance de 2, il suffit de vérifier que $s(r) = 1$. On a, d'après la question 1, pour tout $1 \leqslant k \leqslant r-1$,

$$1 \leqslant \nu(\mathrm{C}_r^k) = \nu(r!) - \nu(k!) - \nu((r-k)!) = s(k) + s(r-k) - s(r). \quad (*)$$

En numération binaire r s'écrit $1\alpha_p\alpha_{p-1}\ldots\alpha_0$. Si r n'est pas une puissance de 2, l'un des α_i est non nul. Par conséquent, si on pose $k = \alpha_p\alpha_{p-1}\ldots\alpha_0$, on obtient $1 \leqslant k \leqslant r-1$, $r-k = 10\ldots0$ (avec k zéros) et clairement,

$$s(r) = s(k) + s(r-k),$$

ce qui contredit l'inégalité $(*)$. Donc r est une puissance de 2.

$\star$ Réciproquement, supposons que r soit une puissance de 2. Ceci entraîne $s(r) = 1$ et, pour $1 \leqslant k \leqslant r-1$,

$$\nu(\mathrm{C}_r^k) = s(k) + s(r-k) - s(r) = s(k) + s(r-k) - 1 \geqslant 1,$$

ce qui traduit bien que les C_r^k sont pairs.

Le lecteur pourra se reporter à l'exercice 4.25 sur les congruences de Lucas pour une autre preuve de cette propriété.

Ainsi, nous avons prouvé que δ est nilpotent si et seulement si la propriété (*ii*) est vérifiée. Cela achève la démonstration. ◁

Une fonction arithmétique est une application de $\mathbb{N}^*$ dans $\mathbb{C}$. Le lecteur aura déjà pu rencontrer l'un ou l'autre des exemples importants suivants : φ l'indicatrice d'Euler, σ (qui à n associe la somme des diviseurs de n), μ la fonction de Möbius, τ (qui à n associe le nombre de diviseurs de n)... Elles ont la propriété suivante : si n et n' sont premiers entre eux, l'image de nn' est le produit de l'image de n par l'image de n'. On dit qu'elles sont multiplicatives.

Ces fonctions sont au centre des exercices suivants.

4.28. Expression de $\sum_{k=1}^{n} \tau(k)$

On note $\tau(n)$ le nombre de diviseurs positifs de l'entier n. Montrer que

$$\sum_{k=1}^{n} \tau(k) = \sum_{k=1}^{n} \mathrm{E}\left(\frac{n}{k}\right) = 2\sum_{k=1}^{r} \mathrm{E}\left(\frac{n}{k}\right) - r^2, \quad \text{où } r = \mathrm{E}(\sqrt{n}).$$

(École normale supérieure)

▷ **Solution.**

• La première égalité provient simplement d'une interversion de sommation :

$$\sum_{k=1}^{n} \tau(k) = \sum_{k=1}^{n} \sum_{d|k} 1 = \sum_{d=1}^{n} \sum_{\substack{1 \leqslant k \leqslant n \\ d|k}} 1 = \sum_{d=1}^{n} \mathrm{E}\left(\frac{n}{d}\right),$$

car $\sum_{\substack{1 \leqslant k \leqslant n \\ d|k}} 1$ compte le nombre de multiples de d dans l'intervalle $[\![1, n]\!]$ et il y en a bien $\mathrm{E}\left(\frac{n}{d}\right)$.

• Pour établir la deuxième expression, on écrit

$$\sum_{k=1}^{n} \tau(k) = \sum_{k=1}^{n} \sum_{d|k} 1 = \sum_{dd' \leqslant n} 1,$$

en posant $d' = \dfrac{k}{d}$. L'ensemble des couples (d, d') vérifiant $dd' \leqslant n$ est la réunion des couples (d, d') tels que $d \leqslant \sqrt{n}$ et des couples (d, d') tels

que $d' \leqslant \sqrt{n}$. L'intersection de ces deux parties étant formée des couples (d, d') avec $d \leqslant \sqrt{n}$ et $d' \leqslant \sqrt{n}$, on a

$$\sum_{k=1}^{n} \tau(k) = \sum_{\substack{dd' \leqslant n \\ d \leqslant \sqrt{n}}} 1 + \sum_{\substack{dd' \leqslant n \\ d' \leqslant \sqrt{n}}} 1 - \sum_{\substack{dd' \leqslant n \\ d, d' \leqslant \sqrt{n}}} 1.$$

Par symétrie, les deux premières sommes sont égales et on a donc

$$\sum_{k=1}^{n} \tau(k) = 2 \sum_{\substack{dd' \leqslant n \\ d \leqslant \sqrt{n}}} 1 - r^2 = 2 \sum_{k=1}^{r} \mathrm{E}\left(\frac{n}{k}\right) - r^2. \quad \triangleleft$$

La première égalité de l'énoncé permet de montrer que le nombre moyen de diviseurs d'un entier entre 1 *et* n, *c'est-à-dire* $\dfrac{1}{n} \sum\limits_{k=1}^{n} \tau(k)$, *est équivalent à* $\ln n$ *en l'infini (cf. l'exercice 3.21 du tome 1 d'analyse pour un développement asymptotique plus précis).*

4.29. Une majoration de σ

Pour $n \in \mathbb{N}^*$, on note $\sigma(n)$ la somme des diviseurs de n. Montrer que $\sigma(n) \leqslant n + n \ln n$.

(École normale supérieure)

Solution.

Soit $\mathcal{D}$ l'ensemble des diviseurs de n. Si $d \in \mathcal{D}$, il existe $k \in [\![1, n]\!]$ tel que $d = \dfrac{n}{k}$. On a donc $\mathcal{D} \subset \left\{ \dfrac{n}{k}, k \in [\![1, n]\!] \right\}$, d'où l'on déduit l'inégalité

$$\sigma(n) \leqslant n \sum_{k=1}^{n} \frac{1}{k}.$$

Le résultat demandé se déduit de $\sum\limits_{k=1}^{n} \dfrac{1}{k} \leqslant 1 + \ln n$, ce qui se démontre en remarquant que $\dfrac{1}{k} \leqslant \displaystyle\int_{k-1}^{k} \frac{1}{t} \mathrm{d}t$ pour $k \in [\![2, n]\!]$. $\triangleleft$

4.30. Équation faisant intervenir σ

On note $\sigma(n)$ la somme des diviseurs de n (n entier > 0).

1. Montrer que $\sigma(n)$ est impair si et seulement si n est de la forme $2^\alpha q^2$ avec q impair.

2. Résoudre l'équation $3\sigma(n) = 4n - 17$.

(École normale supérieure)

▷ **Solution.**

1. L'équivalence est triviale pour $n = 1$. Soit donc un entier $n \geqslant 2$. On peut l'écrire sous la forme $n = 2^{\alpha} p_1^{\alpha_1} p_2^{\alpha_2} \dots p_k^{\alpha_k}$, où les p_i sont des nombres premiers impairs deux à deux distincts (les α_i sont dans $\mathbb{N}^*$ mais α peut être nul). La somme des diviseurs de n a la même parité que la somme des diviseurs impairs de n, c'est-à-dire que le nombre de diviseurs impairs de n. Ceux-ci sont les entiers de la forme $p_1^{\beta_1} p_2^{\beta_2} \dots p_k^{\beta_k}$, avec pour tout i, $0 \leqslant \beta_i \leqslant \alpha_i$. Il y en a donc $(\alpha_1+1)(\alpha_2+1)\dots(\alpha_k+1)$. Il en résulte donc que $\sigma(n)$ est impair si et seulement si tous les α_i sont pairs, ce qui conduit directement au résultat demandé.

2. Supposons l'équation $3\sigma(n) = 4n - 17$ possède une solution $n > 0$. D'après la question précédente, on peut écrire n sous la forme $n = 2^{\alpha} q^2$ avec q impair.

• Si $\alpha \geqslant 1$ alors $\dfrac{n}{2}$ est un diviseur de n et on a $\sigma(n) \geqslant n + \dfrac{n}{2} = \dfrac{3n}{2}$. Mais dans ce cas $3\sigma(n) \geqslant \dfrac{9n}{2} > 4n - 17$.

• On a donc $\alpha = 0$ et $n = q^2$. Si on passe alors modulo 3, on obtient $4n \equiv 17 \equiv 2\ [3]$. Or, $4n = (2q)^2$ est congru à 0 ou 1 modulo 3 (2 n'est pas un carré modulo 3).

Conclusion. L'équation proposée n'a pas de solution. ◁

4.31. Sur la fonction σ

On note toujours $\sigma(n)$ la somme des diviseurs de n.

Pour $p \in \mathbb{N}^*$, on pose $f(p) = \sup\{n \in \mathbb{N}^*, \sigma(n) \leqslant p\}$. Montrer que pour tout $k \in \mathbb{N}^*$, l'équation $p - f(p) = k$ a une infinité de solutions.

(École polytechnique)

▷ **Solution.**

• On a $f(1) = 1$. Pour $p \geqslant 2$, on pose $\mathrm{F} = \{n \in \mathbb{N}^*, \sigma(n) \leqslant p\}$. L'ensemble F contient 1 et est clairement majoré par $p - 1$ car on a $\sigma(n) \geqslant n + 1$ pour $n \geqslant 2$. Donc $f(p)$ est bien définie et $f(p) \leqslant p - 1$.

• On remarque que, si $p - 1$ est premier, $p - 1$ appartient à F et donc que $f(p) = p - 1$. Les nombres de la forme $1 + q$, avec q premier sont donc solutions de l'équation pour $k = 1$. Il y en a bien une infinité.

• Prenons $k \geqslant 2$.

⋆ Analyse. Soit $(p_n)_{n \geqslant 0}$ une suite de nombres premiers à choisir, tendant vers l'infini. On a

$$\sigma(p_n) = 1 + p_n \leqslant p_n + k \text{ et donc } f(p_n + k) \geqslant p_n.$$

On obtient l'encadrement $p_n \leqslant f(p_n + k) \leqslant (p_n + k) - 1$. Pour avoir $f(p_n + k) = p_n$, il suffit donc que, pour tout $1 \leqslant m \leqslant k - 1$, $p_n + m$ ne

soit pas premier (du moins pour n est assez grand). En effet, dans ces conditions, $p_n + m$ admet un diviseur strict plus grand que $\sqrt{p_n + m}$ et

$$\sigma(p_n + m) \geqslant p_n + m + \sqrt{p_n + m} + 1 \geqslant p_n + \sqrt{p_n} + 1.$$

Donc à partir d'un certain n_0, si $n \geqslant n_0$, $\sqrt{p_n} \geqslant k$ et la somme des diviseurs de $p_n + m$ est donc strictement plus grande que $p_n + k$. Par conséquent, si $n \geqslant n_0$, $f(p_n + k) = p_n$. Il s'agit donc de trouver une suite $(p_n)_{n \geqslant 0}$ de nombres premiers, tendant vers l'infini, telle que pour tout $n \geqslant 0$ et tout $1 \leqslant m \leqslant k - 1$, $p_n + m$ ne soit pas premier.

$\star$ Synthèse. Pour $n > k$, on considère p_n le plus grand nombre premier inférieur ou égal à $n! - (k+1)$. La suite (p_n) tend vers l'infini. Pour tout $m \in [\![1, k-1]\!]$, on a

$$p_n < p_n + m \leqslant n! - (k+1) + (k-1) = n! - 2.$$

Si $p_n + m \leqslant n! - (k+1)$ alors, par définition de p_n, $p_n + m$ n'est pas premier puisque $p_n + m > p_n$. Si $n! - k \leqslant p_n + m \leqslant n! - 2$, $p_n + m$ s'écrit $n! - i$ avec $2 \leqslant i \leqslant k$. Dans ces conditions, i divise $n! - i$ et i est bien distinct de $n! - i$ car

$$n! - i - i = n! - 2i \geqslant ni - 2i > ki - 2i = (k-2)i \geqslant 0.$$

Donc $p_n + m = n! - i$ ne peut être premier. Par conséquent, la suite $(p_n)_{n>k}$ répond à ce que nous recherchions : d'après l'analyse, $p_n + k$ sera solution de l'équation $p - f(p) = k$ pour n assez grand. La suite $(p_n)_{n>k}$ tendant vers l'infini, il y a bien une infinité de solutions. $\triangleleft$

Il s'agit dans l'exercice suivant de déterminer les fonctions arithmétiques multiplicatives qui sont croissantes.

4.32. Un théorème d'Erdös (1946)

1. Une fonction $f : \mathbb{N}^* \to \mathbb{R}$ est dite complètement multiplicative si pour tout $(a, b) \in (\mathbb{N}^*)^2$, on a $f(ab) = f(a)f(b)$. Déterminer toutes les applications complètement multiplicatives. Quelles sont celles qui sont croissantes ?

2. Une application $f : \mathbb{N}^* \to \mathbb{R}$ est dite multiplicative si on a $f(ab) = f(a)f(b)$ lorsque a et b sont des entiers naturels premiers entre eux. Montrer qu'une application multiplicative croissante est complètement multiplicative.

(École normale supérieure)

▷ **Solution.**

1. Soit f une application complètement multiplicative. On a pour tout $n \geqslant 1$, $f(n) = f(n)f(1)$. Donc soit f est nulle, soit $f(1) = 1$. On ne s'intéresse dans la suite qu'au cas où f est non nulle.

• L'application f est complètement déterminée par la suite des images des nombres premiers. En effet, on a pour tout entier naturel $n \geqslant 1$,

$$f(n) = \prod_{p \in \mathcal{P}} f(p)^{\nu_p(n)}$$

où $\mathcal{P}$ est l'ensemble des nombres premiers et pour tout $p \in \mathcal{P}$, $\nu_p(n)$ est la valuation p-adique de n (le produit est parfaitement défini puisque la famille $(\nu_p(n))_{p\in\mathcal{P}}$ est à support fini).

Réciproquement, si on se donne une suite quelconque $(u_p)_{p\in\mathcal{P}}$ de nombres réels indicée par l'ensemble $\mathcal{P}$, l'application qui, à tout $n \geqslant 1$, associe $\prod_{p\in\mathcal{P}} u_p^{\nu_p(n)}$ est complètement multiplicative.

• En gardant ces notations, on va maintenant chercher à quelles conditions sur la suite (u_p) l'application f obtenue est croissante. Les applications $n \longmapsto n^\alpha$, pour $\alpha \in \mathbb{R}_+$, conviennent. On va voir que ce sont les seules. Pour p premier, on a $1 \leqslant p$ et donc $f(1) = 1 \leqslant u_p$. Posons $\alpha_p = \dfrac{\ln u_p}{\ln p}$. On a alors $f(p) = u_p = p^{\alpha_p}$. On va prouver que tous les α_p sont égaux. Choisissons $p < q$, deux nombres premiers et $m \in \mathbb{N}^*$. Soit n l'unique entier tel que $p^n < q^m < p^{n+1}$. Par croissance de f, il vient $p^{\alpha_p n} \leqslant q^{\alpha_q m} \leqslant p^{\alpha_p(n+1)}$. En passant au logarithme, et en divisant par $n \ln p$, on obtient

$$\alpha_p \leqslant \alpha_q \frac{m}{n} \times \frac{\ln q}{\ln p} \leqslant \alpha_p \frac{n+1}{n}.$$

On fait tendre m vers l'infini. Alors, $n = \mathrm{E}\Big(\dfrac{m \ln q}{\ln p}\Big)$ tend aussi vers l'infini et $n \sim \dfrac{m \ln q}{\ln p}$. Par passage à la limite dans l'inégalité ci-dessus, il vient $\alpha_p \leqslant \alpha_q \leqslant \alpha_p$, ce qui donne le résultat annoncé.

Conclusion. Les seules applications non nulles complètement multiplicatives et croissantes sont les applications $n \longmapsto n^\alpha$ pour $\alpha \in \mathbb{R}_+$.

2. Soit f une application multiplicative non nulle et croissante. En particulier si $n \in \mathbb{N}^*$, $f(n) \geqslant f(1) = 1$. On souhaite prouver que la quantité $\dfrac{\ln f(a)}{\ln a}$ ne dépend pas de $a \geqslant 2$. Une difficulté supplémentaire provient de ce qu'il est dorénavant impossible d'écrire $f(a^k) = f(a)^k$. On va commencer par estimer $f(a)^k$ en utilisant la monotonie de f.

On a $f(a-1) \leqslant f(a) \leqslant f(a+1)$ avec $a \wedge (a-1) = a \wedge (a+1) = 1$. Ainsi, $f(a)^2 \leqslant f(a)f(a+1) = f(a^2+a)$ et $f(a)^2 \geqslant f(a)f(a-1) = f(a^2-a)$.

Essayons maintenant d'encadrer $f(a)^3$. On a $f(a)^3 \leqslant f(a)f(a^2+a)$. Mais ici a et a^2+a ne sont pas premiers entre eux. On majore donc $f(a^2+a)$ par $f(a^2+a+1)$ et $f(a)^3 \leqslant f(a^3+a^2+a)$. De même $f(a)^3 \geqslant f(a^3-a^2-a)$. On montre alors par une récurrence facile que, pour $a \geqslant 2$ et $k \geqslant 1$,

$$f(a^k - a^{k-1} - \cdots - a - 1) \leqslant f(a)^k \leqslant f(a^k + a^{k-1} + \cdots + a + 1).$$

Pour $a = 2$, on a $a^k - a^{k-1} - \cdots - a - 1 = 1$ pour tout k : la minoration obtenue est triviale. On regardera le cas de l'entier 2 à la fin. Prenons deux entiers a et b supérieurs à 3. Soit $n \geqslant 1$. Notons m_n le plus petit entier tel que

$$1 + a + \cdots + a^n \leqslant b^{m_n} - b^{m_n - 1} - \cdots - b - 1$$

et p_n le plus grand entier tel que

$$1 + b + b^2 + \cdots + b^{p_n} \leqslant a^n - a^{n-1} - \cdots - 1.$$

On a

$$f(a)^n \leqslant f(1 + a + \cdots + a^n) \leqslant f(b^{m_n} - b^{m_n - 1} - \cdots - b - 1) \leqslant f(b)^{m_n}$$

et

$$f(b)^{p_n} \leqslant f(1 + b + b^2 + \cdots + b^{p_n}) \leqslant f(a^n - a^{n-1} - \cdots - 1) \leqslant f(a)^n,$$

de sorte qu'en passant au logarithme, on obtient

$$\frac{p_n}{n} \ln f(b) \leqslant \ln f(a) \leqslant \frac{m_n}{n} \ln f(b).$$

Il ne reste plus qu'à étudier les limites de $\frac{p_n}{n}$ et $\frac{m_n}{n}$ lorsque n tend vers l'infini. Les suites (p_n) et (m_n) tendent vers $+\infty$. Par définition de p_n, on a les inégalités

$$1 + b + b^2 + \cdots + b^{p_n} \leqslant a^n - a^{n-1} - \cdots - 1 \leqslant 1 + b + \cdots + b^{p_n} + b^{p_n+1}.$$

On passe au logarithme. Des équivalences

$$\ln(1 + b + b^2 + \cdots + b^{p_n}) \sim p_n \ln b \quad \text{et} \quad \ln(a^n - a^{n-1} - \cdots - 1) \sim n \ln a,$$

on déduit $\lim\limits_{n \to +\infty} \dfrac{p_n}{n} = \dfrac{\ln a}{\ln b}$. On obtient le même résultat pour $\dfrac{m_n}{n}$. On a donc $\ln f(a) = \dfrac{\ln a}{\ln b} \ln f(b)$. En conséquence, il existe $\alpha \in \mathbb{R}^+$ tel que pour tout $n \geqslant 3$, $f(n) = n^\alpha$.

Reste le cas de $n = 2$. De $f(6) = 6^\alpha = f(2.3) = f(2).f(3) = f(2)3^\alpha$, on tire aussi $f(2) = 2^\alpha$.

Conclusion. Toute application non nulle multiplicative et croissante est de la forme $n \longmapsto n^\alpha$ pour $\alpha \in \mathbb{R}_+$ (théorème d'Erdös). ◁

4.33. Probabilité pour que deux entiers soient premiers entre eux

Pour $n \geqslant 1$, on note r_n la probabilité pour que deux entiers choisis aléatoirement dans $[\![1, n]\!]^2$ soient premiers entre eux. D'autre part, on définit la fonction de Möbius $\mu : \mathbb{N}^* \to \mathbb{Z}$ de la manière suivante : $\mu(1) = 1$, $\mu(n) = 0$ si n est divisible par le carré d'un nombre premier et $\mu(p_1 \dots p_r) = (-1)^r$, si les p_i sont des nombres premiers deux à deux distincts.

1. Montrer que $r_n = \dfrac{1}{n^2} \displaystyle\sum_{d=1}^{n} \mu(d)\, \mathrm{E}\left(\frac{n}{d}\right)^2$.
2. Calculer $\displaystyle\sum_{d|n} \mu(d)$.
3. Montrer que $\displaystyle\lim_{n\to+\infty} r_n = \frac{6}{\pi^2}$.

(**École normale supérieure**)

▷ **Solution.**

1. Pour $n \geqslant 1$, notons A_n l'ensemble des couples $(a, b) \in [\![1, n]\!]^2$ tels que $a \wedge b = 1$. On a $r_n = \dfrac{\operatorname{Card} \mathrm{A}_n}{n^2}$. Soit $p_1, \dots, p_k$ les nombres premiers inférieurs à n et U_i l'ensemble des couples (a, b) de $[\![1, n]\!]^2$ tels que $p_i | a$ et $p_i | b$. Il est clair que A_n est le complémentaire de la réunion des U_i. Rappelons la *formule du crible* donnant le cardinal d'une réunion finie d'ensembles finis.

Lemme. *Soit* $\mathrm{U}_1, \dots, \mathrm{U}_k$ k *ensembles finis. Alors, le cardinal de la réunion des* U_i *est donné par*

$$\operatorname{Card}\left(\bigcup_{i=1}^{k} \mathrm{U}_i\right) = \sum_{\emptyset \neq \mathrm{I} \subset [\![1,k]\!]} (-1)^{1+\operatorname{Card} \mathrm{I}} \operatorname{Card}\left(\bigcap_{i\in \mathrm{I}} \mathrm{U}_i\right).$$

Démonstration. Ce résultat se prouve par récurrence sur k ou à l'aide des fonctions caractéristiques. ◇

Si $\mathrm{I} \subset [\![1, k]\!]$ est non vide, le cardinal de l'intersection $\bigcap_{i\in \mathrm{I}} \mathrm{U}_i$ est égal au nombre de couples de multiples strictement positifs de $\displaystyle\prod_{i\in \mathrm{I}} p_i$ inférieurs ou égaux à n ; il vaut donc $\mathrm{E}\left(\dfrac{n}{\displaystyle\prod_{i\in \mathrm{I}} p_i}\right)^2$. La formule du crible donne alors

$$\begin{aligned}\operatorname{Card} \mathrm{A}_n &= n^2 - \operatorname{Card}\left(\bigcup_{i=1}^{k} \mathrm{U}_k\right) \\ &= n^2 - \sum_{\emptyset \neq \mathrm{I} \subset [\![1,k]\!]} (-1)^{\operatorname{Card} \mathrm{I}+1} \operatorname{Card}\left(\bigcap_{i\in \mathrm{I}} \mathrm{U}_i\right)\end{aligned}$$

$$= n^2 - \sum_{\emptyset \neq I \subset [\![1,k]\!]} (-1)^{\operatorname{Card} I+1} \mathrm{E}\left(\frac{n}{\prod_{i\in I} p_i}\right)^2 = \sum_{d=1}^{n} \mu(d) \mathrm{E}\left(\frac{n}{d}\right)^2.$$

On en déduit la probabilité r_n :

$$\boxed{r_n = \frac{1}{n^2} \sum_{d=1}^{n} \mu(d) \mathrm{E}\left(\frac{n}{d}\right)^2}.$$

2. Si nous notons $\mathrm{S}(n) = \sum_{d|n} \mu(d)$, nous avons $\mathrm{S}(1) = 1$ et nous allons prouver que $\mathrm{S}(n) = 0$, si $n \geqslant 2$. Pour cela, considérons la décomposition de n en en facteurs premiers : $n = \prod_{i=1}^{k} p_i^{\alpha_i}$, où $p_1, \ldots, p_k$ sont des nombres premiers distincts et $\alpha_1, \ldots, \alpha_k$ des entiers strictement positifs. Les seuls diviseurs d de n pour lesquels $\mu(d)$ est non nul sont les produits de nombres premiers distincts pris parmi $p_1, \ldots, p_k$. Pour un tel diviseur, on a $\mu(d) = (-1)^i$, où i est le nombre de diviseurs premiers de d. Or, n possède C_k^i diviseurs d correspondant à un i fixé. On en déduit que

$$\sum_{d|n} \mu(d) = \sum_{i=0}^{k} \mathrm{C}_k^i (-1)^i = (1-1)^k = 0.$$

On conclut que $\boxed{\sum_{d|n} \mu(d) = \begin{cases} 1 \text{ si } n = 1 \\ 0 \text{ si } n \geqslant 2 \end{cases}}$.

3. Pour l'étude asymptotique de r_n, il paraît naturel de remplacer le terme $\frac{1}{n^2}\mathrm{E}\left(\frac{n}{d}\right)^2$ par son équivalent $\frac{1}{d^2}$. La différence entre les deux sommes s'écrit

$$\left| r_n - \sum_{d=1}^{n} \frac{\mu(d)}{d^2} \right| = \left| \sum_{d=1}^{n} \mu(d) \left(\frac{1}{n^2} \mathrm{E}\left(\frac{n}{d}\right)^2 - \frac{1}{d^2} \right) \right|.$$

Comme $\mathrm{E}\left(\frac{n}{d}\right) > \frac{n}{d} - 1$, on a $\frac{1}{n^2} - \frac{2}{dn} < \frac{1}{n^2}\mathrm{E}\left(\frac{n}{d}\right)^2 - \frac{1}{d^2} \leqslant 0$, ce qui donne la majoration

$$\left| r_n - \sum_{d=1}^{n} \frac{\mu(d)}{d^2} \right| \leqslant \sum_{d=1}^{n} \left(\frac{2}{dn} + \frac{1}{n^2} \right) \leqslant \frac{2}{n} \sum_{d=1}^{n} \frac{1}{d} + \frac{1}{n} = \mathrm{O}\left(\frac{\ln n}{n}\right),$$

parce que la somme partielle de la série harmonique est équivalente à $\ln n$. Il en résulte donc que $\lim_{n\to+\infty} r_n = \sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}$ (la série est absolument convergente).

• La quantité $\frac{6}{\pi^2}$ est l'inverse de $\zeta(2) = \sum_{n=1}^{+\infty} \frac{1}{n^2}$. Cela nous invite à calculer le produit des sommes $\sum_{n=1}^{+\infty} \frac{1}{n^2}$ et $\sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}$. Les familles $\left(\frac{\mu(d)}{d^2}\right)_{d\geqslant 1}$ et $\left(\frac{1}{n^2}\right)_{n\geqslant 1}$ sont sommables, donc la suite double $\left(\frac{\mu(d)}{d^2 n^2}\right)_{n,d\geqslant 1}$ l'est aussi et est elle est justifiable du théorème d'associativité :

$$\left(\sum_{d=1}^{+\infty} \frac{\mu(d)}{d^2}\right)\left(\sum_{n=1}^{+\infty} \frac{1}{n^2}\right) = \sum_{d,n\geqslant 1} \frac{\mu(d)}{(dn)^2} = \sum_{\substack{d\geqslant 1\\ d|p}} \frac{\mu(d)}{p^2}$$

$$= \sum_{p\geqslant 1} \sum_{d|p} \frac{\mu(d)}{p^2} = \sum_{p\geqslant 1} \frac{1}{p^2}\left(\sum_{d|p} \mu(d)\right) = 1,$$

d'après le calcul de la question précédente.

Conclusion. $\boxed{\lim_{n\to+\infty} r_n = \frac{6}{\pi^2}}$. ◁

L'importance de la fonction de Möbius provient de la «formule d'inversion» suivante. Si f est une fonction de $\mathbb{N}^*$ dans $\mathbb{C}$ et qu'on définit la fonction $g : \mathbb{N}^* \to \mathbb{C}$ par $g(n) = \sum_{d|n} f(d)$, on peut retrouver f à partir de g puisque, pour tout $n \in \mathbb{N}^*$, on a $f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) g(d)$. Par exemple, pour $n \geqslant 1$, on a $n = \sum_{d|n} \varphi(n)$ (voir 4.17). On en déduit que $\varphi(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) d$.

Les deux exercices suivants proposent d'établir, avec des arguments complètement différents, le fait que tout nombre premier p congru à 1 modulo 4 est somme de deux carrés. C'est le premier pas vers la caractérisation des entiers naturels n s'écrivant comme somme de deux carrés : n est somme de deux carrés si, et seulement si, pour tout nombre premier p congru à 3 modulo 4, l'exposant de p dans la décomposition en produits de facteurs premiers de n est pair.

4.34. Écriture d'un nombre premier comme somme de deux carrés

Soit p un nombre premier impair.

1. Montrer que si p est une somme de deux carrés d'entiers, on a alors $p \equiv 1 \pmod 4$.

On suppose p congru à 1 modulo 4.

2. Dénombrer les carrés dans $(\mathbb{Z}/p\mathbb{Z})^*$.

3. En déduire qu'il existe $n \in \mathbb{Z}$ tel que $n^2 \equiv -1 \pmod{p}$.

4. Démontrer qu'il existe $(a, b) \in \mathbb{Z}^2$ tels que :

$$0 < b < \sqrt{p} \quad \text{et} \quad \left| b\frac{n}{p} - a \right| \leqslant \frac{1}{\sqrt{p}}.$$

5. Montrer que $(bn - ap)^2 + b^2 = p$.

(**École polytechnique**)

Solution.

1. Modulo 4, un carré est congru à 0 ou à 1. Si un entier est somme de deux carrés, il sera donc congru modulo 4 à 0, 1 ou 2.

Comme p est un entier premier impair, il ne peut être congru ni 0, ni à 2 (car il serait alors divisible par 2). Il s'ensuit qu'un entier premier impair, somme de deux carrés d'entiers est congru à 1 modulo 4.

2. L'entier p étant premier, $\mathrm{K} = \mathbb{Z}/p\mathbb{Z}$ est un corps. On a donc, pour $(x, y) \in \mathrm{K}^{*2}$,

$$x^2 = y^2 \Longleftrightarrow (x - y)(x + y) = 0 \Longleftrightarrow x = y \text{ ou } x = -y.$$

Par conséquent, à tout carré de K^*, correspondent exactement deux antécédents dans K^* par l'application $x \longmapsto x^2$ (on a bien $x \neq -x$ pour $x \in \mathrm{K}^*$, puisque la caractéristique de K est $p > 2$). Il y a donc exactement $\dfrac{\mathrm{Card}\,\mathrm{K}^*}{2} = \dfrac{p-1}{2}$ carrés dans K^*.

3. Il suffit de prouver que -1 est un carré dans K. Si $x \in \mathrm{K}^*$ est un carré, on peut écrire $x = y^2$ avec $y \in \mathrm{K}^*$ et d'après le petit théorème de Fermat,

$$x^{\frac{p-1}{2}} = y^{2 \times \frac{p-1}{2}} = y^{p-1} = 1.$$

Donc x est racine du polynôme $\mathrm{P} = \mathrm{X}^{\frac{p-1}{2}} - 1$. Les $\dfrac{p-1}{2}$ carrés non nuls de K sont donc racines de P. Or, P a au plus $\dfrac{p-1}{2}$ racines distinctes ou confondues dans le corps K. Nécessairement, P est scindé et ses racines sont exactement les carrés de K^*.

Comme $p \equiv 1 \pmod{4}$, $\dfrac{p-1}{2}$ est pair, $(-1)^{\frac{p-1}{2}} = 1$ et -1 est un carré dans K.

4. On pose $\mathrm{N} = \mathrm{E}(\sqrt{p}) + 1$ et $\xi = \dfrac{n}{p}$ et on considère les N réels $x_k = k\xi - \mathrm{E}(k\xi)$ de l'intervalle $[0, 1[$, pour $0 \leqslant k \leqslant \mathrm{N} - 1$, et les N intervalles $\left[0, \dfrac{1}{\mathrm{N}}\right[, \left[\dfrac{1}{\mathrm{N}}, \dfrac{2}{\mathrm{N}}\right[, \ldots, \left[\dfrac{\mathrm{N}-1}{\mathrm{N}}, 1\right[$.

• Supposons d'abord que l'un des x_k soit dans l'intervalle $\left[\dfrac{N-1}{N}, 1\right[$. On a $k > 0$ car $x_0 = 0$. En posant $b = k$ et $a = E(k\xi) + 1$, on obtient

$$0 < b \leqslant N - 1 < \sqrt{p} \quad \text{et} \quad \left|b\frac{n}{p} - a\right| = |x_k - 1| \leqslant \frac{1}{N} \leqslant \frac{1}{\sqrt{p}},$$

l'inégalité $N - 1 < \sqrt{p}$ étant stricte car $\sqrt{p} \notin \mathbb{N}$.

• Dans le cas contraire, les N réels x_k sont dans les $N - 1$ intervalles $\left[\dfrac{k}{N}, \dfrac{k+1}{N}\right[$, avec $0 \leqslant k \leqslant N - 2$. D'après le principe des tiroirs de Dirichlet, il existe k et ℓ distincts tels que x_k et x_ℓ soient dans le même intervalle. Supposons par exemple $k < \ell$. Notons alors $b = \ell - k$ et $a = E(\ell\xi) - E(k\xi)$. On a de nouveau $0 < b \leqslant N - 1 < \sqrt{p}$ et

$$\left|b\frac{n}{p} - a\right| = |(\ell - k)\xi - (E(\ell\xi) - E(k\xi))| = |x_\ell - x_k| \leqslant \frac{1}{N} \leqslant \frac{1}{\sqrt{p}}.$$

Dans tous les cas, nous avons montré l'existence de $(a, b) \in \mathbb{Z}^2$ tel que

$$0 < b < \sqrt{p} \quad \text{et} \quad \left|b\frac{n}{p} - a\right| \leqslant \frac{1}{\sqrt{p}}.$$

5. Les inégalités obtenues dans la question 4 impliquent

$$0 < b^2 < p \quad \text{et} \quad (bn - ap)^2 \leqslant p \text{ et donc } 0 < (bn - ap)^2 + b^2 < 2p$$

D'autre part, on a $n^2 + 1 \equiv 0 \pmod{p}$ et donc

$$(bn - ap)^2 + b^2 = b^2(n^2 + 1) - 2abnp + a^2p^2 \equiv 0 \pmod{p}.$$

Comme cet entier appartient à l'intervalle $]0, 2p[$, cela implique l'égalité

$$(bn - ap)^2 + b^2 = p.$$

Conclusion. Tout nombre premier congru à 1 modulo 4 s'écrit comme somme de deux carrés d'entiers. ◁

4.35. Théorème des deux carrés, preuve combinatoire

Soit p premier congru à 1 modulo 4, $p = 4k + 1$. On considère l'ensemble

$$S = \{(a, b, c) \in \mathbb{N}^2 \times \mathbb{Z},\ 4ab + c^2 = p\}.$$

1. Montrer que S est non vide, fini et inclus dans $(\mathbb{N}^*)^2 \times \mathbb{Z}^*$. Prouver que $S_1 = \{(a, b, c) \in S,\ a > b + c\}$ et $S_2 = \{(a, b, c) \in S,\ a < b + c\}$ forment une partition de S. Exhiber une bijection de S_1 sur S_2.

2. Montrer que $f : (a,b,c) \longmapsto (a-b-c, b, -2b-c)$ est une involution de S_1. Chercher ses points fixes et en déduire que le cardinal de S est congru à 2 modulo 4.

3. Montrer que $S_3 = \{(a,b,c) \in S,\ a \neq b\}$ a un cardinal divisible par 4. En déduire que p est somme de deux carrés.

(École normale supérieure)

Solution.

1. Comme $(k,1,1) \in S$, S est non vide. Si $(a,b,c) \in S$ les entiers a, b et $|c|$ sont majorés par p. Donc S est fini. De plus, si $a = 0$ ou $b = 0$, on a $p = c^2$, ce qui est impossible car p est premier et si $c = 0$, $p = 4ab$ est pair ce qui est tout aussi absurde. Donc S est inclus dans $(\mathbb{N}^*)^2 \times \mathbb{Z}^*$.

On a $(k,1,1) \in S_1$ et $(1,k,1) \in S_2$, de sorte que S_1 et S_2 sont non vides. Comme S_1 et S_2 sont visiblement disjoints, il suffit de prouver qu'il n'y a pas de triplet $(a,b,c) \in S$ tel que $a = b+c$. Or, on aurait dans ce cas

$$p = 4b(b+c) + c^2 = (c+2b)^2,$$

ce qui est impossible. On a bien $S_1 \cup S_2 = S$.

D'autre part, on peut vérifier que l'application qui à (a,b,c) associe $(b,a,-c)$ est une bijection de S_1 sur S_2. Les ensembles S_1 et S_2 ont donc même cardinal et $|S| = 2|S_1|$.

2. Montrons pour commencer que si le triplet (a,b,c) est dans S_1, alors $(a-b-c, b, -2b-c)$ est aussi dans S_1. En effet, $a-b-c$ et b sont positifs, $4(a-b-c)b + (-2b-c)^2 = 4ab + c^2 = p$ et $a-b-c > -b-c$ car $a > 0$. Donc f est une application de S_1 dans S_1. Enfin $f \circ f = \mathrm{Id}_{S_1}$ de sorte qu'on a bien une involution.

Le triplet (a,b,c) est fixe par f si $a = a-b-c$ et $c = -2b-c$, c'est-à-dire si $c = -b$. Il vient alors $p = 4ab + b^2 = b(4a+b)$. Comme p est premier, cela impose $b = 1$ et $a = k$. Il y a donc un unique point fixe. Par ailleurs, la décomposition en cycles à supports disjoints de la permutation f est composée uniquement de transpositions (car f est d'ordre 2). La réunion de leurs supports (S_1 privé du point fixe) est donc de cardinal pair. Il en résulte que le cardinal de S_1 est impair et donc que $|S| \equiv 2 \pmod 4$ (puisque $|S| = 2|S_1|$).

3. À tout $(a,b,c) \in S$ avec $a \neq b$, on peut associer les 4 triplets *distincts* (a,b,c), $(a,b,-c)$, (b,a,c) et $(b,a,-c)$ tous dans S. Le cardinal de S_3 est donc divisible par 4. Il résulte de cela et du résultat de la question précédente que $|S_3|$ ne peut être égal à $|S|$ et que S_3 est donc strictement inclus dans S. On peut donc trouver un triplet de la forme (a,a,c) dans S. On a alors $p = 4a^2 + c^2 = (2a)^2 + c^2$ qui est somme de deux carrés. ◁

Cette preuve combinatoire du théorème des deux carrés est très récente. Elle est due à Don Zagier et date de 1990.

On s'intéresse dans ce qui suit aux sommes de quatre carrés.

4.36. Théorème des quatre carrés de Lagrange (1770)

1. Établir que pour tout $(a,b,c,d,x,y,z,t) \in \mathbb{Z}^8$,

$$(a^2+b^2+c^2+d^2)(x^2+y^2+z^2+t^2) = (ax+by+cz+dt)^2 + (ay-bx-ct+dz)^2 + (az+bt-cx-dy)^2 + (at-bz+cy-dx)^2.$$

2. Soit p un nombre premier impair. Montrer qu'il existe (a,b) dans $\mathbb{Z}^2$ tel que p divise $1+a^2+b^2$.

3. Soit E l'ensemble des entiers naturels $n \geqslant 1$ tel que np soit somme de quatre carrés d'entiers. Montrer que E n'est pas vide. On note m son plus petit élément. Montrer que $m < p$, puis que m est impair.

4. Supposons que $1 < m$ et que $mp = a^2+b^2+c^2+d^2$. En considérant les résidus de a,b,c,d modulo m de valeur absolue minimale, construire $m' < m$ appartenant à E. Conclure.

5. En déduire que tout entier naturel peut s'écrire comme somme de quatre carrés.

(École normale supérieure)

▷ **Solution.**

1. Prendre un logiciel de calcul formel ou développer de tête le membre de droite et se convaincre que les doubles produits s'éliminent.

2. Notons que p divise $1+a^2+b^2$ si et seulement si $-\overline{1} = \overline{a}^2 + \overline{b}^2$ dans $\mathbb{Z}/p\mathbb{Z}$. Il nous suffit de montrer que $-\overline{1}$ est somme de deux carrés dans $\mathbb{Z}/p\mathbb{Z}$.

On sait qu'il y a exactement $\dfrac{p+1}{2}$ carrés dans $\mathbb{Z}/p\mathbb{Z}$. Rappelons-en rapidement la preuve. L'application $x \longmapsto x^2$ est un morphisme de groupes de $(\mathbb{Z}/p\mathbb{Z})^*$ dans lui-même, de noyau $\{\pm\overline{1}\}$ et d'image le groupe C des carrés non nuls. Le premier théorème d'isomorphisme permet d'affirmer que C est isomorphe à $(\mathbb{Z}/p\mathbb{Z})^*/\{\pm\overline{1}\}$. Comme $-\overline{1} \neq \overline{1}$ (car p est impair), il y a exactement $\dfrac{p-1}{2}$ carrés non nuls. En rajoutant 0 le compte est bon.

Les ensembles $\{x^2,\ x \in (\mathbb{Z}/p\mathbb{Z})\}$ et $\{-1-y^2,\ y \in (\mathbb{Z}/p\mathbb{Z})\}$ sont donc tous les deux de cardinal $\dfrac{p+1}{2}$, donc ils se coupent. Notons que cela prouve plus généralement que tout élément de $\mathbb{Z}/p\mathbb{Z}$ est somme de deux carrés.

3. La question 2 montre l'existence de $n \in \mathbb{N}^*$ tel que

$$np = 1 + a^2 + b^2 = 0^2 + 1^2 + a^2 + b^2.$$

Donc E n'est pas vide. En fait, on peut choisir a et b avec une valeur absolue minimale, c'est-à-dire vérifiant $|a| < \frac{p}{2}$ et $|b| < \frac{p}{2}$. En effet, si ce n'est pas le cas, on considère a' et b', congrus respectivement à a et b modulo p, tels que $|a'| < \frac{p}{2}$ et $|b'| < \frac{p}{2}$ (les inégalités sont strictes car p est impair). On a alors

$$1 + a'^2 + b'^2 \equiv 1 + a^2 + b^2 \equiv 0 \pmod{p}$$

et il existe $n' \in \mathbb{N}^*$ tel que $1 + a'^2 + b'^2 = n'p$. Si a et b vérifient cette condition supplémentaire, on a $np = 1 + a^2 + b^2 < 1 + \frac{p^2}{2}$. On en déduit que $n < p$ et donc $m < p$, par définition de m.

Supposons par l'absurde que m soit pair. On sait qu'on peut écrire $mp = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$. Les entiers $\alpha, \beta, \gamma, \delta$ sont soit tous pairs, soit tous impairs, soit deux pairs et deux impairs. Quitte à changer l'ordre des termes, on suppose dans le troisième cas que α et β sont pairs et γ et δ sont impairs. Mais alors, dans les trois cas on peut écrire

$$\frac{m}{2}p = \left(\frac{\alpha+\beta}{2}\right)^2 + \left(\frac{\alpha-\beta}{2}\right)^2 + \left(\frac{\gamma+\delta}{2}\right)^2 + \left(\frac{\gamma-\delta}{2}\right)^2,$$

ce qui contredit la minimalité de m. Donc m est impair.

4. Notons $\alpha, \beta, \gamma, \delta$ les résidus respectifs de a, b, c, d introduits par l'énoncé. On a $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{m}$. Soit $m' \in \mathbb{N}$ tel que $mm' = \alpha^2 + \beta^2 + \gamma^2 + \delta^2$. L'entier m' n'est pas nul car sinon m diviserait a, b, c et d, m^2 diviserait mp et m diviserait p, ce qui est impossible, car $1 < m < p$.

Par ailleurs, comme $\alpha^2 + \beta^2 + \gamma^2 + \delta^2 < 4\left(\frac{m}{2}\right)^2 = m^2$ (l'inégalité est stricte car m est impair), on a $0 < m' < m$. D'après la première question, on peut écrire

$$m^2m'p = (\alpha^2 + \beta^2 + \gamma^2 + \delta^2)(a^2 + b^2 + c^2 + d^2) = \mathrm{A}^2 + \mathrm{B}^2 + \mathrm{C}^2 + \mathrm{D}^2,$$

avec

$$\mathrm{A} = a\alpha + b\beta + c\gamma + d\delta \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{m},$$
$$\mathrm{B} = a\beta - b\alpha - c\delta + d\gamma \equiv ab - ab - cd + cd \equiv 0 \pmod{m}$$

et de même, C et D divisibles par m. On a donc

$$m'p = \left(\frac{\mathrm{A}}{m}\right)^2 + \left(\frac{\mathrm{B}}{m}\right)^2 + \left(\frac{\mathrm{C}}{m}\right)^2 + \left(\frac{\mathrm{D}}{m}\right)^2$$

et m' appartient à E. Cela contredit la définition de m. On conclut que $m = 1$.

5. Il est clair que 0, 1 et 2 sont sommes de quatre carrés. D'après la question 4, tout nombre premier impair également. Le théorème de décomposition en facteurs premiers et la question 1 permettent de conclure. ◁

Voici un exercice sur les sommes de trois carrés.

4.37. Lemme de Davenport-Cassels

Soit $n \in \mathbb{N}$. On suppose que n est la somme des carrés de trois rationnels. Montrer que n est la somme des carrés de trois entiers.
(École normale supérieure)

▷ **Solution.**

La propriété est évidente pour $n = 0$. Nous supposerons donc $n > 0$. Nous donnons de ce problème une formulation géométrique : si la sphère S de $\mathbb{R}^3$ d'équation $x^2 + y^2 + z^2 = n$ passe par un point rationnel, alors elle passe aussi par un point entier. Nous raisonnerons par l'absurde et supposerons que S contient un point à coordonnées rationnelles, mais pas de point à coordonnées entières.

Soit $a \in \mathbb{Q}^3 \cap \mathrm{S}$. Il existe $u \in \mathbb{Z}^3$ et $d \geqslant 2$ tel que $a = \dfrac{1}{d}u$. Nous supposerons a et u choisis de telle façon que d soit minimal.

Montrons qu'il existe $a' \in \mathbb{Z}^3$ tel que $\|a - a'\| < 1$ (il s'agit de norme euclidienne). Si $a = (x, y, z)$, on considère les entiers x', y', z' les plus proches de x, y, z respectivement. On obtient $|x - x'| \leqslant \dfrac{1}{2}$, $|y - y'| \leqslant \dfrac{1}{2}$ et $|z - z'| \leqslant \dfrac{1}{2}$. On en déduit que $\|a - a'\| \leqslant \sqrt{\dfrac{1}{4} + \dfrac{1}{4} + \dfrac{1}{4}} = \dfrac{\sqrt{3}}{2} < 1$.

Puisque a n'appartient pas à $\mathbb{Z}^3$, a' est distinct de a. La droite qui joint a à a' coupe la sphère S en a. Elle la recoupe en point a_1 dont nous allons calculer les coordonnées.

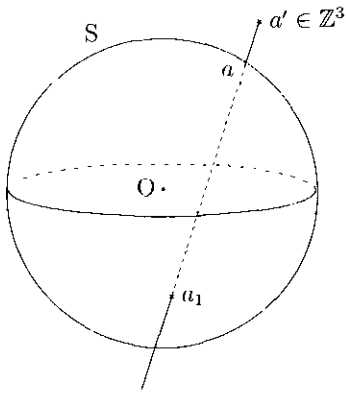
Il existe $\lambda \in \mathbb{R}$ tel que $a_1 = a' + \lambda(a - a')$. On écrit que a_1 appartient à S :

$$n = \|a_1\|^2 = \|a'\|^2 + 2\lambda\langle a', a - a'\rangle + \lambda^2\|a - a'\|^2.$$

Une solution de cette équation du second degré est 1, puisque $a \in \mathrm{S}$. L'autre est donc $\lambda = \dfrac{\|a'\|^2 - n}{\|a - a'\|^2}$. Examinons $\|a - a'\|^2$; on obtient

$$\|a - a'\|^2 = \|a'\|^2 + \|a\|^2 - 2\langle a', a\rangle = \|a'\|^2 + n - \frac{2}{d}\langle a', u\rangle = \frac{d_1}{d},$$

avec $d_1 \in \mathbb{N}^*$, puisque a' et u sont dans $\mathbb{Z}^3$ et $0 < d_1 < d$, puisque $0 \quad \|a - a'\| < 1$.



On en déduit $\lambda = \dfrac{\|a'\|^2 - n}{\|a - a'\|^2} = \dfrac{d(\|a'\|^2 - n)}{d_1}$ et

$$\begin{aligned} a_1 &= a' + \lambda(a - a') = a' + \frac{d(\|a'\|^2 - n)}{d_1} \frac{1}{d}(u - da') \\ &= a' + \frac{\|a'\|^2 - n}{d_1}(u - da'). \end{aligned}$$

Il existe donc $v \in \mathbb{Z}^3$ tel que $a_1 = \dfrac{1}{d_1} v$. Par ailleurs, a_1 appartient à S. Mais on a $0 < d_1 < d$, ce qui contredit la minimalité de d. Nous obtenons la contradiction souhaitée. $\lhd$

4.38. Une équation diophantienne

Soit $n \in \mathbb{N}$, $\alpha \in \mathbb{N}$, avec $\alpha > n \geqslant 2$. Montrer que l'équation

$$x_1^2 + \cdots + x_n^2 = \alpha x_1 \ldots x_n$$

n'a pas de solution entière autre que $(0, \ldots, 0)$.

(École normale supérieure)

$\rhd$ **Solution.**

Nous allons raisonner par l'absurde, en utilisant la méthode de la descente infinie : supposant qu'il existe une solution $(x_1, \ldots, x_n)$, différente de $(0, \ldots, 0)$, nous démontrerons qu'on peut trouver une autre solution $(y_1, \ldots, y_n)$, différente de $(0, \ldots, 0)$, telle que $y_1 + \cdots y_n < x_1 + \cdots + x_n$. Le procédé peut être réitéré. La somme $x_1 + \cdots + x_n$, qui appartient

à $\mathbb{N}$, ne pouvant décroître indéfiniment, cela donnera la contradiction souhaitée.

Si $(x_1, \dots, x_n)$ est une solution différente de $(0, \dots, 0)$, on a, pour tout $i \in [\![1, n]\!]$, $x_i \neq 0$. On peut sans perte de généralité supposer que $x_1 \leqslant \cdots \leqslant x_n$. On va chercher une autre solution de l'équation sous la forme $(x_1, \dots, x_{n-1}, y)$ et vérifier que $y < x_n$. Pour que $(x_1, \dots, x_{n-1}, y)$ soit solution, il faut que y soit racine du polynôme

$$\mathrm{P} = \mathrm{X}^2 - \alpha x_1 \dots x_{n-1} \mathrm{X} + \sum_{i=1}^{n-1} x_i^2.$$

On sait déjà que l'entier x_n est une racine de P. La somme des racines de ce polynôme étant $\alpha x_1 \dots x_{n-1}$, l'autre racine y est entière et strictement positive (car le produit des racines est $\sum_{i=1}^{n-1} x_i^2$). Comparons-la à x_{n-1}. Le calcul de

$$\begin{aligned} \mathrm{P}(x_{n-1}) &= x_{n-1}^2 - \alpha x_1 \dots x_{n-2} x_{n-1}^2 + \sum_{i=1}^{n-1} x_i^2 \\ &= (n - \alpha x_1 \dots x_{n-2}) x_{n-1}^2 + \sum_{i=1}^{n-2} x_i^2 - (n-2) x_{n-1}^2 \end{aligned}$$

montre que $\mathrm{P}(x_{n-1})$ est strictement négatif, car $\sum_{i=1}^{n-2} x_i^2 \leqslant (n-2) x_{n-1}^2$, et $\alpha x_1 \dots x_{n-2} \geqslant \alpha > n$. On en déduit que x_{n-1} est entre x_n et y. On a donc $y < x_{n-1} < x_n$.

On a obtenu une autre solution $(x_1, \dots, x_{n-1}, y)$ de notre équation, différente de $(0, \dots, 0)$ vérifiant $x_1 + \cdots + x_{n-1} + y < x_1 + \cdots + x_{n-1} + x_n$. ◁

Ce chapitre d'arithmétique ne pouvait s'achever sans citer le grand théorème de Fermat : pour $n \geqslant 3$, l'équation $x^n + y^n = z^n$ n'a pas de solution non triviale dans $\mathbb{Z}$. On trouve mention de ce résultat dans une annotation marginale, écrite dans l'ouvrage de Diophante par Fermat, qui déclarait ne pas pouvoir donner sa preuve par manque de place. On est certain aujourd'hui que Fermat ne pouvait pas en avoir une démonstration complète. Mais le mythe était créé et des générations de mathématiciens pendant trois siècles allaient s'évertuer à le démontrer. Le Français Lamé crut son heure de gloire arrivée lorsqu'au siècle dernier, il en présenta une preuve. Malheureusement, il utilisait implicitement la factorialité de certains sous-anneaux de $\mathbb{C}$, qui n'est pas toujours vérifiée. De cette erreur furent tirés des prolongements qui ont contribué à mettre en place la théorie moderne des anneaux et des idéaux.

C'est en 1994 que le mathématicien anglais Wiles achevait ce long chemin dont le bout était éclairé depuis déjà une vingtaine d'années par les avancées de Weil.

L'exercice suivant résout ce qu'on appelle le premier cas du théorème de Fermat pour certains nombres premiers.

4.39. Théorème de Sophie Germain (1823)

Soit p un nombre premier de Sophie Germain, c'est-à-dire un nombre premier impair tel que $q = 2p+1$ soit premier. On se propose de prouver le théorème de Sophie Germain : *il n'existe pas de triplet* $(x, y, z) \in \mathbb{Z}^3$ *tel que* $xyz \not\equiv 0\ [p]$ *et* $x^p + y^p + z^p = 0$.

1. On raisonne par l'absurde. On suppose donné dans la suite un triplet $(x, y, z) \in \mathbb{Z}^3$ tel que $x^p + y^p + z^p = 0$ et $xyz \not\equiv 0\ [p]$. Montrer qu'on peut supposer $\operatorname{pgcd}(x, y, z) = 1$ et qu'alors, x, y, z sont premiers entre eux deux à deux.

2. Montrer qu'il existe deux entiers a et α tels que $y + z = a^p$ et $\sum_{k=0}^{p-1} (-z)^{p-1-k} y^k = \alpha^p$. Établir l'existence de deux entiers b, c tels que $x + y = c^p$ et $x + z = b^p$.

3. Si $m \in \mathbb{Z}$ n'est pas divisible par q, montrer que $m^p \equiv \pm 1\ \ [q]$. En déduire qu'un et un seul des trois entiers x, y, z est divisible par q. On supposera que c'est x.

4. Établir successivement les congruences suivantes, toutes modulo q

$$b^p + c^p - a^p \equiv 0\ \ ;\ \ a \equiv 0\ \ ;\ \ y \equiv c^p \quad \text{et} \quad \alpha^p \equiv p y^{p-1}.$$

Obtenir une contradiction et conclure.

(École normale supérieure)

Solution.

1. Soit $d = \operatorname{pgcd}(x, y, z)$, $x' = \dfrac{x}{d}$, $y' = \dfrac{y}{d}$ et $z' = \dfrac{z}{d}$. On a alors $x'^p + y'^p + z'^p = 0$, $\operatorname{pgcd}(x', y', z') = 1$ et bien entendu $x'y'z' \not\equiv 0\ [p]$. On peut donc supposer $\operatorname{pgcd}(x, y, z) = 1$. Supposons alors que $\operatorname{pgcd}(x, y) > 1$ et notons p_0 un diviseur premier de $\operatorname{pgcd}(x, y)$. Comme p_0 divise $x^p + y^p$, il divise z^p et donc z d'après le lemme d'Euclide, ce qui est contradictoire avec $\operatorname{pgcd}(x, y, z) = 1$. Donc x et y sont premiers entre eux. Le même raisonnement s'applique aux autres couples.

2. Montrons en raisonnant par l'absurde que $y+z$ et $\sum_{k=0}^{p-1} (-z)^{p-1-k} y^k$ sont premiers entre eux. Supposons donc qu'ils ont un codiviseur premier p'. On a classiquement

$$(1) \qquad (y+z)\left(\sum_{k=0}^{p-1}(-z)^{p-1-k}y^k\right) = y^p + z^p = -x^p = (-x)^p.$$

On en déduit qu'alors p'^2 divise $-x^p$ et donc que p' divise x. De plus, comme $y \equiv -z\ [p']$, on a

$$\sum_{k=0}^{p-1}(-z)^{p-1-k}y^k \equiv \sum_{k=0}^{p-1} y^{p-1} \equiv py^{p-1} \equiv 0\ \ [p'],$$

c'est-à-dire que p' divise py^{p-1}. Par le lemme de Gauss, soit $p'|p$, c'est-à-dire $p' = p$, et dans ce cas p divise x, ce qui est exclu par hypothèse, soit p' divise y^{p-1} et divise donc y. Mais alors, x et y ne sont pas premiers entre eux : nouvelle contradiction.

Ainsi, $y+z$ et $\sum_{k=0}^{p-1}(-z)^{p-1-k}y^k$ sont premiers entre eux et de l'égalité (1) ci-dessus, on déduit l'existence de deux entiers a et α tels que

$$y + z = a^p \quad \text{et} \quad \sum_{k=0}^{p-1}(-z)^{p-1-k}y^k = \alpha^p.$$

(Si le produit de deux entiers u et v premiers entre eux est une puissance k-ième, alors u et v sont tous les deux des puissances k-ièmes comme nous l'avons prouvé à l'exercice 4.22).

Par symétrie, on prouve de même qu'il existe deux entiers b, c tels que $x + y = c^p$ et $x + z = b^p$.

3. Soit m un entier non divisible par q. Par le petit théorème de Fermat, on sait que $m^{q-1} \equiv 1\ [q]$, ce qui donne $(m^p)^2 \equiv 1\ [q]$ et donc, comme q est premier, $m^p \equiv 1\ [q]$ ou $m^p \equiv -1\ [q]$ ($\mathbb{Z}/q\mathbb{Z}$ est un corps).

Supposons par l'absurde qu'aucun des trois entiers x, y, z n'est divisible par q. Alors on a $x^p \equiv \pm 1$, $y^p \equiv \pm 1$ et $z^p \equiv \pm 1$ modulo q et en sommant, on obtient que $x^p + y^p + z^p$ est congru à 3, 1, -1 ou -3 modulo q (selon le nombre de signe $+$). C'est absurde (car $q > 5$). Donc l'un des trois entiers est divisible par q. On peut évidemment supposer sans perte de généralité que c'est x. Alors on a $yz \not\equiv 0\ [q]$, puisque x, y, z sont deux à deux premiers entre eux.

4. Toutes les congruences suivantes sont modulo q. On sait d'après la question 2 que $y + z = a^p$, $x + y = c^p$ et $x + z = b^p$. Il en résulte que $b^p + c^p - a^p = 2x \equiv 0$. On a évidemment $y \equiv c^p$ puisque $x \equiv 0$. Par ailleurs, q ne divise pas y, donc ne divise pas c. Il en résulte que $y \equiv \pm 1$. De la même manière, on obtient $z \equiv \pm 1$. Si maintenant q ne divise pas a, on a $a^p \equiv \pm 1$. Mais alors $c^p + b^p - a^p$ est congru à $3, 1, -1$ ou -3 modulo q, ce qui est impossible (car q divise $c^p + b^p - a^p$). On conclut que q divise a.

Regardons maintenant α^p. Sachant que $y+z \equiv a^p \equiv 0$, on en déduit que

$$\alpha^p = \sum_{k=0}^{p-1} (-z)^{p-1-k} y^k \equiv \sum_{k=0}^{p-1} y^{p-1} \equiv p y^{p-1}.$$

Nous avons montré que $y \equiv \pm 1$. Il s'en suit que $\alpha^p \equiv p(-1)^{p-1} \equiv p$ (car $p-1$ est pair). C'est absurde puisque, d'après la question 3, une puissance p ième est congrue à 0, -1 ou 1 modulo q. Nous aboutissons dans tous les cas à une contradiction. Il ne peut exister de triplet $(x,y,z) \in \mathbb{Z}^3$ tel que $xyz \not\equiv 0\ [p]$ et $x^p + y^p + z^p = 0$. ◁

Sophie Germain (1776-1831) est quasiment la seule femme mathématicienne de son temps. Elle suivit les cours de l'École polytechnique par correspondance car les femmes n'y étaient pas admises et c'est sous le pseudonyme masculin de Maurice Leblanc qu'elle écrivit à Gauss pour lui faire part de ses découvertes arithmétiques. C'est dans cette correspondance qu'apparaît le théorème ci-dessus. Le plus grand nombre premier de Sophie Germain connu en 2001 est $109433307 \times 2^{66452} - 1$ *et il comporte* 20013 *chiffres. On conjecture qu'il en existe une infinité.*

Chapitre 5

Polynômes

La théorie des équations polynomiales, qui précède de loin la définition formelle des polynômes, a été le propos essentiel de l'algèbre jusqu'au XIX^{e} *siècle. Elle est à l'origine de nombreuses notions : corps, nombres algébriques... Son développement est lié aux extensions successives de la notion de nombre : introduction des nombres négatifs, des nombres irrationnels, puis des nombres complexes.*

Dès la plus haute Antiquité, on rencontre des exemples de résolutions d'équations. Les Babyloniens savent résoudre l'équation du second degré et les Grecs en font la base même de leur géométrie.

Après l'Antiquité, il faudra attendre le XVI^{e} *siècle pour que des progrès substantiels apparaissent, dus à l'école italienne. Scipione del Ferro, Tartaglia et Cardan apportent la solution de l'équation du troisième degré. L'équation générale est ramenée à la forme réduite* $x^3 + px + q = 0$, *dont une solution s'écrit*

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}.$$

Cette solution soulève des difficultés : si $\dfrac{q^2}{4} + \dfrac{p^3}{27}$ *est négatif, cas où l'équation a des racines — on le sait depuis Archimède — on ne peut pas calculer* x. *Pour lever la difficulté, Cardan introduit timidement de nouveaux nombres, «impossibles» ou «imaginaires». Ferrari et Bombelli résolvent l'équation du quatrième degré.*

Grâce à l'école italienne, la théorie générale des équations algébriques se précise. L'équation étant mise sous la forme $\mathrm{P}(x) = 0$, *on prend conscience de l'importance du degré de* P *pour le nombre de solutions. On découvre que si* a *est une racine de* P, *on peut factoriser par* $x - a$. *Les relations entre les coefficients et les fonctions symétriques des racines d'un polynôme apparaissent chez Viète (1540-1603), mais c'est Girard qui en 1629 leur donne toute leur extension. Suivi par Newton, il exprime les sommes des puissances des racines en fonction des coefficients. L'étude des fonctions symétriques des racines va se développer au* XVII^{e} *siècle avec Waring et au* XIX^{e} *siècle avec Cauchy.*

Au XVII^{e} *siècle, la majorité des mathématiciens est convaincue qu'une équation de degré* n *possède* n *racines, celles-ci pouvant ne pas être réelles, mais il faut attendre d'Alembert pour trouver en 1724 une*

définition précise des nombres complexes (sous la forme $a+\sqrt{-1}b$). En 1799, Gauss fournit plusieurs preuves rigoureuses du «théorème fondamental de l'algèbre» ou «théorème de d'Alembert-Gauss» .

Des progrès sont réalisés également dans l'étude du nombre de racines réelles, et de leur signe. En 1637, Descartes énonce la règle qui porte son nom sur le nombre de racines positives d'un polynôme. On trouve dans l'Algèbre *de Rolle (1690), la propriété suivante : entre deux solutions de l'équation $\mathrm{P}(x)=0$, il existe au moins une solution de l'équation $\mathrm{P}'(x)=0$. C'est Sturm qui formule, en 1829, les résultats les plus précis sur le nombre de racines réelles d'un polynôme.*

Après les succès de l'école italienne au XVI$^{\mathrm{e}}$ *siècle, les mathématiciens se sont attachés à trouver des formules analogues pour les degrés suivants. Les réflexions sur cette question prennent un tour nouveau avec les travaux de Lagrange (1771), qui étudie les permutations des racines d'une équation laissant invariantes certaines fonctions de ces racines. Ces idées sont approfondies par Cauchy et Ruffini. Abel donne une démonstration rigoureuse de l'impossibilité de résoudre par radicaux l'équation générale de degré 5 en 1829. Enfin, en introduisant la notion de groupe, Galois énonce la condition générale à laquelle satisfait toute équation résoluble par radicaux (1831).*

La distinction entre les nombres algébriques, racines d'un polynôme à coefficients entiers, et les autres qu'on nomme transcendants, date du XVII$^{\mathrm{e}}$ *siècle, mais il faut attendre 1844 pour que Liouville démontre l'existence de nombres transcendants et plus longtemps encore pour que soit démontrée la transcendance de e (par Hermite en 1872) et celle de π (par Lindemann en 1882).*

Quant à la définition formelle des polynômes et à l'étude de leur structure, elles cheminèrent tout au long du XIX$^{\mathrm{e}}$ *siècle au rythme lent du processus d'axiomatisation de l'algèbre : par exemple, Dedekind introduit la notion de corps et définit les idéaux vers 1870.*

Les cours modernes remontent l'histoire à l'envers et commencent presque toujours par introduire les polynômes en tant qu'objet formel. Suivant cet usage, les premiers exercices concernent les polynômes formels et la structure d'algèbre que l'on obtient.

5.1. Égalité polynomiale

Dans quels corps a-t-on $\mathrm{X}^4-\mathrm{X}^2+1=(\mathrm{X}^2-5\mathrm{X}+1)(\mathrm{X}^2+5\mathrm{X}+1)$?
(École polytechnique)

▷ **Solution.**
En développant, on obtient

$$(X^2 - 5X + 1)(X^2 + 5X + 1) = X^4 - 23X^2 + 1.$$

Remarquons que si K est un corps qui convient, les polynômes en question sont dans $K_0[X]$, K_0 étant le sous-corps premier de K, autrement dit le plus petit sous-corps de K. On sait que K_0 est isomorphe à $\mathbb{Q}$ (si K est de caractéristique nulle) ou à $\mathbb{Z}/p\mathbb{Z}$ (si K est de caractéristique p, avec p premier).

En caractéristique nulle, $23 \neq 1$, donc K est de caractéristique p premier et on doit avoir $23 \equiv 1 \pmod{p}$. Cela arrive uniquement pour $p = 2$ et $p = 11$.

Réciproquement, dans tout corps de caractéristique 2 ou 11, $23 = 1$ et l'égalité $X^4 - X^2 + 1 = (X^2 - 5X + 1)(X^2 + 5X + 1)$ est vérifiée.

Conclusion. Les corps de caractéristique 2 ou 11 sont les seuls corps répondant au problème. ◁

5.2. Une sous-algèbre de $\mathbb{R}[X]$

Soit A la sous-algèbre de $\mathbb{R}[X]$ engendrée par X^2 et X^3. Montrer que A n'est pas isomorphe à $\mathbb{R}[X]$.

(École normale supérieure)

Solution.

• Pour commencer, explicitons la sous-algèbre A. Pour tout entier $k \geqslant 2$, A contient X^k, car on peut écrire, si k est pair, $X^k = (X^2)^{\frac{k}{2}}$ et si k est impair, $X^k = X^3\,(X^2)^{\frac{k-3}{2}}$. Comme A est une sous-algèbre de $\mathbb{R}[X]$, elle contient l'unité de $\mathbb{R}[X]$ et donc $\mathbb{R}$. Ainsi A contient $\underset{k\neq 1}{\mathrm{Vect}}(X^k)$.

Mais il est clair que $\underset{k\neq 1}{\mathrm{Vect}}(X^k)$ est une sous-algèbre de $\mathbb{R}[X]$ qui contient X^2 et X^3. En effet, c'est par définition un sous-espace vectoriel, qui contient 1 et qui est stable pour la multiplication, car pour tout $(k, \ell) \in (\mathbb{N} \setminus \{1\})^2$, on a $X^k X^\ell = X^{k+\ell}$ et $k + \ell \in \mathbb{N} \setminus \{1\}$. C'est donc A.

• Soit Φ un morphisme d'algèbres de $\mathbb{R}[X]$ dans A. Il est déterminé de manière unique par la donnée de $P = \Phi(X)$. On a, pour tout $Q \in \mathbb{R}[X]$, $\Phi(Q) = Q \circ P$ et en particulier, $\deg \Phi(Q) = \deg P \deg Q$. Si Φ était surjective, $\deg P$ diviserait à la fois 2 et 3 puisque X^2 et X^3 seraient dans l'image. Nécessairement $\deg P$ vaudrait 1 ce qui est exclu pour un polynôme de A.

Le morphisme Φ ne peut pas être un isomorphisme de $\mathbb{R}[X]$ sur A.

Conclusion. A n'est pas isomorphe à $\mathbb{R}[X]$. ◁

On remarque cependant que A *et* $\mathbb{R}[X]$ *ont même corps de fractions.*

On peut aussi démontrer que A *n'est pas factoriel, donc n'est pas principal (cf. exercice 3.10), ce qui fournit une autre solution de l'exer-*

cice. En effet, supposons A *factoriel. On se rappelle que* X *n'est pas dans* A. *On en déduit que* X^2 *et* X^3 *sont irréductibles (car un diviseur d'un élément dans* A *est encore un diviseur dans* $\mathbb{R}[X]$*). On écrit alors*

$$(X^3)^2 = (X^2)^3.$$

On remarque que X^2 *divise le côté droit de l'égalité et est premier avec* X^3*, donc avec son carré, ce qui fournit la contradiction voulue.*

Les exercices suivants ont pour thème commun les propriétés arithmétiques de l'anneau K[X] *où* K *est un corps. Comme* $\mathbb{Z}$*, il s'agit d'un anneau euclidien. Cela explique qu'on y dispose des résultats fondamentaux : principalité, théorèmes de Bezout et de Gauss, décomposition en produit de polynômes irréductibles...*

Lorsque le corps de base est $\mathbb{C}$*, les irréductibles de* $\mathbb{C}[X]$ *sont les polynômes de degré* 1*, et les problèmes de divisibilité sont très simplifiés comme le montrent les trois exercices suivants.*

5.3. Condition de divisibilité (1)

Soit a, b deux réels strictement positifs. Quels sont les entiers naturels n tels que $X^2 - (a^2 + b^2)$ divise $X^{2n} - (a^n + b^n)^2$?
(École Polytechnique)

▷ **Solution.**

Il faut et il suffit que $\pm\sqrt{a^2 + b^2}$ soient racines de $X^{2n} - (a^n + b^n)^2$, ce qui donne $(a^2 + b^2)^n = (a^n + b^n)^2$. Visiblement $n = 2$ convient mais pas 0 et 1.

Soit $\rho = \sqrt{a^2 + b^2}$ et $\theta \in \left]0, \frac{\pi}{2}\right[$ tel que $a + ib = \rho e^{i\theta}$. En divisant par ρ^{2n} et en prenant la racine carrée, on obtient, $1 = \cos^n\theta + \sin^n\theta$. Si $n > 2$, on a $\cos^n\theta + \sin^n\theta < \cos^2\theta + \sin^2\theta = 1$, puisque $\sin\theta$ et $\cos\theta$ appartiennent à $]0, 1[$.

Conclusion. La seule solution est $n = 2$. ◁

5.4. Condition de divisibilité (2)

Donner une condition nécessaire et suffisante sur p et q dans $\mathbb{C}$ pour que $X^8 + X^4 + 1$ divise $X^{8m} + pX^{4m} + q$ où $m \in \mathbb{N}^*$ est fixé.
(École polytechnique)

▷ **Solution.**

Le complexe ω est racine du polynôme $P = X^8 + X^4 + 1$ si et seulement si ω^4 est racine de $X^2 + X + 1$, c'est-à-dire si $\omega^4 = j$ ou $\overline{j}$. Les racines de

P' sont donc simples : il s'agit des racines quatrièmes de j et des racines quatrièmes de $\overline{j}$. Il en résulte que P divise $Q = X^{8m} + pX^{4m} + q$ si et seulement si toute racine ω de P est racine de Q. Or si ω est racine de P, on a

$$Q(\omega) = (\omega^4)^{2m} + p(\omega^4)^m + q = j^{2m} + pj^m + q \text{ ou } \overline{j}^{2m} + p\overline{j}^m + q.$$

La condition nécessaire et suffisante pour que P divise Q est donc

$$j^{2m} + pj^m + q = \overline{j}^{2m} + p\overline{j}^m + q = 0.$$

Deux cas se présentent :

• Si 3 divise m, alors $j^m = \overline{j}^m = 1$ et la condition cherchée est $\boxed{p + q + 1 = 0}$.

• Si m n'est pas divisible par 3, alors $\{j^{2m} + pj^m + q, \overline{j}^{2m} + p\overline{j}^m + q\}$ est égal à $\{j^2 + pj + q, j + pj^2 + q\}$. On obtient $j^2 + pj + q = j + pj^2 + q = 0$ et on trouve finalement la condition $\boxed{p = q = 1}$. ◁

5.5. Condition pour que $(P')^p$ divise P^q

1. Trouver les polynômes $P \in \mathbb{C}[X]$ tels qu'existent p, q dans $\mathbb{N}^*$ tels que $(P')^p$ divise P^q.

2. Même question dans $K[X]$, où K est un corps quelconque de caractéristique nulle.

(**École normale supérieure**)

▷ **Solution.**

1. Excepté le polynôme nul, les constantes ne sont pas solution. On suppose dans la suite que $n = \deg P \geqslant 1$. On peut aussi supposer P unitaire et l'écrire sous la forme $P = \prod_{i=1}^{k} (X - z_i)^{\alpha_i}$ où $z_1, \ldots, z_k$ sont des complexes deux à deux distincts. On a alors

$$P'(X) = nQ(X) \prod_{i=1}^{k} (X - z_i)^{\alpha_i - 1},$$

où Q est un polynôme unitaire de degré $k - 1$ ne s'annulant en aucun z_i.

L'hypothèse implique que toute racine de P' est racine de P. Cela impose $Q = 1$ donc $k = 1$ et $P = (X - z_1)^n$. Réciproquement, tous les polynômes $\lambda(X - z_1)^n$ sont solutions.

2. Là encore, on peut supposer que P est unitaire de degré $n \geqslant 1$. On le décompose en

$$P = \prod_{i=1}^{k} P_i^{\alpha_i},$$

où $P_1, \ldots, P_k$ sont des polynômes irréductibles unitaires distincts et $\alpha_1, \ldots, \alpha_k$ des entiers naturels non nuls. Si R est un polynôme irréductible qui divise P', alors R divise P'^p et donc P^q. Étant irréductible, il divise P. Les diviseurs irréductibles de P' sont donc parmi $P_1, \ldots, P_k$. On va déterminer pour chaque P_i son exposant dans P'.

On a $P = P_i^{\alpha_i} Q_i$, où Q_i est premier avec P_i. On en déduit que

$$P' = \alpha_i P_i^{\alpha_i - 1} P_i' Q_i + P_i^{\alpha_i} Q_i' = P_i^{\alpha_i - 1}(\alpha_i P_i' Q_i + P_i Q_i').$$

Ceci montre que $P_i^{\alpha_i - 1}$ divise P'. Montrons que $P_i^{\alpha_i}$ ne divise pas P'. Pour que $P_i^{\alpha_i}$ divise P', il faudrait que P_i divise $P_i' Q_i$. Le corps K étant de caractéristique nulle, on a $\deg P_i' = \deg P_i - 1$. Ainsi P_i ne divise pas P_i', et puisqu'il est irréductible, il est premier avec P_i'. Il est aussi premier avec Q_i par hypothèse, donc premier avec le produit $P_i' Q_i$: il ne divise pas $P_i' Q_i$ et donc $P_i^{\alpha_i}$ ne divise pas P'.

Ainsi P' s'écrit $P' = n \prod_{i=1}^{k} P_i^{\alpha_i - 1}$. Notons d_i le degré de P_i. On a alors $n = \deg P = \sum_{i=1}^{k} \alpha_i d_i$, et $n - 1 = \deg P' = \sum_{i=1}^{k} (\alpha_i - 1) d_i$. On en déduit que $\sum_{i=1}^{k} d_i = 1$. Chaque P_i étant irréductible donc de degré $d_i \geqslant 1$, on en déduit que $k = 1$ et $d_1 = 1$. Le polynôme P_1 étant de degré 1 et unitaire, il existe $z_1 \in K$ tel que $P_1 = X - z_1$ et donc $P = (X - z_1)^{\alpha_1}$. On trouve donc le même résultat que dans $\mathbb{C}[X]$. ◁

Les exercices ci-après concernent des équations dans l'anneau $\mathbb{C}[X]$.

5.6. Une équation fonctionnelle (1)

Trouver les $P \in \mathbb{R}[X]$ tels que $P(X) = P(1 - X)$.
(École Polytechnique)

▷ **Solution.**

Soit P un polynôme solution. Si α est une racine (complexe) de P, alors $1 - \alpha$ est aussi racine de P. Si α est différent de $\frac{1}{2}$, on a $1 - \alpha \neq \alpha$, et P est divisible par le polynôme

$$(X - \alpha)(X - 1 + \alpha) = X^2 - X + \alpha(1 - \alpha)$$

(on peut noter que tout polynôme de la forme $X^2 - X + \lambda$ est solution du

problème). Notons alors Q le quotient : $P(X) = (X^2 - X + \alpha(1-\alpha))Q(X)$. Si on compose cette identité par $1 - X$, on obtient que $Q(1-X) = Q(X)$. On peut itérer l'argument précédent avec Q. On en déduit que P a la forme suivante :

$$P = R(X) \prod_{i=1}^{p} (X^2 - X + \lambda_i),$$

où les λ_i sont des complexes et où R est un polynôme dont la seule racine complexe est $\frac{1}{2}$. On peut bien entendu supposer P (et donc R) unitaire. On écrit $R = \left(X - \frac{1}{2}\right)^k$. Si on calcule $R(1-X)$ on voit que P est solution si et seulement si k est pair. On peut noter que $\left(X - \frac{1}{2}\right)^{2q} = \left(X^2 - X + \frac{1}{4}\right)^q$. On peut donc conclure que les polynômes unitaires de $\mathbb{C}[X]$ solutions sont ceux qui s'écrivent sous la forme

$$P = \prod_{i=1}^{n} (X^2 - X + \mu_i),$$

où les μ_i sont des complexes pas forcément distincts. Pour obtenir les solutions de $\mathbb{R}[X]$ il suffit de regarder à quelle condition un polynôme de la forme ci-dessus est à coefficients réels. Cela a lieu si et seulement si l'ensemble des μ_i non réels (pris avec multiplicité) est stable par conjugaison. ◁

De l'exercice suivant, nous donnons une solution élémentaire faisant intervenir presque tous les aspects des polynômes : dérivation, considérations arithmétiques, calcul des coefficients... Nous trouverons une autre solution, plus éclairante, de cette équation dans le chapitre 3 (voir l'exercice 3.14), où elle sera interprétée comme la recherche d'unités dans une extension quadratique de l'anneau $\mathbb{C}[X]$.

5.7. Une équation fonctionnelle (2)

Trouver les couples $(P, Q) \in \mathbb{C}[X]^2$ tels que $P^2 = 1 + (X^2 - 1)Q^2$.
(École polytechnique)

▷ **Solution.**

• On commence par une analyse et on se donne une solution (P, Q).

Si $Q = 0$, on obtient $P = \pm 1$. On supposera dans la suite Q non nul. Dans ce cas, on a $n = \deg P \geqslant 1$ et $\deg(1 - P^2) = 2n$ de sorte que $2 + 2\deg(Q) = 2n$, *i.e.* $\deg(Q) = n - 1$. Si a est le coefficient dominant de P et b celui de Q, le coefficient de X^{2n} dans $P^2 + (1 - X^2)Q^2$ est $a^2 - b^2$.

On a donc $a = \pm b$. Observons que $(-P, Q)$, $(P, -Q)$ et $(-P, -Q)$ sont encore des couples solutions et que P et Q sont premiers entre eux par le théorème de Bezout.

En dérivant la relation on obtient

$$2PP' - 2XQ^2 + 2(1 - X^2)QQ' = 0.$$

Il en résulte que $Q|PP'$ et comme P et Q sont premiers entre eux, le théorème de Gauss nous permet d'affirmer que $Q|P'$. Ces deux polynômes étant de même degré, ils sont associés. Et, au vu des coefficients dominants, on a soit $P' = nQ$, soit $P' = -nQ$.

Dans les deux cas, on obtient $n^2QQ' = P'P''$ et $n^2Q^2 = P'^2$. En remplaçant dans la relation obtenue plus haut multipliée par n^2, on en déduit que

$$2(1 - X^2)P'P'' - 2XP'^2 + 2n^2PP' = 0.$$

En simplifiant par $2P' \neq 0$, on obtient l'équation différentielle vérifiée par P

$$(E_n) \qquad n^2P - XP' + (1 - X^2)P'' = 0.$$

Cherchons P sous la forme $a_0 + a_1X + \cdots + a_nX^n$. Après un petit calcul, on voit que l'équation différentielle (E_n) équivaut aux relations suivantes sur les coefficients de P :

(i) pas de contrainte sur a_n ;

(ii) $a_{n-1} = 0$;

(iii) pour tout $k \in [\![0, n-2]\!]$, $a_{k+2} = \dfrac{k^2 - n^2}{(k+1)(k+2)} a_k$.

Il en résulte que tous les coefficients d'indice de parité opposée à celle de n sont nuls (c'est-à-dire que P a la parité de n). Pour les autres on obtient

$$a_{n-2k} = a_n \frac{(-1)^k n(n-k-1)!}{4^k k!(n-2k)!} = a_n \frac{(-1)^k}{4^k} \frac{n}{n-k} C_{n-k}^k.$$

Pour tout $n \geqslant 1$, il existe donc un unique polynôme de degré n, unitaire, solution de l'équation différentielle (E_n). Notons-le A_n. Ses coefficients sont donnés en prenant $a_n = 1$ dans les formules ci-dessus. Il résulte alors de cette analyse, que (P, Q) doit être l'un des couples $(\alpha A_n, \dfrac{\varepsilon}{n} \alpha A_n')$ où $\varepsilon = \pm 1$ et où α est un nombre complexe non nul quelconque.

• Synthèse : soit $P = \alpha A_n$ avec α complexe non nul et $Q = \dfrac{1}{n} P'$ (on peut se limiter à étudier le cas $\varepsilon = 1$). Déterminons les valeurs de α pour lesquelles le couple (P, Q) est effectivement solution. On va essayer de remonter les implications faites lors de l'analyse. En multipliant l'équation (E_n) par $2P'$, on obtient $[n^2P^2]' = [(X^2 - 1)P'^2]'$. Ainsi, il existe une

constante complexe c telle que $n^2P^2 + (1 - X^2)P'^2 = c$. On aimerait avoir $c = n^2$. Or pour calculer c, il suffit de regarder la valeur en 0 du polynôme de gauche : $c = n^2P(0)^2 + P'(0)^2 = n^2\alpha^2A_n(0)^2 + \alpha^2A'_n(0)^2$. Distinguons suivant la parité de n.

⋆ Si $n = 2p$ est pair, le terme de degré 1 de A_n est nul et donc $A'_n(0) = 0$. De plus, en faisant $n = 2p$ et $k = p$ dans les formules précédentes, on obtient

$$A_n(0) = \frac{(-1)^p(2p)(p-1)!}{4^p p!} = \frac{(-1)^p}{2^{2p-1}}.$$

Il y a donc exactement deux valeurs de α, à savoir $\pm 2^{n-1}$, qui conviennent.

⋆ Si $n = 2p + 1$ est impair, cette fois on obtient $A_n(0) = 0$ et $A'_n(0) = (-1)^p \dfrac{n}{4^p}$. Ici encore, on trouve $\alpha = \pm 2^{n-1}$.

En conclusion, les couples solutions de l'équation sont $(\pm 1, 0)$ et les couples $(\pm 2^{n-1}A_n, \pm \dfrac{2^{n-1}}{n} A'_n)$, $n \geqslant 1$. Il s'agit de polynômes réels. ◁

Ces polynômes sont bien connus. Le lecteur aura peut-être reconnu l'équation différentielle vérifiée par le n-ième polynôme de Tchebychev de première espèce T_n. *On a en fait* $T_n = 2^{n-1}A_n$ *et* $U_{n-1} = \dfrac{2^{n-1}}{n} A'_n$, *où* U_{n-1} *est le n-ième polynôme de Tchebychev de seconde espèce. Ces polynômes seront étudiés dans les exercices 5.41 et 5.42.*

On ne manquera pas de rapprocher le théorème de Liouville ci-après du grand théorème de Fermat.

5.8. Un théorème de Liouville (1879)

Soit $n \in \mathbb{N}$, $n \geqslant 3$. Montrer qu'il n'existe pas de polynômes P, Q, R de $\mathbb{C}[X]$ tels que $P^n + Q^n + R^n = 0$ sans que P, Q et R soient tous égaux, à une constante multiplicative près, à un même polynôme.

(École normale supérieure)

Solution.

Notons E l'ensemble des triplets $(P, Q, R) \in \mathbb{C}[X]^3$ vérifiant

$$P^n + Q^n + R^n = 0. \quad (1)$$

Considérons un triplet non nul (P, Q, R) de E. Quitte à diviser P, Q, R par leur pgcd, on peut également supposer que $\text{pgcd}(P, Q, R) = 1$. Il apparaît alors que P, Q, R sont deux à deux premiers entre eux. En effet, si P_0 est un diviseur irréductible de P et Q, P_0 divise $P^n + Q^n = -R^n$ et donc P_0 divise R d'après le lemme d'Euclide. Il en est de même pour

les deux autres cas. Le problème se ramène alors à montrer que P, Q, R sont des polynômes constants.

Supposons, sans perte de généralité, que $\deg(R) \leqslant \deg(Q) \leqslant \deg(P)$. Observons que si R est constant, $P^n + Q^n = \prod_{k=1}^{n} (P - \xi_k Q)$ l'est aussi, $\xi_1, \ldots, \xi_n$ étant les racines n-ièmes de -1. Donc les $P - \xi_k Q$ sont tous constants et on en déduit que P et Q sont constants.

Supposons donc par l'absurde que $\deg R \geqslant 1$. En dérivant la relation (1) et en simplifiant par n, il vient

$$P^{n-1}P' + Q^{n-1}Q' + R^{n-1}R' = 0 \quad (2).$$

En multipliant (1) par R', (2) par R et en faisant la différence, on obtient :

$$P^{n-1}(PR' - P'R) = Q^{n-1}(Q'R - QR').$$

Les polynômes P et Q étant premiers entre eux, on déduit du théorème de Gauss que P^{n-1} divise $Q'R - QR'$. Le polynôme $QR' - Q'R$ n'est pas nul, sinon la fraction $\frac{R}{Q}$ aurait une dérivée nulle, elle serait donc constante et Q et R seraient proportionnels, ce qui est impossible car ils sont premiers entre eux et de degré $\geqslant 1$. En regardant les degrés on en déduit que

$$\begin{aligned} (n-1)\deg P &\leqslant \max(\deg(Q'R), \deg(QR')) = \deg Q + \deg R - 1 \\ &\leqslant 2\deg(P) - 1, \end{aligned}$$

ce qui est absurde car $n \geqslant 3$.

Conclusion. E est l'ensemble des triplets (aP, bP, cP) où $P \in \mathbb{C}[X]$ et où (a, b, c) est un triplet de complexes vérifiant $a^n + b^n + c^n = 0$. ◁

Dans le langage de la géométrie algébrique, le résultat de l'exercice montre que la courbe algébrique $X^n + Y^n = 1$ *n'est pas unicursale*[1]*, c'est-à-dire n'admet pas de paramétrisation rationnelle, pour* $n \geqslant 3$*. En revanche, le cercle est une courbe unicursale et pour* $n = 2$*, l'équation a des solutions non triviales, par exemple* $(1 - X^2)^2 + (2X)^2 = (1 + X^2)^2$.

5.9. Théorème de Mason (1984)

1. Soit A, B, C trois polynômes de $\mathbb{C}[X]$, non constants, premiers entre eux dans leur ensemble et tels que $A + B = C$. Soit m le nombre de racines complexes distinctes de ABC.

1. PERRIN (D.), *Géométrie algébrique, une introduction*, Savoirs actuels, InterÉditions/CNRS Éditions, 1995, p. 1-5.

Montrer que $A\left(\frac{A'}{A}-\frac{C'}{C}\right)=B\left(\frac{C'}{C}-\frac{B'}{B}\right)$.
En déduire que $\max(\deg A, \deg B, \deg C) < m$.

2. Retrouver le résultat de l'exercice précédent.

(École normale supérieure)

Solution.

1. La relation $A+B=C$ conduit à $\frac{C'}{C}=\frac{A'+B'}{A+B}$ qui s'écrit encore

$$A\left(\frac{A'}{A}-\frac{C'}{C}\right)=B\left(\frac{C'}{C}-\frac{B'}{B}\right).$$

Décomposons en éléments simples ces différentes fractions rationnelles. Les polynômes A, B, C sont deux à deux premiers entre eux, car si deux de ces polynômes ont une racine commune z, celle-ci est aussi racine du troisième, car $A+B=C$, et $X-z$ divise $\operatorname{pgcd}(A,B,C)$. Posons

$$A=\alpha\prod_{i=1}^{n_A}(X-a_i)^{\alpha_i}\quad B=\beta\prod_{i=1}^{n_B}(X-b_i)^{\beta_i}\quad C=\gamma\prod_{i=1}^{n_C}(X-c_i)^{\gamma_i},$$

où les a_i, b_i, c_i sont les racines distinctes de A, B, C, respectivement. On a donc $m=n_A+n_B+n_C$. On obtient alors

$$\frac{A'}{A}=\sum_{i=1}^{n_A}\frac{\alpha_i}{X-a_i},\qquad \frac{B'}{B}=\sum_{i=1}^{n_B}\frac{\beta_i}{X-b_i},\qquad \frac{C'}{C}=\sum_{i=1}^{n_C}\frac{\gamma_i}{X-c_i}.$$

Le polynôme $U=\prod_{i=1}^{n_A}(X-a_i)\prod_{i=1}^{n_B}(X-b_i)\prod_{i=1}^{n_C}(X-c_i)$ est un dénominateur commun à ces trois fractions. Il existe donc deux polynômes non nuls P et Q tels que $\frac{A'}{A}-\frac{C'}{C}=\frac{P}{U}$ et $\frac{C'}{C}-\frac{B'}{B}=\frac{Q}{U}$. De plus, les degrés de P et Q sont majorés par $\deg U-1=m-1$ car les fractions sont de degré -1. De l'égalité $AP=BQ$, démontrée plus haut, on déduit, puisque A et B sont premiers entre eux, que A divise Q et B divise P. Il en résulte que $\deg A\leqslant m-1$ et $\deg(B)\leqslant m-1$. Cela vaut aussi pour $C=A+B$. On obtient finalement $\max(\deg A,\deg B,\deg C)\leqslant m-1$.

2. Soit (P, Q, R) un triplet non nul de polynômes premiers entre eux vérifiant $P^n+Q^n+R^n=0$. On a vu dans l'exercice précédent que, si l'un des trois polynômes est constant, les deux autres le sont aussi. Dans le cas contraire, le théorème de Mason conduit à

$$n\max(\deg P,\deg Q,\deg R)<m\leqslant\deg(PQR),$$

et

$$n\max(\deg P,\deg Q,\deg R)<3\max(\deg P,\deg Q,\deg R).$$

C'est impossible si $n \geqslant 3$. $\lhd$

L'analogue du théorème de Mason pour les entiers aurait d'importantes applications en arithmétique : il s'agit de la conjecture abc, formulée par Masser et Oesterlé[2].

Le résultant donne une condition portant sur les coefficients de deux polynômes pour qu'ils soient premiers entre eux.

5.10. Résultant de deux polynômes

Soit F et G deux polynômes non constants à coefficients complexes de degrés respectifs n et m.

1. On considère

$$\Phi : \begin{array}{ccc} \mathbb{C}_{m-1}[X] \times \mathbb{C}_{n-1}[X] & \longrightarrow & \mathbb{C}_{n+m-1}[X] \\ (U, V) & \longmapsto & UF + VG \end{array}.$$

Montrer que Φ est bien définie, qu'elle est linéaire, et donner une condition nécessaire et suffisante pour qu'elle soit injective. Écrire la matrice de Φ dans les bases canoniques.

Le déterminant de cette matrice, noté $\mathrm{Res}(F, G)$, est appelé *résultant* des polynômes F et G.

2. Soit $\Gamma = \{(F(t), G(t)) \in \mathbb{C}^2,\ t \in \mathbb{C}\}$. Établir l'existence de $R \in \mathbb{C}[X, Y]$ tel que, pour tout $(x, y) \in \mathbb{C}^2$,

$$(x, y) \in \Gamma \iff R(x, y) = 0.$$

(École polytechnique)

$\rhd$ **Solution.**

1. • Tout d'abord Φ est bien définie : si $(U, V) \in \mathbb{C}_{m-1}[X] \times \mathbb{C}_{n-1}[X]$, on a $\deg UF \leqslant m - 1 + n$ et $\deg VG \leqslant n - 1 + m$ d'où l'on déduit $\deg(UF + VG) \leqslant m + n - 1$. Ensuite Φ est clairement linéaire.

Supposons Φ injective. Alors Φ est surjective, car

$$\dim\left(\mathbb{C}_{m-1}[X] \times \mathbb{C}_{n-1}[X]\right) = m + n = \dim \mathbb{C}_{n+m-1}[X]$$

et il existe $(U, V) \in \mathbb{C}_{m-1}[X] \times \mathbb{C}_{n-1}[X]$ tel que $UF + VG = \Phi(U, V) = 1$ donc, d'après le théorème de Bezout, F et G sont premiers entre eux.

Réciproquement, supposons que F et G sont premiers entre eux et considérons (U, V) dans $\operatorname{Ker} \Phi$. On obtient $UF = -VG$ donc F divise VG.

2. Pour l'énoncé de cette conjecture nous renvoyons le lecteur intéressé à LANG (S.), *Algebra*. Addison-Wesley, 3e éd., 1993, p. 194-199 ou au chapitre 5 de NATHANSON (M.B.), *Elementary Methods in Number Theory*, GTM 195, Springer-Verlag, 2000.

Comme F est premier avec G, le lemme de Gauss assure que F divise V. Puisque $\deg V < \deg F$, on conclut que $V = 0$. De même, $U = 0$. Donc Φ est injective.

On conclut que Φ est injective si, et seulement si, F et G sont premiers entre eux, ou encore, puisque $\mathbb{C}$ est algébriquement clos si, et seulement si, F et G n'ont aucune racine commune.

• Il s'agit de prendre dans l'espace vectoriel $\mathbb{C}_{m-1}[X] \times \mathbb{C}_{n-1}[X]$ la base $(E_i)_{0 \leqslant i \leqslant n+m-1}$ où E_i vaut $(X^i, 0)$ si $0 \leqslant i \leqslant m-1$ et $(0, X^{i-m})$ si $i \geqslant m$. Sur $\mathbb{C}_{n+m-1}[X]$, il s'agit de la base $(1, X, X^2, \ldots, X^{n+m-1})$. Si $F = a_n X^n + \cdots + a_1 X + a_0$ et $G = b_m X^m + \cdots + b_1 X + b_0$, la matrice de Φ dans ces bases est alors

$$\begin{pmatrix}
a_0 & 0 & \ldots & 0 & b_0 & 0 & \ldots & \ldots & \ldots & 0 \\
a_1 & a_0 & & \vdots & b_1 & b_0 & & & & \vdots \\
a_2 & a_1 & \ddots & 0 & \vdots & \vdots & \ddots & & & \vdots \\
\vdots & \vdots & & a_0 & b_{m-1} & \vdots & \ddots & & & \vdots \\
\vdots & \vdots & & a_1 & b_m & \vdots & & \ddots & & \vdots \\
\vdots & \vdots & & \vdots & 0 & b_m & & & \ddots & 0 \\
a_{n-1} & \vdots & & \vdots & 0 & 0 & \ddots & & b_1 & b_0 \\
a_n & a_{n-1} & & a_{n-m+1} & 0 & 0 & & \ddots & & b_1 \\
0 & a_n & & \vdots & 0 & 0 & & & \ddots & \vdots \\
0 & & \ddots & \vdots & \vdots & \vdots & & & \ddots & \vdots \\
0 & 0 & \ldots & a_n & 0 & 0 & \ldots & \ldots & 0 & b_m
\end{pmatrix}.$$

Le résultant de F et G, Res(F, G), est nul si, et seulement si, F et G ont une racine commune. Le *discriminant de* F est le résultant de F et F'. Il est nul si, et seulement si, F a une racine double. Si $F = aX^2 + bX + c$, on trouve $\mathrm{Res}(F, F') = a(4ac - b^2)$ et si $F = X^3 + pX + q$, on obtient $\mathrm{Res}(F, F') = 4p^3 + 27q^2$.

2. Soit $(x, y) \in \mathbb{C}^2$. Dire que $(x, y) \in \Gamma$ équivaut à dire qu'il existe $t \in \mathbb{C}$ tel que $F(t) = x$ et $G(t) = y$. Cela signifie que $F - x$ et $G - y$ ont une racine commune ce qui donne à l'aide du résultant

$$(x, y) \in \Gamma \iff \mathrm{Res}(F - x, G - y) = 0.$$

On définit R(X, Y) comme étant le déterminant

$$\begin{vmatrix}
a_0 - X & 0 & \dots & 0 & b_0 - Y & 0 & \dots & \dots & 0 \\
a_1 & a_0 - X & & \vdots & b_1 & b_0 - Y & & & \vdots \\
a_2 & a_1 & \ddots & 0 & \vdots & \vdots & \ddots & & \vdots \\
\vdots & \vdots & & a_0 - X & b_{m-1} & \vdots & \ddots & & \vdots \\
\vdots & \vdots & & a_1 & b_m & \vdots & \ddots & & \vdots \\
\vdots & \vdots & & \vdots & 0 & b_m & & \ddots & 0 \\
a_{n-1} & \vdots & & \vdots & 0 & 0 & \ddots & b_1 & b_0 - Y \\
a_n & a_{n-1} & & a_{n-m+1} & 0 & 0 & \ddots & & b_1 \\
0 & a_n & & \vdots & 0 & 0 & & \ddots & \vdots \\
0 & & \ddots & \vdots & \vdots & \vdots & & \ddots & \vdots \\
0 & 0 & \dots & a_n & 0 & 0 & \dots & 0 & b_n
\end{vmatrix}.$$

On a $R \in \mathbb{C}[X, Y]$ et $(x, y) \in \Gamma \iff R(x, y) = 0$. ◁

On peut définir de la même manière le résultant de deux polynômes de $K[X]$ *pour un corps* K *quelconque. Si* K *n'est pas algébriquement clos, l'annulation du discriminant n'est pas une condition suffisante d'existence d'une racine double pour les polynômes de degré* $\geqslant 4$. *Pour* $K = \mathbb{R}$ *et* $F = X^3 + pX + q$, F *possède trois racines réelles si et seulement si son discriminant* $4p^3 + 27q^2$ *est négatif ou nul.*

Bien qu'il y soit question de racines, les arguments arithmétiques sont essentiels dans l'exercice suivant.

5.11. Caractérisation d'un polynôme par les antécédents de deux points distincts

Soit P et Q deux polynômes non constants de $\mathbb{C}[X]$ tels que l'ensemble des racines de P (resp. $P - 1$) soit égal à l'ensemble des racines de Q (resp. $Q - 1$). Montrer que $P = Q$.

(École polytechnique)

▷ **Solution.**

• Si P et Q sont des polynômes, on notera $Z(P)$ l'ensemble des racines de P et $P \wedge Q$ le pgcd de P et Q. Soit $n = \deg P \geqslant 1$ et $m = \deg Q \geqslant 1$. On suppose sans perte de généralité que $n \geqslant m$ et on pose $R = P - Q$. On a $\deg R \leqslant n$ et on va prouver que R est nul en montrant qu'il a plus de $n + 1$ racines.

• Par hypothèse, R s'annule sur $Z(P)$ et sur $Z(P - 1)$, ensembles

évidemment disjoints. Déterminons le cardinal de ces ensembles. Le nombre de racines distinctes de P est égal à $\deg P - \deg(P \wedge P')$. En effet, si $P = \lambda \prod_{i=1}^{p} (X - z_i)^{n_i}$, où les z_i sont deux à deux distincts, on obtient $P \wedge P' = \prod_{i=1}^{p} (X - z_i)^{n_i - 1}$ et $\deg(P \wedge P') = (n_1 - 1) + \cdots + (n_p - 1) = n - p$.

• On a donc

$$\begin{aligned} \mathrm{Card}(Z(P)) &= n - \deg(P \wedge P') \quad \text{et} \\ \mathrm{Card}(Z(P-1)) &= \deg(P-1) - \deg((P-1) \wedge (P-1)') \\ &= n - \deg((P-1) \wedge P'). \end{aligned}$$

Mais comme P et P − 1 sont premiers entre eux, $(P-1) \wedge P'$ et $P \wedge P'$ sont deux diviseurs premiers entre eux de P′. En particulier, on a

$$\deg((P-1) \wedge P') + \deg(P \wedge P') \leqslant n - 1.$$

Il en résulte que

$$\mathrm{Card}(Z(R)) \geqslant 2n - \deg(P \wedge P') - \deg((P-1) \wedge P') \geqslant 2n - (n-1) = n+1.$$

On a donc R = 0, ce qu'il fallait démontrer. ◁

Bien entendu, on peut remplacer 0 *et* 1 *par deux complexes distincts* a *et* b *quelconques : un polynôme non constant* $P \in \mathbb{C}[X]$ *est uniquement déterminé si on connaît les antécédents de* a *et de* b *par* P *(i.e. les racines de* $P - a$ *et de* $P - b$*).*

Les deux exercices suivants utilisent encore le fait que le nombre de racines complexes distinctes de $P \in \mathbb{C}[X]$ *est égal à* $\deg P - \deg(P \wedge P')$.

5.12. Image réciproque d'une partie finie par un polynôme

Soit $P \in \mathbb{C}[X]$ non constant et E un sous-ensemble fini de $\mathbb{C}$. Montrer que $|P^{-1}(E)| \geqslant (|E| - 1)\deg P + 1$.

(École Polytechnique)

Solution.

Si $w \in \mathbb{C}$, le nombre d'antécédents de w par P est le nombre de racines distinctes de $P(X) - w$. Ce nombre vaut $\deg P - \deg(P - w) \wedge P'$. Notons $w_1, \ldots, w_r$ les éléments de E (avec $r = |E|$). On a donc

$$|P^{-1}(E)| = \sum_{i=1}^{r} (\deg P - \deg((P - w_i) \wedge P')) = r \deg P - \sum_{i=1}^{r} \deg((P - w_i) \wedge P')$$

Or, les polynômes $P - w_i$ sont deux à deux premiers entre eux (ils n'ont aucune racine commune). Il en est donc de même des polynômes $\text{pgcd}(P - w_i, P')$ et comme ils divisent tous P', leur produit aussi. Par conséquent, il vient

$$\sum_{i=1}^{r} \deg((P - w_i) \wedge P') = \deg \prod_{i=1}^{r} ((P - w_i) \wedge P') \leqslant \deg P' = \deg P - 1,$$

et $|P^{-1}(E)| \geqslant |E| \deg P - \deg P + 1$ ce qui est le résultat voulu. ◁

5.13. Sur le nombre d'antécédents d'un point par un polynôme

Soit $P \in \mathbb{C}[X]$ de degré $d \geqslant 1$. On note $n(z)$ le nombre de racines de l'équation $P(x) = z$. Donner une expression de $\sum_{z \in \mathbb{C}} (d - n(z))$.

(École Polytechnique)

▷ **Solution.**

Cet exercice utilise encore la formule

$$n(z) = d - \deg((P - z) \wedge P').$$

La somme considérée a l'air infinie, mais il n'y a en fait qu'un nombre fini de nombres complexes z tels que $d - n(z)$ soit non nul (z doit être l'image par P d'une racine de P' sans quoi $P - z$ et P' sont premiers entre eux). Notons $u_1, ..., u_p$ les racines distinctes de P' et n_i l'ordre de multiplicité de u_i. Supposons dans un premier temps que les nombres complexes $P(u_i)$ sont deux à deux distincts. Dans ce cas, on a pour tout i, $(P - P(u_i)) \wedge P' = (X - u_i)^{n_i}$ car u_j n'est pas racine de $P(X) - P(u_i)$ pour $j \neq i$, et u_i est racine d'ordre $n_i + 1$ de $P(X) - P(u_i)$ (il suffit d'écrire la formule de Taylor en u_i pour P pour s'en convaincre). On a alors

$$\begin{aligned} \sum_{z \in \mathbb{C}} (d - n(z)) &= \sum_{z \in \mathbb{C}} \deg((P - z) \wedge P') = \sum_{i=1}^{p} \deg((P - P(u_i)) \wedge P') \\ &= \sum_{i=1}^{p} n_i = \deg P' = d - 1. \end{aligned}$$

Que se passe-t-il maintenant si certains des $P(u_i)$ sont égaux ? Soit z l'une des valeurs $P(u_i)$ et supposons que z soit atteint avec $u_{i_1}, \ldots, u_{i_s}$. Dans ce cas, le pgcd de $P(X) - z$ et P' est égal à $\prod_{j=1}^{s} (X - u_{i_j})^{n_{i_j}}$ et son degré est égal à la somme des ordres de multiplicité des racines u_{i_j}.

Il en résulte que le résultat est identique. Dans tous les cas, la somme cherchée vaut $d-1$. ◁

Si L *est un sur-corps de* K, *deux polynômes* P, Q *de* $K[X]$ *peuvent être regardés comme des polynômes de* $L[X]$. *Le pgcd de* P *et* Q *ne va pas dépendre du corps dans lequel on se place : il est obtenu par l'algorithme d'Euclide et ses coefficients appartiennent au plus petit corps contenant les coefficients de* P *et* Q. *On dira que le pgcd est invariant par extension de corps. L'exercice suivant utilise ce fait.*

5.14. Polynôme rationnel inséparable de degré 5

1. Soit P un polynôme irréductible de $\mathbb{Q}[X]$. Démontrer que P n'a pas de racine double dans $\mathbb{C}$.

2. Soit $P \in \mathbb{Q}[X]$ de degré 5. On suppose que P admet une racine multiple dans $\mathbb{C}$. Montrer que P possède une racine dans $\mathbb{Q}$.

(**École normale supérieure**)

▷ **Solution.**

1. Comme $1 \leqslant \deg P' < \deg P$, les polynômes P' et P sont premiers entre eux dans $\mathbb{Q}[X]$, puisque P est irréductible. Ils le sont donc encore dans $\mathbb{C}[X]$ par invariance du pgcd par extension de corps. Si α était une racine double de P dans $\mathbb{C}$, on aurait $P(\alpha) = P'(\alpha) = 0$ et $X - \alpha$ serait un facteur commun à P et P' dans $\mathbb{C}[X]$. On aboutit à une contradiction.

2. Notons α une racine multiple complexe de P. Supposons par l'absurde, que P n'ait pas de racine rationnelle et regardons sa décomposition en éléments irréductibles dans $\mathbb{Q}[X]$. Il ne peut pas y avoir de facteur de degré 1 ; le polynôme P ne pouvant être irréductible d'après la première question, il s'écrit $P = QR$ avec $\deg Q = 2$, $\deg R = 3$ et Q, R tous deux irréductibles dans $\mathbb{Q}[X]$, puisque sans racine rationnelle.

Les polynômes Q et R ne peuvent avoir α comme racine commune dans $\mathbb{C}$. En effet, si on considère D le pgcd de Q et R dans $\mathbb{Q}[X]$, il est égal à 1 puisque Q et R sont deux polynômes irréductibles non proportionnels. Mais D est aussi le pgcd de Q et R en tant que polynômes de $\mathbb{C}[X]$, du fait de l'invariance du pgcd par extension de corps. Par conséquent, si on avait $Q(\alpha) = R(\alpha) = 0$, $X - \alpha$ diviserait D, qui serait de degré $\geqslant 1$, ce qui n'est pas le cas.

Par conséquent, comme $(X-\alpha)^2$ divise P, $(X-\alpha)^2$ divise soit Q, soit R. Mais alors Q ou R est un polynôme irréductible ayant une racine multiple dans $\mathbb{C}$, ce qui est impossible.

L'hypothèse faite au départ est fausse : P possède donc une racine rationnelle. ◁

Le thème commun aux exercices qui suivent est l'irréductibilité de polynômes à coefficients entiers. Rappelons qu'un polynôme non nul P *de* $\mathbb{Z}[X]$ *est dit irréductible si l'écriture* $P = QR$ *avec* $(Q, R) \in \mathbb{Z}[X]^2$ *impose* $Q = \pm 1$ *ou* $R = \pm 1$.

5.15. Un polynôme irréductible de $\mathbb{Z}[X]$

Soit $n \geqslant 2$ et $a_1, \ldots, a_n$ des éléments de $\mathbb{Z}$ deux à deux distincts. Montrer que le polynôme $P = (X - a_1)(X - a_2)\ldots(X - a_n) - 1$ est irréductible dans $\mathbb{Z}[X]$.

(École normale supérieure)

▷ **Solution.**

Supposons que $P = QR$ où Q et R sont dans $\mathbb{Z}[X]$. On a, pour tout k de $[\![1, n]\!]$, $Q(a_k)R(a_k) = P(a_k) = -1$ et donc, soit $Q(a_k) = 1 = -R(a_k)$, soit $Q(a_k) = -1 = -R(a_k)$, puisque ce sont des entiers. Dans les deux cas, on a $Q(a_k) + R(a_k) = 0$ Le polynôme $Q + R$ s'annule en n points distincts. Si $\deg(Q+R) < n$, on a alors $Q+R = 0$ et $P = -Q^2$ ce qui est absurde car $P(x)$ serait négatif pour tout x réel et de limite égale à $+\infty$ en $+\infty$. On a donc $\deg(Q + R) = n$. Mais alors, Q ou R est constant. Le produit des coefficients dominants de Q et R étant égal à 1, on a donc soit $Q = \pm 1$, soit $R = \pm 1$. Ainsi, P est irréductible dans $\mathbb{Z}[X]$. ◁

L'exercice suivant établit un critère fort utile pour montrer qu'un polynôme à coefficients entiers est irréductible.

5.16. Critère d'Eisenstein

1.a. On dit qu'un polynôme non nul de $\mathbb{Z}[X]$ est *primitif* si le pgcd de ses coefficients est égal à 1. Montrer que le produit de deux polynômes primitifs de $\mathbb{Z}[X]$ est primitif.

b. Pour $A \in \mathbb{Z}[X]$ non nul, on appelle *contenu de* A, et on note $c(A)$ le pgcd des coefficients de A. Soit A et B deux polynômes non nuls de $\mathbb{Z}[X]$. Montrer que $c(AB) = c(A)c(B)$.

2. Soit $A = a_nX^n + \cdots + a_1X + a_0 \in \mathbb{Z}[X]$ et p un nombre premier. On suppose que :

(*i*) p ne divise pas a_n ;

(*ii*) p divise $a_0, a_1, \ldots, a_{n-1}$;

(*iii*) p^2 ne divise pas a_0.

Montrer que A est irréductible dans $\mathbb{Q}[X]$.

(École normale supérieure)

Solution.

1.a. Soit $A = \sum_{k=0}^{n} a_k X^k$, $B = \sum_{k=0}^{m} b_k X^k$ des polynômes à coefficients entiers et $C = \sum_{k=0}^{m+n} c_k X^k = AB$. Supposons A et B primitifs et montrons, en raisonnant par l'absurde, que C est primitif. Si ce n'est pas le cas, il existe un nombre premier p qui divise tous les c_k. Pour tout $P \in \mathbb{Z}[X]$, on note $\overline{P}$ le projeté de P dans $(\mathbb{Z}/p\mathbb{Z})[X]$: si $P = \sum_{k\in\mathbb{N}} s_k X^k$, $\overline{P} = \sum_{k\in\mathbb{N}} \overline{s_k} X^k$ où $\overline{s_k}$ est la classe de s_k modulo p. Comme p divise tous les c_k, on a $\overline{C} = 0$ et donc $\overline{A}\,\overline{B} = \overline{AB} = \overline{C} = 0$. Mais $(\mathbb{Z}/p\mathbb{Z})[X]$ est intègre, puisque $\mathbb{Z}/p\mathbb{Z}$ est un corps, donc on a $\overline{A} = 0$ ou $\overline{B} = 0$. Autrement dit, p divise tous les coefficients de A ou tous les coefficients de B. Ceci est exclu.

On conclut que le produit de deux polynômes primitifs de $\mathbb{Z}[X]$ est encore primitif.

b. On peut écrire $AB = c(A)c(B)\dfrac{A}{c(A)}\dfrac{B}{c(B)}$. Les polynômes $\dfrac{A}{c(A)}$ et $\dfrac{B}{c(B)}$ sont alors primitifs, donc leur produit aussi d'après la question précédente et le contenu de AB est $c(A)c(B)$.

2. Montrons que si A n'est pas irréductible dans $\mathbb{Q}[X]$, alors il peut s'écrire $A = BC$ avec B et C dans $\mathbb{Z}[X]$ de degrés strictement inférieurs à celui de A.

Soit $\alpha = c(A)$ et $A' = \dfrac{1}{\alpha}A \in \mathbb{Z}[X]$; A' est primitif. Comme A est composé, par hypothèse, A' l'est aussi et on peut écrire $A' = B'C'$, avec B' et C' dans $\mathbb{Q}[X]$ de degrés strictement inférieurs à celui de A. Notons β (resp. γ) le produit des dénominateurs des coefficients de B' (resp. C'). Alors les polynômes $B = \beta B'$ et $C = \gamma C'$ sont dans $\mathbb{Z}[X]$ et $\beta\gamma A' = BC$. En passant aux contenus, on obtient $\beta\gamma = \beta\gamma c(A') = c(B)c(C)$. Par conséquent, on a

$$A = \alpha\left(\frac{1}{\beta}B\right)\left(\frac{1}{\gamma}C\right) = \alpha\left(\frac{1}{c(B)}B\right)\left(\frac{1}{c(C)}C\right) = \left(\frac{\alpha}{c(B)}B\right)\left(\frac{1}{c(C)}C\right)$$

et $\dfrac{\alpha}{c(B)}B$ et $\dfrac{1}{c(C)}C$ sont à coefficients entiers de degré strictement inférieur à celui de A.

Passons à la démonstration proprement dite du critère d'Eisenstein. Raisonnons par l'absurde et supposons A non irréductible. D'après ce qui précède, il existe B et C dans $\mathbb{Z}[X]$, de degrés strictement inférieurs à n, tels que $A = BC$. Écrivons alors $B = b_k X^k + \cdots + b_1 X + b_0$ et $C = c_\ell X^\ell + \cdots + c_1 X + c_0$, avec $k = \deg B$ et $\ell = \deg C$. Comme dans la question précédente, on projette l'égalité $A = BC$ dans $\mathbb{Z}/p\mathbb{Z}[X]$. On obtient $\overline{a_n}X^n = \overline{B}\,\overline{C}$. Les polynômes $\overline{B}$ et $\overline{C}$ sont de degrés respectifs k et ℓ car $b_k c_\ell = a_n$ n'étant pas divisible par p, $\overline{b_k} \neq 0$ et $\overline{c_\ell} \neq 0$. Par

unicité de la décomposition en irréductibles dans $(\mathbb{Z}/p\mathbb{Z})[X]$, $\overline{B} = \overline{b_k}X^k$ et $\overline{C} = \overline{c_\ell}X^\ell$. On a alors $\overline{b_0} = \overline{c_0} = 0$ c'est-à-dire $p|b_0$ et $p|c_0$. On en déduit que p^2 divise $a_0 = b_0c_0$ ce qui contredit (iii). $\lhd$

Il résulte du critère d'Eisenstein que le polynôme $X^n - 2$ *est irréductible dans* $\mathbb{Q}[X]$ *pour tout entier* $n \geqslant 1$ *(prendre* $p = 2$*), ce qui prouve qu'il y a dans* $\mathbb{Q}[X]$ *des irréductibles de tout degré. Voici une autre application du critère d'Eisenstein.*

5.17. Irréductibilité de Φ_p dans $\mathbb{Q}[X]$

Soit $\omega = e^{\frac{2i\pi}{p}}$, où p est premier, et $\Phi_p = X^{p-1} + \cdots + X + 1$ (p-ième polynôme cyclotomique).

1. On admet que Φ_p est irréductible dans $\mathbb{Q}[X]$. Démontrer que l'ensemble $\mathcal{I}$ des polynômes annulateurs de ω dans $\mathbb{Q}[X]$ est $\Phi_p\mathbb{Q}[X]$.

2. Montrer que le polynôme $(X+1)^{p-1} + \cdots + (X+1) + 1$ est irréductible dans $\mathbb{Q}[X]$ (on pourra utiliser l'exercice précédent).

3. En déduire que Φ_p est irréductible dans $\mathbb{Q}[X]$.

4. Démontrer que $\mathbb{Q}\left[e^{\frac{2i\pi}{p}}\right] = \{Q(\omega), Q \in \mathbb{Q}[X]\}$ est un corps, appelé *corps cyclotomique*. Quelle est sa dimension comme espace vectoriel sur $\mathbb{Q}$?

(École normale supérieure)

▷ **Solution.**

1. Posons $\mathcal{I} = \{Q \in \mathbb{Q}[X], Q(\omega) = 0\}$. C'est un idéal de $\mathbb{Q}[X]$ en tant que noyau du morphisme d'algèbre $Q \in \mathbb{Q}[X] \longmapsto Q(\omega) \in \mathbb{C}$. Nous savons que tout idéal de $\mathbb{Q}[X]$ est principal. Comme I est non nul, il existe donc un unique polynôme unitaire Q tel que $\mathcal{I} = Q\mathbb{Q}[X]$. Or, nous savons que $\Phi_p(\omega) = 0$, puisque $(\omega - 1)\Phi_p(\omega) = \omega^p - 1 = 1 - 1 = 0$. On en déduit que Q divise Φ_p, puisque $\Phi_p \in \mathcal{I}$. Comme $Q \neq 1$ et comme Φ_p est supposé irréductible, on a nécessairement $Q = \Phi_p$. On conclut

$$\boxed{\mathcal{I} = \Phi_p\mathbb{Q}[X]}.$$

2. Posons $U = (X+1)^{p-1} + \cdots + (X+1) + 1 = \Phi_p(X+1)$, c'est-à-dire

$$\begin{aligned} U &= \frac{1 - (X+1)^p}{1 - (X+1)} = \frac{(X+1)^p - 1}{X} \\ &= X^{p-1} + C_p^{p-1}X^{p-2} + \cdots + C_p^2X + C_p^1 \in \mathbb{Z}[X]. \end{aligned}$$

Pour montrer que U est irréductible, nous allons utiliser le critère d'Eisenstein de l'exercice précédent avec le nombre premier p. Les hypothèses (i) et (iii) sont clairement vérifiées. Il s'agit de vérifier (ii), c'est-à-dire que les C_p^k sont divisibles par p pour $1 \leqslant k \leqslant p-1$. En effet,

si $1 \leqslant k \leqslant p-1$, on a $k!C_p^k = p(p-1)\dots(p-k+1)$. Donc p divise $k!C_p^k$. Comme $k < p$, p est premier avec $k!$ donc divise C_p^k.

On conclut à l'aide du critère d'Eisenstein que U est irréductible.

3. Supposons que Φ_p est composé et posons $\Phi_p = BC$ où B, C sont dans $\mathbb{Q}[X]$ avec $\deg B < \deg \Phi_p$ et $\deg C < \deg \Phi_p$. On en déduit que $U = \Phi_p(X+1) = B(X+1)C(X+1)$, avec $\deg B(X+1) = \deg B < \deg U$ et $\deg C(X+1) = \deg C < \deg U$. Il en résulte que U n'est pas irréductible, ce qui est faux.

Conclusion. $\Phi_p = X^{p-1} + \dots + X + 1$ est irréductible dans $\mathbb{Q}[X]$.

4. Le polynôme Φ_p est le polynôme unitaire de plus petit degré annulant ω. Donc la famille $(1, \omega, \dots, \omega^{p-2})$ est libre sur $\mathbb{Q}$ (toute combinaison linéaire non triviale nulle offrirait un polynôme non nul, de degré strictement inférieur à $p-1$, annulant ω). Si $Q \in \mathbb{Q}[X]$ et si on note R le reste de Q modulo Φ_p, il vient $Q(\omega) = R(\omega)$ puisque $\Phi_p(\omega) = 0$, ce qui montre que

$$\mathbb{Q}[\omega] = \mathrm{Vect}(1, \omega, \dots, \omega^{p-2}).$$

Le système $(1, \omega, \dots, \omega^{p-2})$ est donc une base du $\mathbb{Q}$-espace vectoriel $\mathbb{Q}[\omega]$, d'où l'on déduit

$$\boxed{\dim \mathbb{Q}\left[e^{\frac{2i\pi}{p}}\right] = p-1}.$$

Reste à démontrer que $\mathbb{Q}\left[e^{\frac{2i\pi}{p}}\right]$ est un corps. En premier lieu, $\mathbb{Q}[\omega]$ est l'image par le morphisme d'algèbre $P \longmapsto P(\omega)$ de l'algèbre $\mathbb{Q}[X]$; c'est donc une sous-algèbre de la $\mathbb{Q}$-algèbre $\mathbb{C}$ et en particulier un sous-anneau de $\mathbb{C}$. Soit x un élément non nul de $\mathbb{Q}[\omega]$. Il existe $R \in \mathbb{Q}[X]$ non nul de degré strictement inférieur à $p-1$ tel que $x = R(\omega)$. Comme Φ_p est irréductible, il est premier avec R. D'après le théorème de Bezout, il existe $(U, V) \in \mathbb{Q}[X]^2$ tel que $U\Phi_p + VR = 1$. En évaluant en ω, cela donne $U(\omega) \times 0 + V(\omega)x = 1$, soit $V(\omega)x = 1$ et $V(\omega) \in \mathbb{Q}[\omega]$ est l'inverse de x.

Conclusion. $\mathbb{Q}\left[e^{\frac{2i\pi}{p}}\right]$ est un corps de dimension $p-1$ sur $\mathbb{Q}$. ◁

Les racines de Φ_p sont les racines p-ièmes de l'unité différentes de 1. Plus généralement, pour $n \in \mathbb{N}^*$, on note Π_n l'ensemble des racines primitives n-ièmes de l'unité (rappelons qu'une racine n-ième est primitive si, et seulement si, elle engendre le groupe multiplicatif U_n, et que l'ard $\Pi_n = \varphi(n)$, où φ est l'indicateur d'Euler) et $\Phi_n = \prod_{\xi \in \Pi_n} (X - \xi)$. On peut alors montrer que Φ_n appartient à $\mathbb{Z}[X]$ (ce qui est fait dans l'exercice 4.18) et est irréductible dans $\mathbb{Q}[X]$, ce qui constitue le théorème de Dirichlet [3]. Il en résulte que si ξ est un élément de Π_n, alors $\mathbb{Q}[\xi]$ est un corps de dimension $\varphi(n)$ sur $\mathbb{Q}$.

3. PERRIN (D.), *Cours d'algèbre*. Ellipses, 1996.

5.18. Décomposition de $1 + X + X^2 + \cdots + X^{n-1}$

On suppose que $1 + X + X^2 + \cdots + X^{n-1} = P(X)Q(X)$ où P, Q sont deux polynômes réels unitaires à coefficients positifs ou nuls. Montrer que les coefficients de P et Q sont dans $\{0, 1\}$.

(École normale supérieure)

▷ **Solution.**

• Les racines complexes de $1 + X + X^2 + \cdots + X^{n-1}$ (et donc celles de P ou Q) sont les racines n-ièmes de l'unité différentes de 1. La seule racine réelle est -1, si n est pair. Les facteurs du second degré de P sont de la forme $(X - \alpha)(X - \overline{\alpha}) = X^2 - 2\mathrm{Re}(\alpha)X + 1$, où α est une racine n-ième de l'unité. Comme P est unitaire, il est produit de tels facteurs avec éventuellement $X + 1$ pour n pair, de sorte que le terme constant de P est égal à 1. On remarque d'autre part que si α est racine de P, $\dfrac{1}{\alpha} = \overline{\alpha}$ est aussi racine de P, avec la même multiplicité égale à 1 (un tel polynôme est appelé polynôme réciproque). Les polynômes P et $\widehat{P} = X^{\deg(P)}P\left(\dfrac{1}{X}\right)$ ont donc les mêmes racines, toutes simples : ils sont proportionnels. Étant unitaires, ils sont égaux. Si $P = \sum_{k=0}^{p} a_k X^k$, où p est le degré de P, alors $\widehat{P} = \sum_{k=0}^{p} a_{p-k} X^k$. On en déduit que l'on a, pour tout $k \in [\![0, p]\!]$, $a_{p-k} = a_k$. On obtient évidemment un résultat analogue pour le polynôme Q, dont on notera q le degré et $b_0, \ldots, b_q$ les coefficients.

• Supposons par exemple $p \leqslant q$. Considérons, pour $0 \leqslant k \leqslant p$, le coefficient d'ordre k de PQ ; il est égal à 1. On obtient $\sum_{i=0}^{k} a_{k-i} b_i = 1$. En particulier, pour $k = p$, on a, compte tenu de la symétrie des coefficients de P,

$$1 = \sum_{i=0}^{p} a_{p-i} b_i = \sum_{i=0}^{p} a_i b_i.$$

Sachant que $a_0 = b_0 = 1$ et que tous les coefficients sont positifs, on en déduit que, pour $1 \leqslant i \leqslant p$, on a $a_i b_i = 0$. On observe en particulier que $b_p = 0$ et donc que $q > p$.

On peut ensuite montrer simplement, par récurrence sur k entier entre 0 et p, que a_k et b_k sont dans $\{0, 1\}$. C'est vrai pour $k = 0$ (car $a_0 = b_0 = 1$). Si la propriété est établie jusqu'au rang $k - 1$, alors

$$a_k + b_k = a_k b_0 + a_0 b_k = 1 - \sum_{i=1}^{k-1} a_{k-i} b_i$$

est un entier plus petit que 1 et positif par hypothèse, donc égal à 0 ou 1. Nous avons le résultat voulu puisque nous savons que $a_k = 0$ ou $b_k = 0$.

Considérons ensuite, pour $p+1 \leqslant k \leqslant q$, le coefficient d'ordre k de PQ. On obtient $1 = \sum_{i=k-p}^{k} a_{k-i}b_i$, et donc

$$b_k = 1 - \sum_{i=k-p}^{k-1} a_{k-i}b_i.$$

Nous savons déjà que, pour $0 \leqslant k \leqslant p$, les coefficients a_k et b_k sont dans $\{0,1\}$. Une récurrence semblable à la précédente permet de démontrer que $b_k \in \{0,1\}$, pour $p+1 \leqslant k \leqslant q$. ◁

On peut se demander s'il existe de telles décompositions du polynôme $1 + X + X^2 + \cdots + X^{n-1}$. *L'exercice précédent démontre que si* n *est premier, ce polynôme est irréductible dans* $\mathbb{Q}[X]$. *Il faut donc avoir* $P = 1$ *ou* $Q = 1$. *Si* $n = ab$ *avec* a *et* b *entiers naturels,* $a \neq 1$ *et* $b \neq 1$, *on peut écrire*

$$\begin{aligned} 1 + X + X^2 + \cdots + X^{n-1} &= \frac{X^n - 1}{X - 1} = \frac{(X^a)^b - 1}{X - 1} \\ &= \frac{X^a - 1}{X - 1}\left(\sum_{i=0}^{b-1} X^{ai}\right) = \sum_{i=0}^{a-1} X^i \times \sum_{i=0}^{b-1} X^{ai}, \end{aligned}$$

ce qui est une décomposition non triviale du type considéré.

Nous allons maintenant nous intéresser aux fonctions polynomiales et à leurs propriétés. Si K *est un corps et* A *une partie infinie de* K, *il y un isomorphisme entre l'algèbre* K[X] *et l'algèbre des fonctions polynomiales de* A *dans* K. *Cela permet d'identifier ces deux algèbres, ce que font les exercices suivants.*

5.19. Polynômes complexes d'image réelle

Quels sont les polynômes complexes P dont l'image est incluse dans $\mathbb{R}$?

(École normale supérieure)

▷ **Solution.**

Si $P \in \mathbb{C}[X]$ est de degré supérieur ou égal à 1, la fonction polynôme associée à P est une surjection de $\mathbb{C}$ dans lui-même. En effet, pour tout complexe w, l'équation $P(z) = w$ admet une solution par le théorème de d'Alembert.

Conclusion. Si $P(\mathbb{C}) \subset \mathbb{R}$, c'est que P est un polynôme constant réel. $\lhd$

L'exercice suivant s'intéresse aux sommes de deux carrés dans l'anneau $\mathbb{R}[X]$. La situation est nettement plus simple que dans $\mathbb{Z}$.

5.20. Sommes de deux carrés dans $\mathbb{R}[X]$

Soit $P \in \mathbb{R}[X]$. Montrer que $P(x) \geqslant 0$ pour tout $x \in \mathbb{R}$ si, et seulement si, P s'écrit $A^2 + B^2$ avec A et B dans $\mathbb{R}[X]$.
(École polytechnique)

$\rhd$ **Solution.**

Si P est somme de deux carrés de polynômes réels, il est évident que $P(x) \geqslant 0$ pour tout x réel.

Réciproquement, supposons que $P(x) \geqslant 0$ pour tout x réel. Le cas P constant étant trivial, supposons $\deg P \geqslant 1$. En considérant la limite de P en $+\infty$ qui ne peut être que $+\infty$, il apparait que le coefficient dominant de P est positif. Si a est une racine réelle de P d'ordre n, P s'écrit $(X-a)^n Q$ avec $Q(a) \neq 0$. En particulier, Q étant continu sur $\mathbb{R}$, il garde un signe constant au voisinage de a. Comme P ne change pas de signe en a, n est nécessairement pair. Par conséquent, P se décompose sur $\mathbb{C}$ de la manière suivante

$$P = \lambda \prod_{i=1}^{p} (X - a_i)^{\alpha_i} \prod_{j=1}^{q} (X - \lambda_j)^{n_j} (X - \overline{\lambda_j})^{n_j},$$

où $a_1, \ldots, a_p$ sont les racines réelles de P, $\alpha_1, \ldots, \alpha_p$ sont des entiers pairs et $\lambda \in \mathbb{R}_+^*$ (dans le second produit on a associé chaque racine non réelle avec sa conjuguée qui a le même ordre de multiplicité puisque P est réel). Si on pose

$$C = \sqrt{\lambda} \prod_{i=1}^{p} (X - a_i)^{\frac{\alpha_i}{2}} \prod_{j=1}^{q} (X - \lambda_j)^{n_j},$$

on constate que $P = C\overline{C}$. En écrivant $C = A + iB$, avec A et B dans $\mathbb{R}[X]$, on obtient $P = (A + iB)(A - iB) = A^2 + B^2$. $\lhd$

L'exercice suivant caractérise les polynômes réels prenant des valeurs positives sur $[-1, 1]$.

5.21. Polynômes positifs sur $[-1, 1]$

Soit $P \in \mathbb{R}[X]$ tel que $P(x) \geqslant 0$, pour tout x dans $[-1, 1]$.

1. On suppose que P est de degré inférieur ou égal à 2. Montrer qu'on peut écrire $P = \alpha(X - a)^2 + \beta(1 - X^2)$, avec $\alpha \geqslant 0$, $\beta \geqslant 0$ et $a \in [-1, 1]$.

2. On revient au cas général. Montrer l'existence de A et B dans $\mathbb{R}[X]$ tel que $P = A^2 + (1 - X^2)B^2$.

(École polytechnique)

▷ **Solution.**

1. Supposons d'abord que P s'annule sur $[-1, 1]$, en a.

Si $a \in]-1, 1[$, P ne change pas de signe en a, donc a ne peut pas être racine simple de P. Si P est de degré 2, il s'écrit donc $\alpha(X - a)^2$, avec $\alpha > 0$; sinon, c'est le polynôme nul, ce qui revient à prendre $\alpha = 0$. Si $a \in]-1, 1[$, P a donc la forme voulue, avec $\beta = 0$.

Examinons maintenant le cas où P s'annule seulement en $a = \pm 1$. Quitte à considérer le polynôme $P(-X)$, on peut supposer que $a = 1$. Il existe alors $(b, c) \in \mathbb{R}^2$ tel que $P = (1 - X)(bX + c)$. De l'hypothèse, on déduit que $bx + c \geqslant 0$ pour tout $x \in [-1, 1]$, ce qui équivaut à $b + c \geqslant 0$ et $-b + c \geqslant 0$. Alors P est de la forme voulue, avec $a = 1$, si et seulement si $P = (1 - X)(\alpha(1 - X) + \beta(1 + X))$, c'est-à-dire $b = -\alpha + \beta$ et $c = \alpha + \beta$, ce qui s'obtient en posant $\alpha = \frac{1}{2}(c - b)$ et $\beta = \frac{1}{2}(c + b)$. Les réels α et β sont bien positifs d'après les conditions trouvées sur b et c.

Dans le cas où P est strictement positif sur $[-1, 1]$, nous allons montrer qu'en lui soustrayant un terme de la forme $\beta(1 - X^2)$, on peut se ramener au cas précédent. Soit E l'ensemble des $k > 0$ tels que $P - k(1 - X^2)$ soit positif ou nul sur $[-1, 1]$. Puisque P possède sur $[-1, 1]$ un minimum strictement positif m, E n'est pas vide : si $0 < k \leqslant m$, k appartient à E. Cet ensemble est majoré, par exemple par $P(0)$. Soit β sa borne supérieure. Le polynôme $Q = P - \beta(1 - X^2)$ s'annule sur $[-1, 1]$, sinon le même raisonnement appliqué à Q permettrait d'exhiber $k > 0$ tel que $Q - k(1 - X^2)$ soit positif ou nul sur $[-1, 1]$, ce qui contredirait la définition de β. Si a est une racine de Q sur $[-1, 1]$, a est différent de ± 1, sinon P s'annulerait en ± 1. Le polynôme Q est positif sur $[-1, 1]$ et s'annule en $a \in]-1, 1[$. Nous avons montré qu'alors il existe $\alpha \geqslant 0$ tel que $Q = \alpha(X - a)^2$, ce qui conduit à $P = \alpha(X - a)^2 + \beta(1 - X^2)$, avec $\alpha \geqslant 0$ et $\beta \geqslant 0$, ce qui est le résultat demandé.

2. Montrons pour commencer que P peut s'écrire comme produit de polynômes de degré inférieur ou égal à 2, positifs sur $[-1, 1]$. Notons $a_1, \ldots, a_k$ les racines de P appartenant à $]-1, 1[$. Puisque P ne change pas de signe en a_i, la multiplicité de a_i est paire ; on la note $2n_i$. Le

polynôme P se décompose donc en

$$P = \prod_{i=1}^{k} (X - a_i)^{2n_i} \prod_{j=1}^{\ell} P_j.$$

où $P_1, P_2, \ldots, P_\ell$ sont des polynômes irréductibles, donc de degré inférieur ou égal à 2, ne s'annulant pas sur $]-1,1[$. Comme P est positif sur $[-1,1]$, on en déduit que $\prod_{j=1}^{\ell} P_j$ est strictement positif sur $]-1,1[$. Quitte à multiplier certains P_j par -1, on peut supposer que $P_1, \ldots, P_\ell$ sont tous strictement positifs sur $]-1,1[$ et donc positifs sur $[-1,1]$. On a montré que P s'écrit comme produit de polynômes de la forme $(X-a_i)^2$ ou P_j, tous positifs sur $[-1,1]$ et de degré inférieur ou égal à 2.

D'après la question 4, chacun de ces polynômes peut s'écrire

$$\alpha(X-a)^2 + \beta(1-X^2) = (\sqrt{\alpha}(X-a))^2 + (1-X^2)(\sqrt{\beta})^2,$$

ce qui est de la forme $A^2 + (1-X^2)B^2$. Pour conclure, il reste à montrer qu'un produit de polynômes de cette forme est encore un polynôme ayant la même forme. Le montrer pour deux polynômes est d'ailleurs suffisant. Cela résulte alors de l'identité de Lagrange, valable dans tout anneau :

$$(a^2+b^2)(c^2+d^2) = (ac-bd)^2 + (ad+bc)^2.$$

Formellement, on est tenté d'écrire, pour des polynômes A, B, C, D,

$$\begin{aligned} Q &= \left(A^2 + (1-X^2)B^2\right)\left(C^2 + (1-X^2)D^2\right) \\ &= \left(A^2 + (\sqrt{1-X^2}B)^2\right)\left(C^2 + (\sqrt{1-X^2}D)^2\right) \\ &= \left(AC - (1-X^2)BD\right)^2 + \left(\sqrt{1-X^2}(AD+BC)\right)^2 \\ &= \left(AC - (1-X^2)BD\right)^2 + (1-X^2)(AD+BC)^2. \end{aligned}$$

L'égalité entre Q et le dernier terme peut se vérifier directement. Elle démontre le résultat voulu. Elle tient moins du miracle qu'il n'en paraît puisque ce calcul se justifierait en se plaçant sur une «extension» de $\mathbb{R}[X]$ dans laquelle $1-X^2$ aurait une racine. ◁

Il est évident que réciproquement, tout polynôme à coefficients réels qui s'écrit $A^2 + (1-X^2)B^2$ est positif sur $[-1,1]$.

Voici encore un exercice sur les polynômes réels positifs. On n'oubliera pas qu'un polynôme réel est en particulier une fonction de classe C^∞ à laquelle on peut appliquer tous les résultats classiques de l'analyse réelle.

5.22. Polynôme positif

Soit $P \in \mathbb{R}[X]$ tel que pour tout $x \in \mathbb{R}$, $P(x) \geqslant 0$. Soit n le degré de P et $Q = P + P' + \cdots + P^{(n)}$. Montrer que pour tout $x \in \mathbb{R}$, on a $Q(x) \geqslant 0$.

(École normale supérieure)

▷ **Solution.**

Le polynôme P étant positif, n est pair et le coefficient dominant de P est positif. Le degré de Q vaut alors n et son coefficient dominant est celui de P. Il en résulte que Q tend vers $+\infty$ en $+\infty$ et en $-\infty$ de sorte que Q est minoré sur $\mathbb{R}$. Si m est sa borne inférieure, il existe $A > 0$ tel que $|Q(x)| \geqslant m + 1$ si $|x| \geqslant A$, ce qui montre que m est aussi la borne inférieure de Q sur le compact $[-A, A]$. La fonction Q étant continue, cette borne inférieure est atteinte en un point x_0. On a alors $Q'(x_0) = 0$. On remarque que $Q - Q' = P$. On a donc $Q(x_0) = P(x_0) \geqslant 0$. D'où le résultat. ◁

L'exercice suivant étudie les racines modulo p premier d'un polynôme à coefficients entiers.

5.23. Diviseurs d'un polynôme de $\mathbb{Z}[X]$

Soit $P \in \mathbb{Z}[X]$ et p premier. On dit que p est un *diviseur* de P s'il existe un entier $n \in \mathbb{N}$ tel que $P(n) \equiv 0 \pmod{p}$ et $P(n) \neq 0$. Montrer que tout polynôme non constant a une infinité de diviseurs.

(École normale supérieure)

▷ **Solution.**

On peut supposer que le coefficient constant a_0 de P n'est pas nul. Sinon, on écrit $P = X^k Q$, où k est la valuation de P et il suffit de démontrer la propriété pour Q.

On note D l'ensemble des diviseurs de P. L'ensemble des entiers n tels que $|P(n)| \leqslant 1$ est fini car P n'est pas constant. Si n est un entier naturel tel que $|P(n)| \geqslant 2$, alors $P(n)$ possède un diviseur premier p qui appartient à D : D n'est pas vide.

Montrons que le cardinal de D n'est pas fini en raisonnant par l'absurde. Si $D = \{p_1, p_2, \ldots, p_\ell\}$ ($\ell \in \mathbb{N}^*$) considérons, pour $m \in \mathbb{N}^*$, le produit $n = m \cdot |a_0| \cdot p_1 \ldots p_\ell$. Il existe $N_m \in \mathbb{Z}$ tel que $P(n) = a_0 N_m$ et $N_m \equiv 1 \pmod{p_1 \ldots p_\ell}$. L'entier $|N_m|$ tend vers $+\infty$ avec m, donc pour m assez grand, on a $|N_m| \geqslant 2$. Alors N_m possède alors un diviseur premier p qui appartient à D. D'où la contradiction cherchée, car on devrait avoir $N_m \equiv 1 \pmod{p}$. ◁

On s'intéresse maintenant aux polynômes à coefficients réels qui laissent $\mathbb{Z}$ stable.

5.24. Polynômes de Hilbert

On pose $H_0 = 1$ et, pour $n \in \mathbb{N}^*$,

$$H_n = \frac{X(X-1)\dots(X-n+1)}{n!}.$$

1. Montrer que, pour tout $n \in \mathbb{N}$, on a $H_n(\mathbb{Z}) \subset \mathbb{Z}$. En déduire que le produit de n entiers consécutifs dans $\mathbb{Z}$ est divisible par $n!$.

2. Soit $P \in \mathbb{R}[X]$, avec $\deg P \leqslant n$. Montrer qu'il y a équivalence entre :

(i) $P(\mathbb{Z}) \subset \mathbb{Z}$;

(ii) $P(k) \in \mathbb{Z}$, pour $k = 0, 1, \dots, n$;

(iii) il existe $(\lambda_0, \dots, \lambda_n) \in \mathbb{Z}^{n+1}$ tel que $P = \sum_{k=0}^{n} \lambda_k H_k$.

(École polytechnique)

▷ **Solution.**

1. Soit $n \in \mathbb{N}$. On a

$$H_n(k) = \begin{cases} 0 & \text{si} \quad 0 \leqslant k \leqslant n-1 \\ C_k^n & \text{si} \quad k \geqslant n \\ (-1)^n C_{n-k-1}^n & \text{si} \quad k < 0. \end{cases}$$

Pour tout k entier, le produit des entiers $k, k+1, \dots, k+n-1$ vaut donc $n!\, H_n(k-n+1)$, qui est un multiple de $n!$.

Il convient de remarquer que $P(\mathbb{Z}) \subset \mathbb{Z}$ *n'implique pas* $P \in \mathbb{Z}[X]$, *comme le prouve l'exemple des polynômes* H_n.

2. Bien entendu *(i)* implique *(ii)*. Si P est de la forme $\sum_{k=0}^{n} \lambda_k H_k$, avec $(\lambda_0, \dots, \lambda_n) \in \mathbb{Z}^{n+1}$, alors la question 1 montre que $P(\mathbb{Z}) \subset \mathbb{Z}$. Donc *(iii)* implique *(i)*.

Il reste à prouver que *(ii)* implique *(iii)*. Chaque polynôme H_k étant de degré k, la famille $(H_0, \dots, H_n)$ est une base de $\mathbb{R}_n[X]$. Il existe donc $(\lambda_0, \dots, \lambda_n) \in \mathbb{R}^{n+1}$ tel que $P = \sum_{k=0}^{n} \lambda_k H_k$. Montrons, par récurrence sur $k \in [\![0, n]\!]$ que $\lambda_k \in \mathbb{Z}$.

- On a $P(0) = \lambda_0 \in \mathbb{Z}$.
- Si, pour $k \in [\![1, n]\!]$, $(\lambda_0, \dots, \lambda_{k-1}) \in \mathbb{Z}^k$, on a alors

$$\mathrm{P}(k) = \sum_{\ell=0}^{k} \lambda_\ell \mathrm{H}_\ell(k) = \sum_{\ell=0}^{k} \lambda_\ell \mathrm{C}_k^\ell \text{ et donc } \lambda_k = \mathrm{P}(k) - \sum_{\ell=0}^{k-1} \lambda_\ell \mathrm{C}_k^\ell \in \mathbb{Z},$$

car $\mathrm{P}(k), \lambda_0, \ldots, \lambda_{k-1}$ sont dans $\mathbb{Z}$, ce qui achève la démonstration. ◁

Si K *est un corps,* $x_1, \ldots, x_n$ *et* $y_1, \ldots, y_n$ *des éléments de* K*, les* x_i *étant deux à deux distincts, il existe un unique polynôme* $\mathrm{P} \in \mathrm{K}_{n-1}[\mathrm{X}]$ *tel que l'on ait, pour* $1 \leqslant i \leqslant n$, $\mathrm{P}(x_i) = y_i$*. C'est le polynôme d'interpolation de Lagrange relatif à* $(x_i, y_i)_{1 \leqslant i \leqslant n}$*. Il s'écrit*

$$\mathrm{P} = \sum_{i=1}^{n} y_i \mathrm{L}_i, \quad \text{où} \quad \mathrm{L}_i = \frac{\prod_{j \neq i} (\mathrm{X} - x_j)}{\prod_{j \neq i} (x_i - x_j)}.$$

On peut remarquer que si $\Phi(\mathrm{X}) = \prod_{i=1}^{n} (\mathrm{X} - x_i)$*, alors* $\mathrm{L}_i(\mathrm{X}) = \dfrac{\Phi(\mathrm{X})}{(\mathrm{X} - x_i)\Phi'(x_i)}$.

Nous appellerons les polynômes L_i *les polynômes interpolateurs de base relatifs au* n*-uplet* $(x_1, \ldots, x_n)$*.*

Il en résulte en particulier qu'un polynôme $\mathrm{P} \in \mathrm{K}_{n-1}[\mathrm{X}]$ *est uniquement déterminé par ses valeurs en* n *points distincts* $x_1, \ldots, x_n$ *de* K*. Les valeurs* $\mathrm{P}(x_i)$ *sont alors les coordonnées de* P *dans la base de* $\mathrm{K}_{n-1}[\mathrm{X}]$ *formée des polynômes* $\mathrm{L}_1, \ldots, \mathrm{L}_n$*. Cette idée est à la base de l'exercice suivant, qui regroupe en fait trois petits exercices sur le même thème.*

5.25. Interpolation de Lagrange

1. Soit $\mathrm{P} \in \mathbb{C}[\mathrm{X}]$ de degré $n \geqslant 1$. On suppose qu'il existe $k \in \mathbb{Z}$ tel que $\mathrm{P}(k)$, $\mathrm{P}(k+1)$, ... , $\mathrm{P}(k+n)$ soient dans $\mathbb{Z}$. Montrer que pour tout $x \in \mathbb{Z}$, $\mathrm{P}(x) \in \mathbb{Z}$.

2. Soit $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$ de degré $n \geqslant 1$. On note N le pgcd de $\mathrm{P}(0)$, $\mathrm{P}(1)$, ..., $\mathrm{P}(n)$. Montrer que N divise $\mathrm{P}(x)$ pour tout $x \in \mathbb{Z}$.

3. Soit $\mathrm{P} \in \mathbb{R}[\mathrm{X}]$ de degré n tel que $\mathrm{P}(0)$, $\mathrm{P}(1)$, $\mathrm{P}(4)$, ..., $\mathrm{P}(n^2)$ soient dans $\mathbb{Z}$. Montrer que, pour tout $a \in \mathbb{Z}$, $\mathrm{P}(a^2) \in \mathbb{Z}$.

(École normale supérieure)

▷ **Solution.**

1. Quitte à considérer le polynôme Q tel que $\mathrm{Q}(\mathrm{X}) = \mathrm{P}(\mathrm{X} + k)$, on peut supposer que $k = 0$. On note $\mathrm{L}_0, \mathrm{L}_1, \ldots, \mathrm{L}_n$ les polynômes interpolateurs de base pour le $(n+1)$-uplet $(0, 1, \ldots, n)$. On a alors, pour tout entier x, $\mathrm{P}(x) = \sum_{i=0}^{n} \mathrm{P}(i)\mathrm{L}_i(x)$. Pour montrer le résultat voulu, il suffit de prouver que $\mathrm{L}_i(x) \in \mathbb{Z}$ pour tout entier x et tout $i \in [\![0, n]\!]$.

On a, pour $i \in [\![0, n]\!]$ et $x \in \mathbb{Z}$,

$$\mathrm{L}_i(x) = \frac{\prod\limits_{j\in[\![0,n]\!]\setminus\{i\}}(x-j)}{\prod\limits_{j\in[\![0,n]\!]\setminus\{i\}}(i-j)} = \frac{(-1)^{n-i}\prod\limits_{j=0}^{i-1}(x-j)}{i!}\times\frac{\prod\limits_{j=i+1}^{n}(x-j)}{(n-i)!}.$$

Le produit de k entiers consécutifs est divisible par $k!$ (c'est ce que montre la première question de l'exercice 5.24) donc $\mathrm{L}_i(x)$ est entier.

2. On a encore $\mathrm{P}(x) = \sum\limits_{i=0}^{n}\mathrm{P}(i)\mathrm{L}_i(x)$ pour tout $x\in\mathbb{Z}$. Comme N divise les $\mathrm{P}(i)$ et que les $\mathrm{L}_i(x)$ sont entiers d'après 1, il est clair que N divise $\mathrm{P}(x)$.

3. Soit $\mathrm{Q}(\mathrm{X}) = \mathrm{P}(\mathrm{X}^2)$. On a $\deg(\mathrm{Q}) = 2n$ et $\mathrm{Q}(-n)$, $\mathrm{Q}(-n+1)$, ..., $\mathrm{Q}(-1)$, $\mathrm{Q}(0)$, $\mathrm{Q}(1)$,...,$\mathrm{Q}(n)$ sont dans $\mathbb{Z}$. D'après la question 1, Q prend des valeurs entières sur $\mathbb{Z}$: c'est le résultat voulu. ◁

Observons que la première question fournit une autre solution de la question 2 de l'exercice 5.24 pour (ii) $\Longrightarrow$ (i). L'exercice qui suit utilise aussi l'interpolation de Lagrange.

5.26. Polynômes complexes envoyant surjectivement $\mathbb{Q}$ sur $\mathbb{Q}$

Soit $\mathrm{P}\in\mathbb{C}[\mathrm{X}]$. Trouver une condition nécessaire et suffisante pour que P induise une surjection de $\mathbb{Q}$ sur $\mathbb{Q}$.

(École normale supérieure)

▷ **Solution.**

Un polynôme non nul $\mathrm{P}\in\mathbb{C}[\mathrm{X}]$ vérifiant $\mathrm{P}(\mathbb{Q})\subset\mathbb{Q}$ est à coefficients rationnels. En effet, si n est son degré, on a $\mathrm{P} = \mathrm{P}(0)\mathrm{L}_0 + \cdots + \mathrm{P}(n)\mathrm{L}_n$, où les L_i sont les polynômes interpolateurs de base pour le $(n+1)$-uplet $(0,1,\ldots,n)$ (et les L_i sont tous dans $\mathbb{Q}[\mathrm{X}]$). On suppose donc $\mathrm{P}\in\mathbb{Q}[\mathrm{X}]$. Il est évident que tous les polynômes $\mathrm{P}\in\mathbb{Q}[\mathrm{X}]$ de degré 1 conviennent. On va en fait montrer que ce sont les seuls. Il est clair que les polynômes constants ne conviennent pas. Montrons qu'il en est de même des polynômes de degré $\geqslant 2$.

Soit $\mathrm{P}\in\mathbb{Q}[\mathrm{X}]$ de degré $n\geqslant 2$. Quitte à multiplier P par un entier non nul (ce qui ne modifie pas la surjectivité), on peut supposer que $\mathrm{P}\in\mathbb{Z}[\mathrm{X}]$. Écrivons $\mathrm{P} = \sum\limits_{i=0}^{n}a_i\mathrm{X}^i$, avec $(a_0,\ldots,a_n)\in\mathbb{Z}^{n+1}$. Si $\frac{p}{q}$ est un élément non nul de $\mathbb{Q}$, écrit sous forme irréductible, alors $q^n\mathrm{P}\left(\frac{p}{q}\right) = \sum\limits_{i=0}^{n}a_ip^iq^{n-i}$. Considérons un entier premier m. Si $\mathrm{P}\left(\frac{p}{q}\right) = \frac{1}{m}$, alors m divise q^n et donc q, puisque m est premier. Alors m^n divise q^n et en particulier comme $n\geqslant 2$, m divise $\frac{q^n}{m} = \sum\limits_{i=0}^{n}a_ip^iq^{n-i}$. On en déduit que m divise

$a_n p^n$ et donc a_n, car p et q sont premiers entre eux. Un nombre rationnel de la forme $\frac{1}{m}$, où m est un nombre premier ne divisant pas a_n, n'a donc pas d'antécédent par P. Donc $\mathrm{P}(\mathbb{Q})$ n'est pas égal à $\mathbb{Q}$.

Conclusion. Les seuls polynômes complexes qui envoient $\mathbb{Q}$ sur $\mathbb{Q}$ surjectivement sont les polynômes à coefficients rationnels de degré 1. ◁

Soit $n \in \mathbb{N}^*$ et U_n l'ensemble des racines n-ièmes de l'unité. Pour k dans $\mathbb{Z}$, on a $\sum_{w \in \mathrm{U}_n} w^k = 0$ si $k \notin n\mathbb{Z}$ et $\sum_{w \in \mathrm{U}_n} w^k = n$ si $k \in n\mathbb{Z}$. Il en résulte que si $\mathrm{P} \in \mathbb{C}_{n-1}[\mathrm{X}]$, $\frac{1}{n} \sum_{w \in \mathrm{U}_n} \mathrm{P}(w) = \mathrm{P}(0)$. A l'aide d'une similitude, on en déduit facilement que, pour tout polygone régulier de $\mathbb{C}$ et tout $\mathrm{P} \in \mathbb{C}_{n-1}[\mathrm{X}]$, la valeur de P au centre du polygone (qui est l'isobarycentre des sommets) est la moyenne des valeurs de P aux sommets du polygone. L'exercice qui suit montre qu'un polygone $(z_1, \ldots, z_n)$ qui vérifie cette propriété est nécessairement régulier.

5.27. Caractérisation des polygones réguliers

On considère $n+1$ nombres complexes $z_0, z_1, \ldots, z_n$ distincts tels que pour tout polynôme $\mathrm{P} \in \mathbb{C}_{n-1}[\mathrm{X}]$, on ait $\mathrm{P}(z_0) = \frac{1}{n} \sum_{k=1}^{n} \mathrm{P}(z_k)$. Montrer que $z_1, \ldots, z_n$ sont les sommets d'un polygone régulier de centre z_0. **(École normale supérieure)**

▷ **Solution.**

On observe que z_0 est l'isobarycentre de $z_1, \ldots, z_n$ (pour le montrer, il suffit de prendre $\mathrm{P} = \mathrm{X}$).

Nous allons exploiter l'hypothèse sur une base de $\mathbb{C}_{n-1}[\mathrm{X}]$. Nous proposons deux solutions de l'exercice.

• Une première méthode consiste à choisir une base de $\mathbb{C}_{n-1}[\mathrm{X}]$ où la somme des $\mathrm{P}(z_k)$ se calcule facilement : la base $(\mathrm{L}_1, \ldots, \mathrm{L}_n)$ des polynômes interpolateurs du n-uplet $(z_1, \ldots, z_n)$ définis, pour $i = 1, \ldots, n$, par

$$\mathrm{L}_i = \frac{\prod_{j \neq i} (\mathrm{X} - z_j)}{\prod_{j \neq i} (z_i - z_j)}.$$

L'hypothèse est que, pour tout $i \in [\![1, n]\!]$, $\mathrm{L}_i(z_0) = \frac{1}{n}$. Si on introduit le polynôme $\mathrm{Q}(\mathrm{X}) = (\mathrm{X} - z_1)(\mathrm{X} - z_2) \ldots (\mathrm{X} - z_n)$, cette égalité s'écrit

$$\frac{\mathrm{Q}(z_0)}{z_0 - z_i} = \frac{1}{n} \mathrm{Q}'(z_i).$$

Cela implique que le polynôme $Q'(X)(X-z_0)-n[Q(X)-Q(z_0)]$ s'annule en chaque z_i, $i=1,\dots,n$. Ce polynôme étant de degré inférieur ou égal à $n-1$, il est nul. On montre alors que $Q(X)-Q(z_0)=(X-z_0)^n$. En effet, on remarque que

$$\left(\frac{Q(X)-Q(z_0)}{(X-z_0)^n}\right)'=\frac{Q'(X)(X-z_0)-n(Q(X)-Q(z_0))}{(X-z_0)^{n+1}}=0.$$

La fraction $\dfrac{Q(z)-Q(z_0)}{(z-z_0)^n}$ est donc constante et, puisque le coefficient dominant de Q est 1, elle vaut 1. On obtient, pour tout $i\in[\![1,n]\!]$,

$$(z_i-z_0)^n=Q(z_i)-Q(z_0)=-Q(z_0).$$

Conclusion. Les complexes z_1-z_0, …, z_n-z_0 sont les racines n-ièmes de $-Q(z_0)$ et $z_1,\dots,z_n$ forment donc un polygone régulier centré en z_0.

• Une autre idée consiste à exploiter l'hypothèse sur une base de $\mathbb{C}_{n-1}[X]$ où $P(z_0)$ se calcule facilement : la base $((X-z_0)^p)_{0\leqslant p\leqslant n-1}$. On obtient pour $p\in[\![1,n-1]\!]$, $0=\dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(z_k-z_0)^p$. Les sommes de Newton $S_1,S_2,\dots,S_{n-1}$ relatives à $z_1-z_0,z_2-z_0,\dots,z_n-z_0$, définies par $S_p=\displaystyle\sum_{i=1}^{n}(z_k-z_0)^p$ sont donc nulles. D'après les formules de Newton, qui sont démontrées dans l'exercice 5.31, on en déduit que, si $\sigma_1,\dots,\sigma_n$ sont les fonctions symétriques élémentaires de $z_1-z_0,\dots,z_n-z_0$, on a, pour $1\leqslant p\leqslant n-1$,

$$(-1)^{p+1}p\sigma_p=S_{p-1}+\cdots+\sigma_{p-1}S_1=0.$$

On a donc $\sigma_1=\cdots=\sigma_{n-1}=0$ et $z_1-z_0,\dots,z_n-z_0$ sont les racines du polynôme $X^n+(-1)^n\sigma_n$, c'est-à-dire les racines n-ièmes de $(-1)^{n+1}\sigma_n$. On conclut comme dans la solution précédente. ◁

Dans l'exercice suivant, il est encore question de fonctions polynomiales : il s'agit des fonctions induites par un polynôme à coefficients dans $\mathbb{Z}$ ou $\mathbb{Q}$ sur $\mathbb{Z}/p\mathbb{Z}$ ou $\mathbb{Z}/p^i\mathbb{Z}$.

5.28. Polynômes entiers et fonctions polynomiales induites sur $\mathbb{Z}/p^i\mathbb{Z}$

Soit p un nombre premier.

1. On note $\overline{u}$ la classe de l'entier u dans $\mathbb{Z}/p\mathbb{Z}$. Pour $P\in\mathbb{Z}[X]$, on note $\Phi(P)\in\mathcal{F}(\mathbb{Z}/p\mathbb{Z},\mathbb{Z}/p\mathbb{Z})$ l'application qui à toute classe $\overline{u}$ associe la classe $\overline{P(u)}$. Étudier la surjectivité de Φ et son noyau.

2. Pour $n \in \mathbb{N}^*$, on note k_n le produit des entiers non multiples de p qui sont compris entre 1 et n. On pose $\mathrm{H}_0 = 1$ et, pour $n \geqslant 1$, $\mathrm{H}_n = \dfrac{1}{k_n}\mathrm{X}(\mathrm{X}-1)\ldots(\mathrm{X}-n+1)$. Les polynômes H_n engendrent un sous-groupe additif de $\mathbb{Q}[\mathrm{X}]$ noté S. Soit $i \in \mathbb{N}^*$. Maintenant on note $\overline{u}$ la classe de l'entier u dans $\mathbb{Z}/p^i\mathbb{Z}$.

a. Soit $\mathrm{Q} \in \mathrm{S}$. Montrer que $\mathrm{Q}(\mathbb{Z}) \subset \mathbb{Z}$ et que $u \equiv u' \,[p^i]$ implique $\mathrm{Q}(u) \equiv \mathrm{Q}(u') \,[p^i]$.

b. Pour $\mathrm{Q} \in \mathrm{S}$ on note $\Psi(\mathrm{Q}) \in \mathcal{F}(\mathbb{Z}/p^i\mathbb{Z}, \mathbb{Z}/p^i\mathbb{Z})$ l'application qui à toute classe $\overline{u}$ associe la classe $\overline{\mathrm{Q}(u)}$. Étudier la surjectivité de Ψ et son noyau.

(**École normale supérieure**)

▷ **Solution.**

1. L'application qui à $\mathrm{P} = \sum_{n\in\mathbb{N}} a_n\mathrm{X}^n \in \mathbb{Z}[\mathrm{X}]$ associe $\overline{\mathrm{P}} = \sum_{n\in\mathbb{N}} \overline{a_n}\mathrm{X}^n$ élément de $(\mathbb{Z}/p\mathbb{Z})[\mathrm{X}]$ est un morphisme d'anneaux. Comme l'application qui, à un polynôme de $(\mathbb{Z}/p\mathbb{Z})[\mathrm{X}]$, associe sa fonction polynomiale est aussi un morphisme d'anneaux, il en va de même de la composée Φ.

• Montrons que Φ est surjective. Si $f \in \mathcal{F}(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/p\mathbb{Z})$, nous savons qu'il existe un unique polynôme $\mathrm{A} \in (\mathbb{Z}/p\mathbb{Z})[\mathrm{X}]$ de degré inférieur ou égal à $p-1$ tel que pour tout $i \in [\![0, p-1]\!]$, $\mathrm{A}(\overline{i}) = f(\overline{i})$ (polynôme d'interpolation de Lagrange).

Soit $(b_1, \ldots, b_{p-1}) \in \mathbb{Z}^p$ tel que $\mathrm{A} = \sum_{j=0}^{p-1} \overline{b_j}\mathrm{X}^j$,. On pose $\mathrm{P} = \sum_{j=1}^{p-1} b_j\mathrm{X}^j$. Par construction, on a, pour tout $i \in [\![0, p-1]\!]$, $\overline{\mathrm{P}(i)} = \mathrm{A}(\overline{i}) = f(\overline{i})$. Ceci montre que $\Phi(\mathrm{P}) = f$. L'application Φ est surjective.

Ce qui précède montre que toute application de $\mathbb{Z}/p\mathbb{Z}$ dans lui-même est polynomiale.

• D'après le petit théorème de Fermat, on a $x^p - x = 0$ pour tout $x \in \mathbb{Z}/p\mathbb{Z}$, autrement dit $\Phi(\mathrm{X}^p - \mathrm{X}) = 0$. Le polynôme $\mathrm{X}^p - \mathrm{X}$ appartient à $\operatorname{Ker}\Phi$. Si $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$, on effectue la division euclidienne de P par $\mathrm{X}^p - \mathrm{X}$. On obtient $\mathrm{P} = (\mathrm{X}^p - \mathrm{X})\mathrm{Q} + \mathrm{R}$, avec $\mathrm{Q} \in \mathbb{Z}[\mathrm{X}]$ et $\mathrm{R} \in \mathbb{Z}_{p-1}[\mathrm{X}]$. On a alors $\Phi(\mathrm{P}) = \Phi(\mathrm{Q})\Phi(\mathrm{X}^p - \mathrm{X}) + \Phi(\mathrm{R}) = \Phi(\mathrm{R})$. Ceci montre que P appartient à $\operatorname{Ker}\Phi$ si et seulement si R est dans $\operatorname{Ker}\Phi$.

On écrit $\mathrm{R} = \sum_{j=0}^{p-1} b_j\mathrm{X}^j$, avec $(b_0, \ldots, b_{p-1}) \in \mathbb{Z}^p$ et on considère son projeté $\overline{\mathrm{R}} = \sum_{j=0}^{p-1} \overline{b_j}\mathrm{X}^j$ dans $\mathbb{Z}/p\mathbb{Z}[\mathrm{X}]$. Si R est dans $\operatorname{Ker}\Phi$, tout élément de $\mathbb{Z}/p\mathbb{Z}$ est racine du polynôme $\overline{\mathrm{R}}$. Celui-ci étant de degré inférieur ou égal à $p-1$, c'est le polynôme nul. Ainsi, pour tout $j \in [\![0, p-1]\!]$, $\overline{b_j} = \overline{0}$, c'est-à-dire que p divise b_j. On en déduit qu'il existe $\mathrm{T} \in \mathbb{Z}_{p-1}[\mathrm{X}]$ tel que

$R = pT$ et $P = (X^p - X)Q + pT$. Il est clair, réciproquement, que tout polynôme ayant cette forme appartient à $\operatorname{Ker}\Phi$, et plus généralement tous les polynômes de la forme $(X^p - X)Q + pT$ avec $(Q, T) \in \mathbb{Z}[X]^2$ (qui se ramènent à la forme précédente en divisant T par $X^p - X$).

Conclusion. Le noyau de Φ est l'idéal de $\mathbb{Z}[X]$ engendré par les polynômes $X^p - X$ et p.

2.a. Les éléments de S sont les combinaisons linéaires à coefficients entiers des polynômes H_n. Il suffit donc de démontrer la propriété pour les polynômes H_n $(n \geqslant 1)$.

Montrons que $H_n(\mathbb{Z}) \subset \mathbb{Z}$ pour commencer[4]. Soit $x \in \mathbb{Z}$.

- Si $0 \leqslant x \leqslant n-1$, alors $H_n(x) = 0$.
- Si $x \geqslant n$, alors $H_n(x) = \dfrac{x!}{k_n(x-n)!} = \dfrac{n!}{k_n} C_x^n \in \mathbb{Z}$ car, par construction, k_n divise $n!$.
- Enfin, si $x < 0$, alors $H_n(x) = \dfrac{(-1)^n(n-x+1)!}{k_n(-x)!} = \dfrac{(-1)^n n!}{k_n} C_{n-x+1}^n$

et de nouveau $H_n(x) \in \mathbb{Z}$.

Le polynôme $k_n H_n$ est à coefficients entiers. Si $u \equiv u' \,[p^i]$, alors d'après les propriétés des congruences, on a $k_n H_n(u) \equiv k_n H_n(u') \,[p^i]$. Autrement dit, p^i divise $k_n H_n(u) - k_n H_n(u')$. Or, par hypothèse, k_n est premier avec p, donc avec p^i. On en déduit, d'après le théorème de Gauss, que p^i divise $H_n(u) - H_n(u')$, c'est-à-dire que $H_n(u) \equiv H_n(u') \,[p^i]$.

b. On peut remarquer que Ψ est un morphisme de groupes additifs (on ne dispose pas d'une structure d'anneau sur S).

- Soit $d \in \mathbb{N}$ et $Q = \sum_{n=0}^{d} a_n H_n$, avec $(a_0, \ldots, a_d) \in \mathbb{Z}^{d+1}$, un élément de S. D'après ce qui précède, on a, pour $0 \leqslant j \leqslant d$,

$$Q(j) = \sum_{n=0}^{j} a_n \frac{n!}{k_n} C_j^n.$$

Supposons que $Q \in \operatorname{Ker}\Psi$. On a alors $Q(j) \equiv 0 \,[p^i]$, pour tout j de $[\![0, d]\!]$. On obtient, en particulier, pour $j = 0$ et $j = 1$, $a_0 \equiv 0 \,[p^i]$ et $a_0 + \dfrac{a_1}{k_1} \equiv 0 \,[p^i]$ et donc $a_0 \equiv \dfrac{a_1}{k_1} \equiv 0 \,[p^i]$. Supposons démontré que pour tout $n \in [\![1, j-1]\!]$, on a $\dfrac{a_n n!}{k_n} \equiv 0 \,[p^i]$. On obtient alors

$$\frac{a_j j!}{k_j} = Q(j) - \sum_{n=0}^{j-1} a_n \frac{n!}{k_n} C_j^n \equiv 0 \,[p^i],$$

ce qui démontre, par récurrence, que pour tout $j \in [\![0, d]\!]$, $\dfrac{a_j j!}{k_j} \equiv 0 \,[p^i]$.

4. En fait, le résultat découle directement de l'exercice 5.24 puisque H_n est un multiple entier d'un polynôme de Hilbert.

Montrons réciproquement, que si cette condition est réalisée, alors $Q \in \operatorname{Ker}\Psi$. En effet, si $0 \leqslant j \leqslant d$, on a $a_j H_j(x) = 0$ pour $0 \leqslant x \leqslant j-1$ et $a_j H_j(x) = \dfrac{a_j j!}{k_j} C_x^j \equiv 0\ [p^i]$, pour $x \geqslant j$. On a donc, pour tout $x \in \mathbb{N}$, $Q(x) \equiv 0\ [p^i]$ et $\Psi(Q) = 0$. Le polynôme Q est dans $\operatorname{Ker}\Psi$.

Notons enfin que la condition trouvée équivaut à $a_j j! \equiv 0[p^i]$, pour tout j de $[\![0, d]\!]$, puisque k_j est premier avec p. On conclut :

$$\boxed{\operatorname{Ker}\Psi = \left\{ \sum_{j=0}^{d} a_j H_j, d \in \mathbb{N}, (a_0, \ldots, a_d) \in \mathbb{Z}^{d+1}, \forall j \in [\![0, d]\!]\ a_j j! \equiv 0[p^i] \right\}.}$$

Remarquons que pour tout $j \in \mathbb{N}$, $\nu_p(a_j j!) = \nu_p(a_j) + \nu_p(j!) \geqslant \nu_p(j!)$. Si $\nu_p(j!) \geqslant i$, alors la condition sur a_j est automatiquement réalisée. Soit j_0 le plus petit des entiers j tels que $\nu_p(j!) \geqslant i$. Il résulte de ce qui précède que le sous-groupe de S engendré par les polynômes H_j, $j \geqslant j_0$ est inclus dans le noyau.

• On peut décrire autrement les éléments du noyau. Soit $Q \in \operatorname{Ker}\Psi$. La condition $\dfrac{a_j j!}{k_j} \equiv 0\ [p^i]$ signifie qu'il existe $b_j \in \mathbb{Z}$ tel que $a_j = \dfrac{b_j k_j}{j!} p^i$. Le polynôme Q s'écrit donc $p^i \sum_{j=0}^{d} b_j H'_j$, où les polynômes H'_j sont les polynômes de Hilbert définies par $H'_0 = 1$ et pour tout $n \in \mathbb{N}^*$,

$$H'_n = \frac{1}{n!} X(X-1)\ldots(X-n+1).$$

Il est classique de montrer que les polynômes qui s'écrivent comme combinaisons linéaires à coefficients entiers des polynômes de Hilbert sont les polynômes de $\mathbb{Q}[X]$ qui envoient $\mathbb{Z}$ dans $\mathbb{Z}$ (voir exercice 5.24). On obtient finalement

$$\boxed{\operatorname{Ker}\Psi = \{p^i Q, Q \in \mathbb{Q}[X], Q(\mathbb{Z}) \subset \mathbb{Z}\}}.$$

• Soit $f \in \Psi(S)$ et $Q \in S$ tel que $\Psi(Q) = f$. Notons comme précédemment $Q = \sum_{n=0}^{d} a_n H_n$. La fonction $\Psi(Q)$ ne dépend que des restes des a_j modulo p^i. On peut donc supposer que les coefficients a_j sont dans $[\![0, p^i - 1]\!]$. D'autre part, l'étude du noyau montre qu'on peut prendre $d < j_0$. En notant

$$S' = \left\{ Q \in S, Q = \sum_{j=0}^{j_0-1} a_j H_j, (a_0, \ldots, a_{j_0-1}) \in [\![0, p^i - 1]\!]^{j_0} \right\},$$

on a donc $\operatorname{Im}\Psi = \Psi(S')$ et $\operatorname{Card}(\operatorname{Im}\Psi) \leqslant \operatorname{Card}(S') = (p^i)^{j_0}$.

Pour préciser $\operatorname{Card}(S')$, il faut évaluer j_0 (défini ci-dessus). Il résulte de l'exercice 4.21 que

$$\nu_p((p^i-1)!) = \sum_{k=1}^{i-1} E\left(\frac{p^i - 1}{p^k}\right) = \sum_{k=1}^{i-1} (p^{i-k} - 1) = \frac{p^i - p}{p-1} - i + 1.$$

⋆ De la formule du binôme on déduit que, si $i \geqslant 2$, on a

$$\nu_p((p^i-1)!) \geqslant \frac{1+(p-1)i+\dfrac{i(i-1)}{2}(p-1)^2-p}{p-1} - i + 1 \geqslant \frac{i(i-1)(p-1)}{2}$$

et $\nu_p((p^i-1)!) \geqslant i$ si $i \geqslant 3$ ou $p \geqslant 3$. On a alors $j_0 \leqslant p^i - 1$ et $\mathrm{Card}(\mathrm{Im}\,\Psi) \leqslant (p^i)^{p^i-1}$. Comme $\mathrm{Card}(\mathcal{F}(\mathbb{Z}/p^i\mathbb{Z}, \mathbb{Z}/p^i\mathbb{Z})) = (p^i)^{p^i}$, l'application Ψ n'est pas surjective si $i \geqslant 3$ ou $i = 2$ et $p \geqslant 3$.

⋆ Dans le cas où $i = 1$, on a $j_0 = p$. Montrons qu'alors la restriction de Ψ à S′ est injective. Si $Q = \sum_{j=0}^{p-1} a_j H_j$ et $Q' = \sum_{j=0}^{p-1} b_j H_j$ sont deux éléments de S′ ayant même image par Ψ, on obtient, d'après l'étude du noyau qui précède, pour tout $j \in [\![0, p-1]\!]$, $(a_j - b_j)\dfrac{j!}{k_j} \equiv 0 \ [p]$. Mais ici $k_j = j!$, car $j < p$. On a donc $a_j \equiv b_j[p]$, *i.e.* $a_j = b_j$, car a_j et b_j sont dans $[\![0, p-1]\!]$. La restriction de Ψ à S′ est donc une bijection de S′ sur $\mathrm{Im}\,\Psi$. On en déduit que $\mathrm{Card}(\mathrm{Im}\,\Psi) = p^p = \mathrm{Card}(\mathcal{F}(\mathbb{Z}/p\mathbb{Z}, \mathbb{Z}/p))$: l'application Ψ est surjective.

⋆ Enfin, dans le cas où $i = p = 2$, on a $j_0 = 4$. On vérifie que les éléments de $Q = \sum_{j=0}^{3} a_j H_j$ et $Q' = \sum_{j=0}^{3} b_j H_j$ de S′ ont même image par Ψ si et seulement si $a_0 = b_0$, $a_1 = b_1$, $a_2 \equiv b_2[2]$ et $a_3 \equiv b_3[2]$. En notant

$$S'' = \left\{ Q \in S, Q = \sum_{j=0}^{3} a_j H_j,\ (a_0, a_1) \in [\![0,3]\!]^2,\ (a_2, a_3) \in [\![0,1]\!]^2 \right\},$$

on a donc $\mathrm{Im}\,\Psi = \Psi(S'')$ et $\mathrm{Card}(\mathrm{Im}\,\Psi) = \mathrm{Card}(S'') = 4 \cdot 4 \cdot 2 \cdot 2 = 64$ et $\mathrm{Card}(\mathcal{F}(\mathbb{Z}/p'\mathbb{Z}, \mathbb{Z}/p^i\mathbb{Z})) = 4^4 = 256$: Ψ n'est pas surjective. ◁

Les exercices suivants ont pour thème les relations entre les coefficients et les fonctions symétriques des racines d'un polynôme.

5.29. Images des racines d'un polynôme

Soit $P = X^4 + aX^2 + bX + c \in \mathbb{C}[X]$.

1. Donner une condition nécessaire et suffisante sur a, b, c pour que les racines de P forment un parallélogramme.

2. Même question avec un rectangle.

(École polytechnique)

▷ **Solution.**

1. Notons x_1, x_2, x_3 et x_4 les racines du polynôme P. Leurs images dans le plan complexe forment un parallélogramme si on peut les or-

donner de telle manière que $x_1 + x_2 = x_3 + x_4$ (relation équivalente à $x_1 - x_3 = x_4 - x_2$). Le coefficient de X^3 est nul donc la somme σ_1 des racines de P est nulle et la condition équivaut à $x_1 + x_2 = x_3 + x_4 = 0$. En considérant σ_3 on obtient

$$b = -\sigma_3 = -x_1x_2(x_3 + x_4) - x_3z_4(x_1 + x_2) = 0.$$

Réciproquement, si $b = 0$ on obtient $P = X^4 + aX^2 + c = Q(X^2)$, où $Q = X^2 + aX + c$. En notant z_1 et z_2 les racines de Q, les racines de P sont les racines carrées de z_1 et z_2 ; elles sont deux à deux opposées et leurs images forment un parallélogramme de centre 0.

2. Un rectangle est un parallélogramme dont les diagonales ont la même longueur. D'après la question précédente, dont on reprend les notations, les racines de P forment un rectangle si et seulement si $b = 0$ et $|z_1| = |z_2|$. On cherche donc à quelle condition les racines de Q ont même module. En notant ω une racine carrée de $\Delta = a^2 - 4c$, on souhaite avoir $|-a-\omega| = |-a+\omega|$, ce qui équivaut à $a\overline{\omega} + \overline{a}\omega = 0$, à $\overline{a}\omega$ imaginaire pur et donc à $\overline{a}^2\omega^2 = \overline{a^2}\Delta$ réel négatif. Ceci est réalisé si $a = 0$. Sinon, on écrit

$$\overline{a^2}\Delta = |a|^2\frac{\Delta}{a^2} = |a|^2\left(1 - \frac{4c}{a^2}\right).$$

Cette expression est un réel négatif si et seulement si $\dfrac{c}{a^2}$ est un réel supérieur ou égal à $\dfrac{1}{4}$.

Conclusion. Les racines de P forment un rectangle si et seulement si $b = 0$ et soit $a = 0$, soit $\dfrac{c}{a^2}$ est un réel supérieur ou égal à $\dfrac{1}{4}$. ◁

5.30. Un calcul de $\zeta(2)$

1. Soit $n \in \mathbb{N}$. Montrer qu'il existe un polynôme P_n tel que, pour tout $t \in \left]0, \dfrac{\pi}{2}\right[$,

$$P_n(\operatorname{cotan}^2 t) = \frac{\sin(2n+1)t}{\sin^{2n+1} t}.$$

2. Expliciter les racines de P_n et calculer leur somme.

3. En observant que l'on a $\operatorname{cotan}^2 t \leqslant \dfrac{1}{t^2} \leqslant 1 + \operatorname{cotan}^2 t$ pour tout $t \in \left]0, \dfrac{\pi}{2}\right[$, déterminer la valeur de $\zeta(2) = \displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2}$.

(École polytechnique)

▷ **Solution.**

1. On applique la formule de Moivre pour calculer $\sin(2n+1)t$ pour tout $t \in \left]0, \frac{\pi}{2}\right[$. On a

$$\sin(2n+1)t = \operatorname{Im}(\cos t + i\sin t)^{2n+1} = \sum_{k=0}^{n} \mathrm{C}_{2n+1}^{2k+1}(-1)^k \sin^{2k+1} t \cos^{2(n-k)} t.$$

En divisant par $\sin^{2n+1} t \neq 0$, on obtient bien $\dfrac{\sin(2n+1)t}{\sin^{2n+1} t} = \mathrm{P}_n(\operatorname{cotan}^2 t)$, où P_n est le polynôme $\mathrm{P}_n(\mathrm{X}) = \sum_{k=0}^{n} \mathrm{C}_{2n+1}^{2k+1}(-1)^k \mathrm{X}^{n-k}$.

Le polynôme P_n est unique car si Q répond également au problème, on a $\mathrm{P}_n(x) = \mathrm{Q}(x)$ pour tout $x > 0$, puisque x peut s'écrire $x = \operatorname{cotan}^2 t$ avec $t = \operatorname{arccotan}(\sqrt{x}) \in \left]0, \frac{\pi}{2}\right[$.

2. Le polynôme P_n est de degré n et il résulte de la première question que pour tout $k \in [\![1, n]\!]$, $x_k = \operatorname{cotan}^2\left(\dfrac{k\pi}{2n+1}\right)$ est racine de P_n. Ces n racines étant deux à deux distinctes, il s'agit des n racines de P_n. La somme des racines de P_n est l'opposé du coefficient de X^{n-1} divisé par le coefficient dominant. Ainsi,

$$x_1 + \cdots + x_n = \frac{\mathrm{C}_{2n+1}^{3}}{\mathrm{C}_{2n+1}^{1}} = \frac{n(2n-1)}{3}.$$

3. La double inégalité proposée découle de l'encadrement bien connu $\sin t \leqslant t \leqslant \tan t$ valable pour tout $t \in \left[0, \frac{\pi}{2}\right[$. Si on doit le prouver on peut soit faire une étude de fonction, soit invoquer un argument de convexité. On écrit alors l'inégalité pour $t = \dfrac{k\pi}{2n+1}$ et on somme pour k variant de 1 à n. Il vient à l'aide de la question précédente

$$\frac{n(2n-1)}{3} \leqslant \sum_{k=1}^{n} \frac{(2n+1)^2}{\pi^2 k^2} \leqslant n + \frac{n(2n-1)}{3}.$$

On divise alors par $\dfrac{(2n+1)^2}{\pi^2}$ et on fait tendre n vers l'infini. On trouve un résultat sans doute bien connu du lecteur : $\boxed{\zeta(2) = \sum_{n=1}^{+\infty} \frac{1}{n^2} = \frac{\pi^2}{6}}$. ◁

Parmi les fonctions polynomiales symétriques en n variables les plus utilisées, figurent les sommes de Newton $\mathrm{S}_p = \sum_{i=1}^{n} x_i^p$. Les formules de Newton sont des relations entre les sommes de Newton et les fonctions symétriques élémentaires.

5.31. Formules de Newton (1707)

Soit n un entier $\geqslant 2$, K un corps commutatif, $x_1, \ldots, x_n$ des éléments de K. On considère, pour $p \in \mathbb{N}$, la somme $S_p = \sum_{i=1}^{n} x_i^p$. On note $\sigma_1, \ldots, \sigma_n$ les fonctions symétriques élémentaires de $x_1, \ldots, x_n$, définies par

$$\sigma_k = \sum_{1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n} x_{i_1} x_{i_2} \ldots x_{i_n}.$$

Démontrer qu'on a les relations suivantes :

1. $S_p - \sigma_1 S_{p-1} + \cdots + (-1)^{n-1}\sigma_{n-1} S_{p-n} + (-1)^n \sigma_n S_{p-n} = 0$ pour $p \geqslant n$.

2. $S_p - \sigma_1 S_{p-1} + \cdots + (-1)^{p-1}\sigma_{p-1} S_1 + (-1)^p p\sigma_p = 0$ pour $1 \leqslant p \leqslant n-1$.

(École polytechnique)

▷ **Solution.**

1. Considérons le polynôme $P = \prod_{i=1}^{n} (X - x_i) \in K[X]$. Il a pour expression

$$P = X^n + \sum_{i=1}^{n} (-1)^i \sigma_i X^{n-i}.$$

Pour $1 \leqslant i \leqslant n$ et $p \geqslant n$, on a $P(x_i) = x_i^n - \sigma_1 x_i^{n-1} + \cdots + (-1)^p \sigma_p = 0$ et donc en multipliant par x_i^{p-n},

$$x_i^p - \sigma_1 x_i^{p-1} + \cdots + (-1)^n \sigma_n x_i^{p-n} = 0.$$

En additionnant, pour $1 \leqslant i \leqslant n$, on obtient

$$\boxed{S_p - \sigma_1 S_{p-1} + \cdots + (-1)^n \sigma_n S_{p-n} = 0}.$$

2. Soit p entier tel que $1 \leqslant p \leqslant n-1$. Ce second cas est nettement plus difficile. Pour commencer, on compare S_p et $\sigma_1 S_{p-1}$. On obtient

$$S_p = \sigma_1 S_{p-1} - \sum_{\substack{1 \leqslant i_1, i_2 \leqslant n \\ i_1 \neq i_2}} x_{i_1} x_{i_2}^{p-1}.$$

On compare ensuite cette dernière somme et $\sigma_2 S_{p-2}$. On a, cette fois,

$$\sigma_2 S_{p-2} = \sum_{\substack{1 \leqslant i_1, i_2 \leqslant n \\ i_1 \neq i_2}} x_{i_1} x_{i_2}^{p-1} + \sum_{\substack{1 < i_1 < i_2 \leqslant n \\ i_3 \neq i_1, i_2}} x_{i_1} x_{i_2} x_{i_3}^{p-2}.$$

Plus généralement, on pose, pour $0 \leqslant k \leqslant p-1$,

$$A_k = \sum_{\substack{1 \leqslant i_1 < \cdots < i_k \leqslant n \\ i_{k+1} \neq i_1, i_2, \ldots, i_k}} x_{i_1} \ldots x_{i_k} x_{i_{k+1}}^{p-k}.$$

On remarque qu'en particulier $A_0 = S_p$ et $A_{p-1} = p\sigma_p$. On obtient alors, pour $1 \leqslant k \leqslant p-1$,

$$\sigma_k S_{p-k} = A_{k-1} + A_k.$$

En multipliant la première relation par -1, la seconde par $(-1)^2$, ... et la $(p-1)$-ième par $(-1)^{p-1}$, on trouve

$$\sum_{k=1}^{p-1} (-1)^k \sigma_k S_{p-k} = -A_0 + (-1)^{p-1} A_{p-1} = -S_p + (-1)^{p-1} p\sigma_p,$$

ce qui, si on fait tout passer dans le membre de gauche de l'égalité, nous donne le résultat voulu.

De ce résultat, nous proposons une preuve alternative fondée sur les séries génératrices.

Soit $\alpha \in \mathbb{C}$. On peut écrire le développement en série entière

$$\frac{1}{1-\alpha x} = \sum_{p=0}^{+\infty} \alpha^p x^p,$$

valable pour $|\alpha x| < 1$. Si on écrit ce développement pour $\alpha = x_i$, pour $1 \leqslant i \leqslant n$ et si on somme les relations obtenues, on obtient

$$\sum_{i=1}^{n} \frac{1}{1-x_i x} = \sum_{p=0}^{+\infty} S_p x^p,$$

développement valable dans un voisinage de 0. La somme $\displaystyle\sum_{i=1}^{n} \frac{1}{1-X_i x}$ est le développement en éléments simples d'une fraction rationnelle, que nous allons déterminer. On sait que si on pose $P = (X-x_1)\ldots(X-x_n)$, on obtient

$$\frac{P'}{P} = \sum_{i=1}^{n} \frac{1}{X-x_i} \quad \text{et donc} \quad \frac{XP'}{P} = \sum_{i=1}^{n} \frac{1}{1-x_i/X},$$

puis en substituant $1/X$ à X,

$$\sum_{i=1}^{n} \frac{1}{1-x_i X} = \frac{P'(1/X)}{XP(1/X)} = \frac{X^{n-1}P'(1/X)}{X^n P(1/X)}.$$

Puisque $P = \displaystyle\sum_{k=0}^{n} (-1)^k \sigma_k X^{n-k}$ et $P' = \displaystyle\sum_{k=0}^{n-1} (-1)^k (n-k)\sigma_k X^{n-k-1}$ (en posant $\sigma_0 = 1$), on obtient

$$\sum_{i=1}^{n} \frac{1}{1-x_i \mathrm{X}} = \frac{\sum_{k=0}^{n-1} (-1)^k (n-k)\sigma_k \mathrm{X}^k}{\sum_{k=0}^{n} (-1)^k \sigma_k \mathrm{X}^k}.$$

Par conséquent, pour x voisin de 0, on peut écrire

$$\left(\sum_{k=0}^{n-1} (-1)^k (n-k)\sigma_k x^k\right) = \left(\sum_{p=0}^{+\infty} \mathrm{S}_p x^p\right)\left(\sum_{k=0}^{n} (-1)^k \sigma_k x^k\right).$$

D'après l'unicité des coefficients d'une série entière et la règle du produit de Cauchy, le coefficient de x^p (pour $1 \leqslant p \leqslant n-1$) vaut

$$\begin{aligned}(-1)^p \sigma_p (n-p) &= \sum_{k=0}^{p} \mathrm{S}_k (-1)^{p-k} \sigma_{p-k} \\ &= \mathrm{S}_p - \sigma_1 \mathrm{S}_{p-1} + \sigma_2 \mathrm{S}_{p-2} - \cdots + (-1)^p \sigma_p \underbrace{\mathrm{S}_0}_{=n}\end{aligned}$$

et finalement

$$\boxed{\mathrm{S}_p - \sigma_1 \mathrm{S}_{p-1} + \cdots + (-1)^{p-1} \sigma_{p-1} \mathrm{S}_1 + (-1)^p p \sigma_p = 0}. \quad \lhd$$

Ces formules permettent de déterminer les sommes de Newton de proche en proche. Inversement, si le corps est de caractéristique nulle (il faut pouvoir diviser par les entiers), la connaissance des sommes de Newton permet de déterminer les fonctions symétriques élémentaires. Une utilisation de cela est faite dans l'exercice 5.27.

L'exercice suivant est l'occasion de rappeler que toute expression rationnelle et symétrique de $x_1, x_2, \ldots, x_n$ peut s'exprimer rationnellement[5] en fonction de $\sigma_1, \ldots, \sigma_n$, fonctions symétriques élémentaires de $x_1, \ldots, x_n$. Ici, on se contente de le constater sur un exemple particulier.

5.32. Polynômes réels scindés

On considère E l'ensemble des polynômes réels scindés sur $\mathbb{R}$ et à coefficients dans $\{0, -1, 1\}$. On souhaite décrire E.

1. Montrer qu'il suffit de connaître les éléments de E unitaires et qui ne s'annulent pas en 0.

5. Le lecteur pourra trouver une preuve de ce théorème dans ARNAUDIÈS (J.-M.) & BRAYSSE (H.), *Cours de Mathématiques, Algèbre 1*, Dunod, 1987, Chapitre X.

2. Pour n réels strictement positifs $a_1, \dots, a_n$, démontrer l'inégalité $\sum_{1 \leqslant i,j \leqslant n} \frac{a_i}{a_j} \geqslant n^2$.

3. Soit P un polynôme unitaire non constant dans E tel que $P(0) \neq 0$. Démontrer que $\deg P \leqslant 3$. En déduire P.

(**École normale supérieure**)

▷ **Solution.**

1. Il est évident que P est dans E si et seulement si $-P$ est dans E. Il nous suffit donc de connaître les éléments unitaires de E. Si 0 est racine de $P \in E$ avec une multiplicité k, on peut écrire $P = X^k Q$ avec $Q \in \mathbb{R}[X]$. Or, Q est scindé et ses coefficients sont aussi dans $\{0, -1, 1\}$. Réciproquement, si Q est dans E, alors $X^k Q \in E$ pour tout entier k.

En résumé, les éléments non nuls de E s'écrivent $\pm X^k Q$, où $k \in \mathbb{N}$ et où Q est un élément unitaire de E n'ayant pas 0 comme racine.

2. Cette inégalité résulte du fait que, si x, y sont des réels strictement positifs, on a $\frac{x}{y} + \frac{y}{x} \geqslant 2$. Dans la somme $\sum_{1 \leqslant i,j \leqslant n} \frac{a_i}{a_j}$, on regroupe alors le terme (i, j) avec le terme (j, i) pour $i \neq j$. Comme le quotient vaut 1 lorsque $i = j$, il vient

$$\sum_{1 \leqslant i,j \leqslant n} \frac{a_i}{a_j} \geqslant n + 2C_n^2 = n^2.$$

3. Notons n le degré de P. Écrivons $P = X^n - \sigma_1 X^{n-1} + \cdots + (-1)^n \sigma_n$ où les σ_k sont les fonctions symétriques élémentaires des racines de P. Notons $x_1, \dots, x_n$ ces racines distinctes ou confondues. Par hypothèse, elles sont toutes réelles et non nulles.

L'inégalité de la question précédente appliquée aux x_i^2 donne

$$S = \sum_{1 \leqslant i,j \leqslant n} \frac{x_i^2}{x_j^2} \geqslant n^2.$$

Or S est une expression rationnelle des x_i et elle va s'exprimer à l'aide des σ_k. Comme ces derniers sont majorés par 1, on pourra en déduire une majoration de n. Plus précisément, on a

$$\begin{aligned} S = \sum_{1 \leqslant i,j \leqslant n} \frac{x_i^2}{x_j^2} &= \left(\sum_{i=1}^n x_i^2 \right) \left(\sum_{i=1}^n \frac{1}{x_i^2} \right) \\ &= (\sigma_1^2 - 2\sigma_2) \left(\left(\frac{\sigma_{n-1}}{\sigma_n} \right)^2 - 2 \frac{\sigma_{n-2}}{\sigma_n} \right). \end{aligned}$$

On a alors clairement $S \leqslant 3 \times 3 = 9$ et donc $n \leqslant 3$. On peut alors mener une recherche exhaustive.

⋆ Pour $n = 1$, $X - 1$ et $X + 1$ conviennent pour P.

⋆ Pour $n = 2$ on obtient $X^2 - 1$, $X^2 + X - 1$, $X^2 - X - 1$.

⋆ Pour $n = 3$, l'étude du cas d'égalité dans la majoration obtenue plus haut donne $\sigma_1 = \pm 1$, $\sigma_2 = -1$, $\sigma_3 = \pm 1$ et $\sigma_1 = -\sigma_3$. On obtient alors les polynômes $X^3 + X^2 - X - 1$ et $X^3 - X^2 - X + 1$, qui conviennent tous les deux. ◁

Le théorème rappelé avant l'exercice précédent se généralise aux polynômes à coefficients dans un anneau. Ainsi, si P *est un polynôme symétrique en* $x_1, \ldots, x_n$ *à coefficients dans* $\mathbb{Z}$, *il existe* Q *à coefficients dans* $\mathbb{Z}$ *tel que* $P(x_1, \ldots, x_n) = Q(\sigma_1, \ldots, \sigma_n)$. *On pourra exploiter cela dans l'exercice qui suit.*

5.33. Un théorème de Kronecker

Soit P un polynôme unitaire de $\mathbb{Z}[X]$ dont les racines complexes sont toutes de module inférieur ou égal à 1. On suppose $P(0) \neq 0$. Montrer que toutes les racines de P sont des racines de l'unité.

(**École normale supérieure**)

▷ **Solution.**

• Notons $z_1, z_2, \ldots, z_n$ les racines de P comptées avec leur multiplicité, $n \geqslant 1$ désignant le degré de P, et $\sigma_1, \ldots, \sigma_n$ les fonctions symétriques élémentaires des z_i. On a donc

$$P(X) = X^n - \sigma_1 X^{n-1} + \cdots + (-1)^n \sigma_n$$

et les σ_i sont dans $\mathbb{Z}$. Comme $|z_i| \leqslant 1$ pour tout $i \in [\![1, n]\!]$, on obtient, pour $1 \leqslant p \leqslant n$, la majoration

$$|\sigma_p| = \left| \sum_{I \in \mathcal{P}_p([\![1,n]\!])} \prod_{i \in I} z_i \right| \leqslant \sum_{I \in \mathcal{P}_p([\![1,n]\!])} 1 = \operatorname{Card} \mathcal{P}_p([\![1, n]\!]) = C_n^p.$$

Ceci montre que l'ensemble Ω_n des polynômes unitaires de $\mathbb{Z}[X]$, de degré n, dont les racines complexes sont de module inférieur ou égal à 1 est fini.

• L'idée est alors de considérer pour tout $k \in \mathbb{N}^*$ les polynômes $P_k(X) = (X - z_1^k)(X - z_2^k) \ldots (X - z_n^k)$. Ce sont des polynômes unitaires de degré n dont les racines z_i^k sont de module inférieur ou égal à 1. Pour tout $r \in [\![1, n]\!]$, le coefficient de X^{n-r} dans P_k est $(-1)^r \sigma_r(z_1^k, \ldots, z_n^k)$. Il s'agit d'un polynôme symétrique en $z_1, z_2, \ldots, z_n$ à coefficients dans $\mathbb{Z}$. Il est donc égal à un polynôme à coefficients entiers en les fonctions symétriques élémentaires des z_i. Il en résulte que, pour tout entier k, on a $P_k \in \mathbb{Z}[X]$ et donc que $P_k \in \Omega_n$.

Comme Ω_n est fini, et que chaque élément de Ω_n admet au plus n racines complexes distinctes, l'ensemble des racines des éléments de Ω_n est aussi fini. En particulier, pour tout $i \in [\![1, n]\!]$, l'application $k \longmapsto z_i^k$ ne peut pas être injective. Il existe donc deux entiers $k \neq \ell$ tels que $z_i^k = z_i^\ell$. Comme z_i est supposé non nul, il s'agit d'une racine de l'unité. ◁

Si $P \in \mathbb{C}[X]$ *est unitaire on appelle mesure de Mahler*[6] *de* P *et on note* M(P), *le produit* $\prod_{P(z)=0} \max(1, |z|)$. *Le théorème de Kronecker montre donc que si* $P \in \mathbb{Z}[X]$ *est un polynôme unitaire de mesure égale à* 1, *ses racines sont des racines de l'unité (ou* 0*). Un célèbre problème de Mahler consiste à chercher s'il existe* $\varepsilon > 0$ *tel que le résultat reste vrai pour les polynômes de mesure inférieure à* $1 + \varepsilon$.

Dans $\mathbb{C}[X]$ *tout polynôme est scindé. Il n'en est rien dans* $\mathbb{R}[X]$ *et c'est l'objet de nombreux exercices, des plus simples aux plus sophistiqués, de chercher combien de racines réelles possède un polynôme réel et s'il est scindé. Il apparaît, dans les exercices suivants, que la recherche des racines réelles d'un polynôme de* $\mathbb{R}[X]$ *fait essentiellement appel à des techniques d'analyse.*

5.34. Racines réelles de $nX^n - X^{n-1} - \cdots - X - 1$

Combien $P_n = nX^n - X^{n-1} - \cdots - X - 1$ possède-t-il de racines réelles ?

(École normale supérieure)

▷ **Solution.**

On observe que 1 est racine de P_n pour tout $n \in \mathbb{N}^*$. Pour $n \geqslant 2$, on met en facteur $X - 1$ pour obtenir

$$P_n = (X-1)(nX^{n-1} + (n-1)X^{n-2} + \cdots + 2X + 1) = (X-1)Q_n'$$

avec $Q_n = X^n + X^{n-1} + \cdots + X + 1$. On a $Q_n = \dfrac{X^{n+1} - 1}{X - 1}$ de sorte qu'en dérivant, il vient

$$Q_n' = \frac{nX^{n+1} - (n+1)X^n + 1}{(X-1)^2}.$$

L'étude des variations du numérateur est enfantine et montre que Q_n' admet exactement une racine réelle (négative) si n est pair et n'admet aucune racine réelle si n est impair.

6. Le lecteur trouvera dans l'exercice 5.47 une inégalité (due à Landau) qui fournit une majoration de M(P).

Conclusion. Si n est pair, P_n admet 2 racines réelles (dont 1) et si n est impair, 1 est l'unique racine réelle de P_n. ◁

5.35. Un polynôme scindé sur $\mathbb{R}$

Montrer que, pour tous $n \geqslant 1$ et $\theta \in \mathbb{R}$ non multiple de π, le polynôme $P = \sum_{k=0}^{n} C_n^k \sin k\theta \, X^k$ a toutes ses racines réelles.

(**École polytechnique**)

▷ **Solution.**

Compte tenu de la factorisation $X^n - Y^n = \prod_{k=0}^{n-1} (X - e^{\frac{2ik\pi}{n}} Y)$, on peut écrire

$$\begin{aligned}
2iP(X) &= \sum_{k=0}^{n} C_n^k 2i \sin k\theta \, X^k = \sum_{k=0}^{n} C_n^k \left((e^{i\theta}X)^k - (e^{-i\theta}X)^k \right) \\
&= \left(1 + e^{i\theta}X\right)^n - \left(1 + e^{-i\theta}X\right)^n \\
&= \prod_{k=0}^{n-1} \left(1 + e^{i\theta}X - e^{i\frac{2k\pi}{n}}(1 + e^{-i\theta}X)\right) \\
&= \prod_{k=0}^{n-1} \left(1 - e^{i\frac{2k\pi}{n}} + (e^{i\theta} - e^{i\frac{2k\pi}{n}} e^{-i\theta})X\right).
\end{aligned}$$

On en déduit que si x, nombre complexe, est racine de P, l'un des facteurs $\left(1 - e^{i\frac{2k\pi}{n}} + (e^{i\theta} - e^{i\frac{2k\pi}{n}} e^{-i\theta})x\right)$ est nul. On a nécessairement $e^{i\theta} \neq e^{i\frac{2k\pi}{n}} e^{-i\theta}$; car sinon $1 = e^{i\frac{2k\pi}{n}}$ et $e^{i\theta} = e^{-i\theta}$, autrement dit $\theta \in \pi\mathbb{Z}$. On est donc en droit d'écrire

$$x = \frac{e^{i\frac{2k\pi}{n}} - 1}{e^{i\theta} - e^{i\frac{2k\pi}{n}} e^{-i\theta}} = \frac{e^{i\frac{k\pi}{n}} - e^{-i\frac{k\pi}{n}}}{e^{i\left(\theta - \frac{k\pi}{n}\right)} - e^{i\left(-\theta + \frac{k\pi}{n}\right)}} = \frac{\sin \dfrac{k\pi}{n}}{\sin\left(\theta - \dfrac{k\pi}{n}\right)}.$$

Le polynôme P n'a donc que des racines réelles. ◁

Le théorème de Rolle est un instrument efficace pour trouver des racines aux polynômes réels, comme l'illustre l'exercice suivant.

5.36. Dénombrement de racines réelles

Soit $P \in \mathbb{R}[X]$, un polynôme admettant n racines réelles simples strictement supérieures à 1. On pose

$$Q(X) = (1+X^2)P(X)P'(X) + X(P(X)^2 + P'(X)^2).$$

Montrer que Q a au moins $2n-1$ racines réelles distinctes.

(École polytechnique)

▷ **Solution.**

On a $Q = PP'X^2 + (P^2 + P'^2)X + PP'$, expression qui se factorise en

$$Q = (XP + P')(XP' + P)$$

(regarder Q comme un «polynôme du second degré en X»).

• Le polynôme $XP' + P$ est le polynôme dérivé de XP. Ce dernier admet comme racines celles de P, que l'on note $1 < x_1 < x_2 < \cdots < x_n$, et la racine $x_0 = 0$. Le théorème de Rolle nous assure de l'existence d'une racine de $XP' + P$ dans chaque intervalle $]x_i, x_{i+1}[$, $i \in [\![0, n-1]\!]$. Cela nous donne déjà n racines distinctes.

• Pour le second facteur, on introduit la fonction définie sur $\mathbb{R}$ par $f(x) = e^{x^2/2}P(x)$. Elle est de classe C^∞ et $f'(x) = e^{x^2/2}(P'(x) + xP(x))$. Le théorème de Rolle nous donne donc ici aussi $n-1$ racines distinctes pour $XP + P'$ (une dans chaque intervalle $]x_i, x_{i+1}[$, $i \in [\![1, n-1]\!]$).

• Il reste enfin à vérifier que ces dernières racines diffèrent des précédentes. Or si $P(x) + xP'(x) = 0$ et $P'(x) + xP(x) = 0$, il vient $(1-x^2)P(x) = 0$. Mais par construction les racines obtenues ci-dessus ne sont ni des racines de P, ni égales à ± 1. D'où le résultat : Q admet au moins $2n-1$ racines réelles distinctes. ◁

On retiendra le résultat classique redémontré dans l'exercice suivant : si $P \in \mathbb{R}[X]$ est scindé sur $\mathbb{R}$, alors il en est de même de P'.

5.37. Dérivation et polynômes réels scindés

Soit $(P, Q) \in \mathbb{R}[X]^2$, $Q = \sum a_k X^k$. On pose $R = \sum a_k P^{(k)}$. Montrer que si P et Q sont scindés sur $\mathbb{R}$, alors R est scindé sur $\mathbb{R}$.

(École normale supérieure)

▷ **Solution.**

Si D désigne l'opérateur de dérivation sur $\mathbb{R}[X]$, on observe qu'en fait $R = Q(D)(P)$. Si Q est constant le résultat est trivial. Dans la suite, Q sera donc supposé de degré $p \geqslant 1$ et unitaire (ce qui ne nuit

nullement à la généralité). Il s'écrit donc $Q(X) = (X - \lambda_1)\dots(X - \lambda_p)$ (les λ_i n'étant pas nécessairement deux à deux distincts). On obtient alors $Q(D) = (D - \lambda_1 Id) \circ \cdots \circ (D - \lambda_p Id)$ dans l'algèbre $\mathcal{L}(\mathbb{R}[X])$. Il suffit donc de prouver le résultat suivant : pour tout $P \in \mathbb{R}[X]$ scindé et tout réel λ, $P' - \lambda P$ est aussi scindé.

Si P est constant, c'est clair. Supposons donc $n = \deg P \geqslant 1$ et notons $x_1 < \cdots < x_k$ les racines distinctes de P, de multiplicités respectives $\alpha_1, \dots, \alpha_k$.

• Le cas $\lambda = 0$ est classique : les x_i sont racines de P' d'ordre $\alpha_i - 1$, ce qui fournit $\sum_{i=1}^{k} (\alpha_i - 1) = n - k$ racines de P' comptés avec multiplicités. De plus, si $k > 1$, l'application du théorème de Rolle sur chacun des intervalles $[x_i, x_{i+1}]$, $1 \leqslant i \leqslant k-1$, fournit $k-1$ autres racines de P', deux à deux distinctes et distinctes des précédentes. Ce qui fait le bon compte.

• Le cas λ quelconque se traite de même : x_i est toujours racine de $P' - \lambda P$ de multiplicité $\alpha_i - 1$. Considérons alors la fonction définie sur $\mathbb{R}$ par $f(x) = P(x)e^{-\lambda x}$. On a $f'(x) = (P'(x) - \lambda P(x))e^{-\lambda x}$ et le théorème de Rolle qui s'applique encore sur les intervalles $[x_i, x_{i+1}]$ donne à nouveau $k-1$ racines supplémentaires. On a donc $(n-k)+(k-1) = n-1$ racines comptées avec multiplicité pour $P' - \lambda P$. Ce polynôme étant de degré n, il est nécessairement scindé (le facteur qui reste étant de degré 1).

Remarquons que si P est à racines simples, $P' - \lambda P$ l'est également. Il en résulte que si P est scindé à racines simples, R l'est aussi. ◁

5.38. Un théorème de Laguerre

Soit P et Q deux polynômes non nuls de $\mathbb{R}[X]$, scindés sur $\mathbb{R}$. On suppose que les racines de Q ne sont pas dans l'intervalle $[0, \deg P]$. On note $P = \sum_{k=0}^{n} a_k X^k$ et $R = \sum_{k=0}^{n} a_k Q(k) X^k$. Montrer que R est scindé sur $\mathbb{R}$.

(École polytechnique)

▷ **Solution.**

Sans perte de généralité, on peut supposer P et Q unitaires. On note n le degré de P et q celui de Q.

Une première constatation est qu'il suffit de montrer le résultat pour un polynôme Q de degré 1 (c'est trivial si Q est constant), car le résultat s'étend ensuite facilement par récurrence sur q. En effet, supposons le résultat vrai au rang $q-1$. On écrit $Q = (X - \alpha)S$, où $S \in \mathbb{R}[X]$ est de degré $q-1$ et α est une racine de Q. D'après l'hypothèse de récurrence,

le polynôme $T = \sum_{k=0}^{n} a_k S(k) X^k$ est scindé sur $\mathbb{R}$ et d'après le cas $q = 1$, le polynôme $R = \sum_{k=0}^{n} (a_k S(k))(k-\alpha)X^k = \sum_{k=0}^{n} a_k Q(k) X^k$ est scindé sur $\mathbb{R}$.

On suppose donc dans la suite que $Q = X - \alpha$, où par hypothèse $\alpha \notin [0, \deg P] = [0, n]$. On a alors, avec $P = \sum a_k X^k$,

$$R = \sum_{k=0}^{n} a_k (k-\alpha) X^k = \sum_{k=0}^{n} a_k k X^k - \alpha \sum_{k=0}^{n} a_k X^k = XP' - \alpha P.$$

Le coefficient de X^n dans R est $a_n(n-\alpha) \neq 0$ de sorte que R est un polynôme de degré n.

On observe qu'une racine non nulle x_0 d'ordre k de P est racine d'ordre $k-1$ de R. En revanche, si 0 est racine d'ordre k de P, 0 reste racine d'ordre k de R. En effet, si $P = X^k P_1$ avec $P_1(0) \neq 0$, on obtient $P' = X^{k-1}(kP_1 + XP_1')$ et $R = X^k[(k-\alpha)P_1 + XP_1']$ et le polynôme entre crochets ne s'annule pas en 0 car $\alpha \neq k$.

On considère donc l'ensemble des racines de P auquel on rajoute, s'il n'est pas racine, le réel 0. On note $\alpha_1 < \alpha_2 < \cdots < \alpha_r$ les éléments de cet ensemble et i_0 l'entier entre 1 et r tel que $\alpha_{i_0} = 0$. On note s_i l'ordre de α_i comme racine de P pour tout i compris entre 1 et r. Si $i \neq i_0$, on a $s_i \geqslant 1$, mais s_{i_0} peut être nul.

D'après ce qui est dit plus haut, pour $i \neq i_0$, α_i est racine d'ordre $s_i - 1$ de R ; $0 = \alpha_{i_0}$ est racine d'ordre s_{i_0} de R. Comme $s_1 + \cdots + s_r = n$, pour que R soit scindé sur $\mathbb{R}$, il suffit de trouver $r-1$ autres racines de R, distinctes des α_i, puisque

$$(s_1 - 1) + \cdots + s_{i_0} + \cdots + (s_r - 1) + (r-1) = n.$$

On remarque que $R = XP\left(\dfrac{P'}{P} - \dfrac{\alpha}{X}\right)$. En écrivant

$$P' = \sum_{i=1}^{r} s_i (X - \alpha_i)^{s_i - 1} \prod_{k \neq i} (X - \alpha_k)^{s_k} \quad \text{et} \quad \frac{P'}{P} = \sum_{i=1}^{r} \frac{s_i}{X - \alpha_i},$$

on obtient

$$R = XP\left(\frac{s_1}{X - \alpha_1} + \cdots + \frac{s_{i_0} - \alpha}{X} + \cdots + \frac{s_r}{X - \alpha_r}\right).$$

Il suffit donc de prouver que la fonction

$$f : x \longmapsto \left(\frac{s_1}{x - \alpha_1} + \cdots + \frac{s_{i_0} - \alpha}{x} + \cdots + \frac{s_r}{x - \alpha_r}\right)$$

s'annule $r-1$ fois sur son domaine de définition $D = \mathbb{R} \backslash \{\alpha_1, \ldots, \alpha_r\}$. Nous distinguons deux cas.

⋆ Premier cas : $\alpha < 0$.

La fonction f est continue sur D. Comme les s_i sont strictement positifs ainsi que $s_{i_0} - \alpha$, on a, pour $1 \leqslant i \leqslant r-1$,

$$\lim_{\alpha_i^+} f = +\infty \quad \text{et} \quad \lim_{\alpha_{i+1}^-} f = -\infty$$

et d'après le théorème des valeurs intermédiaires, f s'annule sur $]\alpha_i, \alpha_{i+1}[$. D'où $r-1$ autres racines : c'est ce qu'on voulait.

⋆ Deuxième cas : $\alpha > n$.

On a $s_{i_0} - \alpha < 0$. Comme dans le premier cas, f s'annule sur $]\alpha_i, \alpha_{i+1}[$ pour $i \neq i_0 - 1$ et $i \neq i_0$. Mais on ne peut plus affirmer que f s'annule sur $]\alpha_{i_0-1}, \alpha_{i_0}[$ et sur $]\alpha_{i_0}, \alpha_{i_0+1}[$. Par contre, on a

$$f(x) \underset{-\infty}{\sim} \frac{s_1 + \cdots + s_{i_0} - \alpha + \cdots + s_r}{x} = \frac{n-\alpha}{x} \to 0^+ \quad \text{et} \quad \lim_{\alpha_1^-} f = -\infty,$$

si $\alpha_1 \neq 0$, *i.e.* $i_0 \neq 1$. D'après le théorème des valeurs intermédiaires, f s'annule sur $]-\infty, \alpha_1[$ si $i_0 \neq 1$. De manière analogue, f s'annule sur $]\alpha_r, +\infty[$ si $i_0 \neq r$. On a donc encore $r-1$ nouvelles racines (si $i_0 = 1$ ou r, f s'annule sur $r-2$ intervalles $]\alpha_i, \alpha_{i+1}[$).

Dans les deux cas, nous avons en tout $r-1$ racines nouvelles et R est scindé. ◁

5.39. Plans vectoriels de polynômes scindés

Soit (P, Q) un couple de polynômes réels.

1. On suppose que P et Q sont scindés à racines simples et que leurs racines sont entrelacées, c'est-à-dire que si α et β sont deux racines de l'un ($\alpha < \beta$), il y a au moins une racine de l'autre dans l'intervalle $]\alpha, \beta[$. Montrer que le polynôme $\lambda P + \mu Q$ reste scindé à racines simples lorsque (λ, μ) décrit $\mathbb{R}^2$.

2. Montrer que si pour tout $(\lambda, \mu) \in \mathbb{R}^2$, le polynôme $\lambda P + \mu Q$ est scindé à racines simples, alors les racines de P et Q sont entrelacées.

(**École normale supérieure**)

▷ **Solution.**

1. • De l'hypothèse, on déduit qu'entre deux racines consécutives de P (resp. Q) il y a exactement une racine de Q (resp. P) et donc que P et Q n'ont aucune racine commune. Ceci montre que, soit les degrés de P et Q sont égaux, soit ils diffèrent de 1. Supposons que $\deg(P) \leqslant \deg(Q)$, notons n le degré de Q et $x_1, \ldots, x_n$ ses racines ($x_1 < \cdots < x_n$). On peut supposer P et Q unitaires. La propriété à démontrer est évidente

si $\lambda = 0$ ou $\mu = 0$; on suppose désormais $\lambda\mu \neq 0$. Le polynôme $\lambda P + \mu Q$ est alors de degré n sauf si $\deg(P) = \deg(Q) = n$ et si $\lambda + \mu = 0$.

• α est racine de $\lambda P + \mu Q$ si $Q(\alpha) \neq 0$ et $\dfrac{P(\alpha)}{Q(\alpha)} = -\dfrac{\mu}{\lambda}$. Nous allons donc étudier l'équation $F(x) = 0$, où F est la fraction rationnelle $\dfrac{P}{Q}$.

Par hypothèse, pour $1 \leqslant i \leqslant n-1$, Q ne change pas de signe et P change une fois de signe sur l'intervalle $]x_i, x_{i+1}[$. La fraction F change donc également de signe sur $]x_i, x_{i+1}[$. Puisque x_i et x_{i+1} sont des pôles simples de F, on en déduit que

$$\left(\lim_{x_i^+} F = -\infty \text{ et } \lim_{x_{i+1}^-} F = +\infty\right) \text{ ou } \left(\lim_{x_i^+} F = +\infty \text{ et } \lim_{x_{i+1}^-} F = -\infty\right).$$

Dans tous les cas, pour tout $i \in [\![1, n-1]\!]$, la fonction F prend une fois la valeur $-\dfrac{\mu}{\lambda}$ sur $]x_i, x_{i+1}[$, d'après le théorème des valeurs intermédiaires. La fraction F prend déjà $n-1$ fois la valeur $-\dfrac{\mu}{\lambda}$.

De plus, on a d'une part $\lim_{x_1^-} F = -\infty$ et $\lim_{x_n^+} F = +\infty$ ou $\lim_{x_1^-} F = +\infty$ et $\lim_{x_n^+} F = -\infty$, et d'autre part $\lim_{-\infty} F = \lim_{+\infty} F = 0$ (si $\deg P = n-1$) ou $\lim_{-\infty} F = \lim_{+\infty} F = 1$ (si $\deg P = n$). Si $\deg P = n-1$, alors F prend une fois la valeur $-\dfrac{\mu}{\lambda} \neq 0$, soit sur $]-\infty, x_1[$, soit sur $]x_n, +\infty[$, toujours d'après le théorème des valeurs intermédiaires. Cela reste vrai quand $\deg P = n$, sauf si $-\dfrac{\mu}{\lambda} = 1$.

Concluons. Si $\deg P = n-1$ ou $-\dfrac{\mu}{\lambda} \neq 1$, le polynôme $\lambda P + \mu Q$ possède n racines réelles distinctes. Il est donc scindé. Si $\deg P = n$ et si $-\dfrac{\mu}{\lambda} = 1$, le polynôme $\lambda P + \mu Q$ est de degré $n-1$ et possède $n-1$ racines distinctes. Il est encore scindé. Dans tous les cas, $\lambda P + \mu Q$ est scindé à racines simples.

2. • On voit en prenant $(\lambda, \mu) = (1, 0)$ ou $(0, 1)$ que P et Q sont scindés à racines simples. Montrons qu'ils n'ont aucune racine commune, en raisonnant par l'absurde. Si α est une racine commune à P et Q, il existe des polynômes P_1 et Q_1 tels que $P = (X-\alpha)P_1$ et $Q = (X-\alpha)Q_1$. On a alors, pour tout $(\lambda, \mu) \in \mathbb{R}^2$, $\lambda P + \mu Q = (X-\alpha)(\lambda P_1 + \mu Q_1)$. Mais α n'est racine ni de P_1, ni de Q_1. Si on choisit λ et μ tels que $\dfrac{\lambda}{\mu} = -\dfrac{Q_1(\alpha)}{P_1(\alpha)}$, alors α est racine double de $\lambda P + \mu Q$, ce qui est contraire à l'hypothèse.

• Montrons que les racines de P et Q sont entrelacées en raisonnant par l'absurde. Supposons par exemple qu'il existe deux racines consécutives de P, x et x' telles que Q ne s'annule pas sur le segment $[x, x']$ (on sait déjà qu'il ne s'annule ni en x ni en x'). La fonction $F = \dfrac{P}{Q}$ est donc définie et dérivable sur $[x, x']$; elle s'annule en x et x'. D'après le théorème de Rolle, il existe $\xi \in]x, x'[$, tel que $F'(\xi) = 0$. Alors ξ est

racine multiple de la fraction rationnelle $F - F(\xi)$. En multipliant par Q, on obtient que ξ est racine multiple de $P - F(\xi)Q$. C'est contraire à l'hypothèse. Les racines de P et Q sont donc entrelacées. $\lhd$

L'exercice suivant étudie dans un cas simple la question de la continuité des racines d'un polynôme par rapport à ses coefficients.

5.40. L'ouvert des polynômes scindés à racines simples sur $\mathbb{R}$ dans l'ensemble des polynômes unitaires de degré n

Soit $P = X^n + a_1X^{n-1} + \cdots + a_n$, un polynôme à coefficients réels scindé sur $\mathbb{R}$, à racines simples. Soit $Q = X^n + b_1X^{n-1} + \cdots + b_n$ un polynôme à coefficients réels. Montrer que si les b_i sont proches des a_i, alors Q est scindé sur $\mathbb{R}$, à racines simples.

(École polytechnique)

$\rhd$ **Solution.**

Notons $x_1, \ldots, x_n$ les n racines, rangées par ordre croissant, du polynôme P. Celui-ci s'annule et change de signe en $x_1, \ldots, x_n$, puisque ces racines sont simples. Considérons des réels $y_1, \ldots, y_n, y_{n+1}$ tels que $y_1 < x_1 < y_2 < \cdots < x_n < y_{n+1}$. On a alors $P(y_k)P(y_{k+1}) < 0$ pour tout $k \in [\![1, n]\!]$. La fonction

$$\Phi : (b_1, \ldots, b_n) \in \mathbb{R}^n \longmapsto (Q(y_1)Q(y_2)\ldots, Q(y_n)Q(y_{n+1})) \in \mathbb{R}^n$$

est continue puisque polynomiale. Par hypothèse, $\Phi(a_1, \ldots, a_n)$ appartient à $(\mathbb{R}_-^*)^n$. Cet ensemble étant un ouvert, on aura, pour $(b_1, \ldots, b_n)$ proche de $(a_1, \ldots, a_n)$ (pour une norme quelconque de $\mathbb{R}^n$, celles-ci étant équivalentes), $\Phi(b_1, \ldots, b_n) \in (\mathbb{R}_-^*)^n$, c'est-à-dire $Q(y_k)Q(y_{k+1}) < 0$ pour tout $k \in [\![1, n]\!]$. Alors la fonction Q change de signe, donc s'annule, entre y_k et y_{k+1}, pour tout $k \in [\![1, n]\!]$. Le polynôme Q s'annule donc n fois : il est scindé sur $\mathbb{R}$, à racines simples. $\lhd$

En choisissant y_k et y_{k+1} tels que $y_{k+1} - y_k \leqslant \varepsilon$, on montre que, pour $b_1, \ldots, b_n$ proche de $a_1, \ldots, a_n$, Q possède une racine z_k telle que $|x_k - z_k| \leqslant \varepsilon$. Ceci démontre que, pour $k \in [\![1, n]\!]$, la fonction, définie sur un voisinage de $(a_1, \ldots, a_n)$, qui à $(b_1, \ldots, b_n)$ associe la k-ième racine de Q (celles-ci étant rangées par ordre croissant) est continue.

Si cette démonstration ne s'applique plus, le théorème reste vrai dans $\mathbb{C}[X]$. On peut utiliser par exemple le discriminant (cf exercice 5.10) : celui-ci est une fonction polynomiale, donc continue, des coefficients de P ; si P n'a que des racines simples, son discriminant n'est pas nul et si les coefficients de Q sont proches de ceux de P, son discriminant reste non nul.

On ne saurait mieux terminer cette série d'exercices sur les polynômes scindés que par l'étude d'une famille de polynômes scindés sur $\mathbb{R}$, aux riches propriétés : les polynômes de Tchebychev.

5.41. Polynômes de Tchebychev

Soit $n \in \mathbb{N}^*$.

1. Montrer qu'il existe un unique polynôme $T_n \in \mathbb{R}[X]$ tel que, pour tout réel x, $T_n(\cos x) = \cos(nx)$. Calculer $T_{n+1} + T_{n-1}$.

Les polynômes T_n sont appelés *polynômes de Tchebychev de première espèce.*

2. Montrer que $T_n \in \mathbb{Z}[X]$, préciser son degré, son coefficient dominant. Déterminer les racines de T_n ainsi que les extrema de la fonction $x \longmapsto T_n(x)$ sur $[-1, 1]$.

3. Montrer que pour tout polynôme P à coefficients réels unitaire et de degré n, on a $\sup_{x\in[-1,1]} |P(x)| \geqslant \dfrac{1}{2^{n-1}}$, avec égalité si et seulement si $P = \dfrac{1}{2^{n-1}} T_n$.

(École normale supérieure)

▷ **Solution.**

1. L'existence de T_n peut s'obtenir à l'aide de la formule de Moivre et de celle du binôme de Newton. On obtient

$$\cos(nx) = \mathrm{Re}(\cos x + i \sin x)^n = \sum_{0\leqslant 2k\leqslant n} C_n^{2k} (-1)^k \cos^{n-2k} x \, (1-\cos^2 x)^k,$$

ce qui prouve l'existence et donne l'expression suivante de T_n

$$T_n(X) = \sum_{k=0}^{E(n/2)} C_n^{2k} (X^2-1)^k X^{n-2k}.$$

Pour l'unicité, il suffit de dire que si R_n est un second polynôme qui convient, on a $(T_n - R_n)(\cos x) = 0$ pour tout réel x. Il en résulte que $T_n - R_n$ est identiquement nul sur $[-1, 1]$. C'est donc le polynôme nul.

Les formules de trigonométrie montrent que, pour tout $x \in \mathbb{R}$,

$$\begin{aligned}(T_{n+1} + T_{n-1})(\cos x) &= \cos(n+1)x + \cos(n-1)x \\ &= 2\cos x \cos nx = 2\cos x T_n(\cos x).\end{aligned}$$

Donc le polynôme $T_{n+1} - 2XT_n + T_{n-1}$ est identiquement nul sur $[-1, 1]$. Il en résulte qu'il est nul et on a la relation de récurrence :

$$\boxed{T_{n+1} - 2XT_n + T_{n-1} = 0}.$$

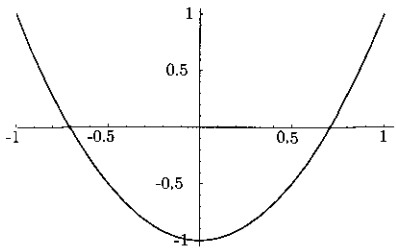
2. • On voit sur l'expression trouvé ci-dessus que $T_n \in \mathbb{Z}[X]$ et que $\deg(T_n) \leqslant n$. Comme le coefficient de X^n vaut $\sum_{k=0}^{E(n/2)} C_n^{2k} = 2^{n-1}$, on a $\deg(T_n) = n$.

On peut aussi montrer ces résultats par récurrence à l'aide de la relation de récurrence obtenue en 1.

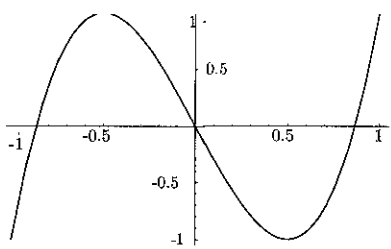
• Posons pour $k \in \mathbb{Z}$, $\theta_k = \frac{\pi}{2n} + \frac{k\pi}{n}$. On a évidemment $\cos(n\theta_k) = 0$ pour tout k. Il en résulte que $x_k = \cos\theta_k$ est une racine de T_n. Quand k décrit $\mathbb{Z}$, x_k prend n valeurs distinctes, obtenues en faisant varier k de 0 à $n-1$ par exemple. Comme $\deg(T_n) = n$, on dispose de toutes les racines de T_n, qui est donc scindé à racines simples sur $\mathbb{R}$. Notons que l'on a $-1 < x_{n-1} < \cdots < x_1 < x_0 < 1$.

• Comme $T_n(\cos\theta) = \cos n\theta$, on a $|T_n(z)| \leqslant 1$ pour tout $z \in [-1, 1]$. Il y a égalité si et seulement si z est de la forme $z_k = \cos\left(\frac{k\pi}{n}\right)$, avec $0 \leqslant k \leqslant n$. Plus précisément, $T_n(z_k) = 1$ si k est pair et $T_n(z_k) = -1$ si k est impair. Notons qu'en particulier, $T_n(1) = 1$ et $T_n(-1) = (-1)^n$.

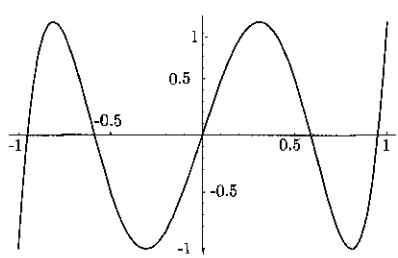
On voit que $-1 = z_n < x_{n-1} < z_{n-1} < \cdots < z_1 < x_0 < z_0 = 1$ et que la fonction T_n est strictement monotone entre deux z_k consécutifs. On a tracé les graphes sur $[-1, 1]$ de quelques polynômes.



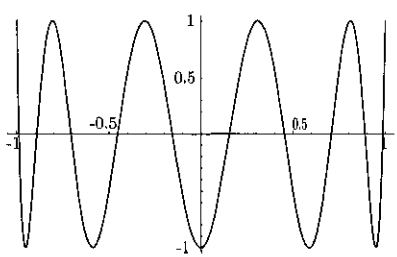
$n = 2$



$n = 3$



$n = 5$



$n = 10$

3. Pour tout polynôme P, on pose $\|P\| = \sup_{x \in [-1,1]} |P(x)|$. On considère

le polynome $P_n = \frac{1}{2^{n-1}} T_n$. Il est bien unitaire et de degré n. On a, d'après la question précédente, $\|P_n\| = 2^{1-n}$.

• Supposons qu'il existe un polynôme unitaire P de degré n tel que $\|P\| < \|P_n\|$. Étudions le polynôme $Q = P_n - P$. On a $\deg Q \leqslant n-1$ car P_n et P sont tous les deux unitaires de degré n. L'idée est de regarder les valeurs prises par Q aux points z_k (définis dans la question précédente). On a $Q(z_0) = P_n(z_0) - P(z_0) = \frac{1}{2^{n-1}} - P(z_0) > 0$ car $\|P\| < \|P_n\| = \frac{1}{2^{n-1}}$. De même $Q(z_1) < 0$, $Q(z_2) > 0$... Il résulte alors du théorème des valeurs intermédiaires que Q s'annule au moins une fois dans chaque intervalle $]z_{k+1}, z_k[$ pour $0 \leqslant k \leqslant n-1$. Ce qui donne n racines distinctes de Q. Comme Q est de degré $\leqslant n-1$, on a $Q = 0$, ce qui est absurde. On a donc $\|P\| \geqslant \|P_n\|$ pour tout polynôme P unitaire de degré n.

• On montre maintenant qu'il n'y a égalité dans $\|P\| \geqslant \|P_n\|$ que pour $P = P_n$. Reprenons le raisonnement précédent. Soit P unitaire de degré n tel que $\|P\| = \|P_n\|$ et $Q = P_n - P$. On a $\deg(Q) \leqslant n-1$ et $Q(z_0) \geqslant 0$, $Q(z_1) \leqslant 0$, $Q(z_2) \geqslant 0$... (on doit remplacer les inégalités strictes par des inégalités larges). Si toutes les inégalités sont strictes, Q est nul comme précédemment.

La situation est plus délicate si l'un des $Q(z_k)$ est nul. Si $1 \leqslant k \leqslant n-1$ et $Q(z_k) = 0$, alors $P(z_k) = \pm \frac{1}{2^{n-1}}$ et z_k est un extremum local de P. On a donc $P'(z_k) = 0$ et comme $P_n'(z_k) = 0$, on trouve $Q'(z_k) = 0$: z_k est une racine multiple de Q.

Montrons par récurrence sur $k \in [\![0, n-1]\!]$ que Q possède au moins k (resp. $k+1$) racines comptées avec leur multiplicité dans l'intervalle $[z_k, z_0] = [z_k, 1]$ si $Q(z_k) \neq 0$ (resp. $Q(z_k) = 0$). Il n'y a rien à prouver si $k = 0$. Supposons $k \geqslant 1$.

⋆ Si $Q(z_k) = 0$, z_k est une racine au moins double et comme il y en a au moins $k-1$ dans $[z_{k-1}, 1]$ (par hypothèse de récurrence), on en trouve donc au moins $k+1$ dans $[z_k, 1]$.

⋆ Si $Q(z_k) \neq 0$ et $Q(z_{k-1}) = 0$, l'hypothèse de récurrence assure l'existence de k zéros dans $[z_{k-1}, 1] \subset [z_k, 1]$.

⋆ Si $Q(z_k) \neq 0$ et $Q(z_{k-1}) \neq 0$, on a $Q(z_k)Q(z_{k-1}) < 0$ et Q s'annule sur $]z_k, z_{k-1}[$. Dans ce cas encore, on a donc au moins k racines dans $[z_k, 1]$.

La récurrence est achevée. Si $Q(z_{n-1}) = 0$, Q a donc au moins n racines dans $[-1, 1]$. Dans le cas contraire, $Q(-1)Q(z_{n-1}) \leqslant 0$ et Q s'annule sur $[-1, z_{n-1}[$. Comme on a déjà $n-1$ racines dans $[z_{n-1}, 1]$, on obtient dans ce cas aussi n racines dans $[-1, 1]$. Le polynôme Q est donc nul ce qui donne le résultat. ◁

La démonstration faite dans cet exercice peut se généraliser. On dit

qu'une fonction continue g équi-oscille sur $n+1$ points de $[a,b]$, s'il existe des points $z_0 < z_1 < \cdots < z_n$ de $[a,b]$ tels que, pour tout $i \in [\![0,n]\!]$, $|g(z_i)| = \|g\| = \sup_{t\in[a,b]} |g(t)|$ et $g(z_{i+1}) = -g(z_i)$ si $i < n$. On a vu que le polynôme T_n équi-oscille sur $n+1$ points de $[-1,1]$. On peut démontrer le théorème suivant : pour $f \in \mathcal{C}([a,b],\mathbb{R})$, il existe un polynôme unique (dit de meilleure approximation) qui réalise le minimum de distance de f à $\mathbb{R}_{n-1}[\mathrm{X}]$ au sens de la norme uniforme. Cet unique polynôme q est caractérisé par le fait que la fonction $f - q$ équi-oscille sur au moins $n+1$ points de $[a,b]$. L'exercice a montré que si on prend $[a,b] = [-1,1]$, alors $q = \frac{\mathrm{T}_n}{2^{n-1}} - \mathrm{X}^n$ réalise le minimum de distance de X^n à $\mathbb{R}_{n-1}[\mathrm{X}]$.

On peut démontrer, comme dans la première question de l'exercice précédent, que, pour $n \in \mathbb{N}$, il existe $\mathrm{U}_n \in \mathbb{R}[\mathrm{X}]$ de degré n tel que, pour tout $\theta \in \mathbb{R}$, $\sin(n+1)\theta = \sin\theta \mathrm{U}_n(\cos\theta)$. Le polynôme U_n est appelé n-ième polynôme de Tchebychev de seconde espèce. En dérivant la relation de définition de T_n, on obtient $\mathrm{T}'_n = n\mathrm{U}_{n-1}$.

La résolution de l'exercice suivant nécessite la connaissance des polynômes de Tchebychev.

5.42. Inégalités de Bernstein et de Markov

1. *Inégalité de Schur (1919).* Soit $\mathrm{P} \in \mathbb{R}_{n-1}[\mathrm{X}]$. Montrer que

$$[\forall x \in [-1,1],\ \sqrt{1-x^2}|\mathrm{P}(x)| \leqslant 1] \Longrightarrow [\forall x\ \in [-1,1],\ |\mathrm{P}(x)| \leqslant n].$$

(Indication : $x_0, x_1, \ldots, x_{n-1}$ étant les racines du n-ième polynôme de Tchebychev T_n, décomposer P dans la base des polynômes interpolateurs relatifs aux x_i.)

2. *Inégalité de Bernstein (1912).* Soit

$$\mathrm{Q}(\theta) = \sum_{k=0}^{n} (\lambda_k \cos k\theta + \mu_k \sin k\theta),$$

où les λ_k et μ_k sont réels. On suppose que $\sup_{\theta\in\mathbb{R}} |\mathrm{Q}(\theta)| \leqslant 1$. Montrer que $\sup_{\theta\in\mathbb{R}} |\mathrm{Q}'(\theta)| \leqslant n$.

3. *Inégalité de Markov (1889).* Soit $\mathrm{P} \in \mathbb{R}_n[\mathrm{X}]$. On suppose que $\sup_{|x|\leqslant 1} |\mathrm{P}(x)| \leqslant 1$. Montrer que $\sup_{|x|\leqslant 1} |\mathrm{P}'(x)| \leqslant n^2$.

(École normale supérieure)

▷ **Solution.**

1. Le n-ième polynôme de Tchebychev est de degré n et ses racines sont les $x_i = \cos\left(\dfrac{(2i+1)\pi}{2n}\right)$, où $i \in [\![0, n-1]\!]$. Notons $\theta_i = \dfrac{(2i+1)\pi}{2n}$. Les polynômes interpolateurs de base pour le n-uplet $(x_0, \dots, x_{n-1})$ sont (cf. les rappels précédant l'exercice 5.25) les polynômes

$$\mathrm{L}_i(\mathrm{X}) = \frac{\Phi(\mathrm{X})}{\Phi'(x_i)(\mathrm{X} - x_i)} \ (0 \leqslant i \leqslant n-1), \text{ où } \Phi = \prod_{i=0}^{n-1}(\mathrm{X} - x_i).$$

On a donc

$$\mathrm{P} = \sum_{i=0}^{n-1} \mathrm{P}(x_i)\frac{\Phi}{\Phi'(x_i)(\mathrm{X} - x_i)} = \sum_{i=0}^{n-1} \mathrm{P}(x_i)\frac{\mathrm{T}_n}{\mathrm{T}_n'(x_i)(\mathrm{X} - x_i)},$$

puisque T_n est proportionnel à Φ. On va calculer les valeurs $\mathrm{T}_n'(x_i)$.

Par définition de T_n, on a, pour tout θ, $\mathrm{T}_n(\cos\theta) = \cos n\theta$ et donc, en dérivant, $-\sin\theta \mathrm{T}_n'(\cos\theta) = -n\sin n\theta$. On en déduit $\mathrm{T}_n'(x_i) = n\dfrac{\sin n\theta_i}{\sin\theta_i}$. On obtient

$$\sin n\theta_i = \sin\left((2i+1)\frac{\pi}{2}\right) = (-1)^i \quad \text{et} \quad \sin\theta_i = \sqrt{1 - x_i^2},$$

car θ_i est dans $]0, \pi[$, et donc $\mathrm{T}_n'(x_i) = \dfrac{(-1)^i n}{\sqrt{1 - x_i^2}}$, pour tout i de $[\![0, n-1]\!]$. On obtient finalement

$$\mathrm{P} = \mathrm{T}_n \sum_{i=0}^{n-1} \frac{\mathrm{P}(x_i)}{\mathrm{T}_n'(x_i)(\mathrm{X} - x_i)} = \frac{\mathrm{T}_n}{n} \sum_{i=0}^{n-1} (-1)^i \sqrt{1 - x_i^2}\frac{\mathrm{P}(x_i)}{\mathrm{X} - x_i}.$$

De l'égalité précédente, on déduit que, pour tout $x \in [-1, 1] \backslash \{x_0, \dots, x_{n-1}\}$, on a

$$|\mathrm{P}(x)| \leqslant \frac{|\mathrm{T}_n(x)|}{n} \sum_{i=0}^{n-1} \sqrt{1 - x_i^2}\frac{|\mathrm{P}(x_i)|}{|x - x_i|} \leqslant \frac{|\mathrm{T}_n(x)|}{n} \sum_{i=0}^{n-1} \frac{1}{|x - x_i|}.$$

Si $x \in [-1, x_{n-1}[\cup]x_0, 1]$, alors tous les $x - x_i$ ont même signe. On peut donc écrire

$$|\mathrm{P}(x)| \leqslant \frac{|\mathrm{T}_n(x)|}{n} \left| \sum_{i=0}^{n-1} \frac{1}{x - x_i} \right|.$$

Mais $x_0, \dots, x_{n-1}$ étant les n racines de T_n, on reconnaît dans la somme ci-dessus $\dfrac{\mathrm{T}_n'(x)}{\mathrm{T}_n(x)}$. On a donc $|\mathrm{P}(x)| \leqslant \dfrac{1}{n}|\mathrm{T}_n'(x)|$. Or on a vu que, pour

tout $\theta \in \mathbb{R}$ tel que $\sin\theta \neq 0$, on a $\mathrm{T}'_n(\cos\theta) = \dfrac{n\sin n\theta}{\sin\theta}$. De l'inégalité, valable pour tout $\theta \in \mathbb{R}$, $|\sin n\theta| \leqslant n|\sin\theta|$ (qui se montre facilement par récurrence sur n), on déduit que $|\mathrm{T}'_n(\cos\theta)| \leqslant n^2$. On obtient enfin, pour tout $x \in [-1, x_{n-1}[\cup]x_0, 1]$,

$$|\mathrm{P}(x)| \leqslant \frac{1}{n}|\mathrm{T}'_n(x)| \leqslant n,$$

ce qui est l'inégalité voulue.

Reste à examiner le cas où $x \in [x_{n-1}, x_0] = \left[-\cos\dfrac{\pi}{2n}, \cos\dfrac{\pi}{2n}\right]$. On a alors

$$|\mathrm{P}(x)| \leqslant \frac{1}{\sqrt{1-x^2}} \leqslant \frac{1}{\sin\dfrac{\pi}{2n}}.$$

L'inégalité $|\mathrm{P}(x)| \leqslant n$ est vérifiée si on a $\sin\dfrac{\pi}{2n} \geqslant \dfrac{1}{n}$. C'est vrai car on a, plus généralement, pour tout $x \in \left[0, \dfrac{\pi}{2}\right]$, $\sin x \geqslant \dfrac{2}{\pi}x$; cette inégalité résulte de la concavité de la fonction sinus sur $\left[0, \dfrac{\pi}{2}\right]$.

2. On va appliquer la question 1 à un polynôme P approprié. Soit θ un réel. Pour majorer $|\mathrm{Q}'(\theta)|$, on utilise la définition de la dérivée comme limite du taux d'accroissement, mais on considère $\mathrm{Q}(\theta+h) - \mathrm{Q}(\theta-h)$, plutôt que $\mathrm{Q}(\theta+h) - \mathrm{Q}(\theta)$, pour des raisons de symétrie. Les formules de trigonométrie usuelles donnent, pour $h \in \mathbb{R}$,

$$\begin{aligned}\mathrm{Q}(\theta+h) - \mathrm{Q}(\theta-h) &= \sum_{k=0}^{n} \lambda_k(\cos(k(\theta+h)) - \cos(k(\theta-h)) \\ &\quad +\mu_k(\sin(k(\theta+h)) - \sin(k(\theta-h)) \\ &= 2\sum_{k=1}^{n}(-\lambda_k\sin k\theta + \mu_k\cos k\theta)\sin kh.\end{aligned}$$

On sait que, pour $k \in \mathbb{N}^*$ et $h \in \mathbb{R}$, on a $\sin kh = \sin h\,\mathrm{U}_{k-1}(\cos h)$, où U_{k-1} est le $(k-1)$-ième polynôme de Tchebychev de seconde espèce ; il est de degré $k-1$. On a donc, θ étant fixé, pour tout réel h,

$$\begin{aligned}\mathrm{Q}(\theta+h) - \mathrm{Q}(\theta-h) &= 2\sin h\left(\sum_{k=1}^{n}(-\lambda_k\sin k\theta + \mu_k\cos k\theta)\,\mathrm{U}_{k-1}(\cos h)\right) \\ &= 2\sin h\,\mathrm{P}(\cos h),\end{aligned}$$

où P appartient à $\mathbb{R}_{n-1}[\mathrm{X}]$.

Montrons que P vérifie l'hypothèse de la question 1.

Pour tout $x \in [-1, 1]$, il existe $h \in [0, \pi]$ tel que $x = \cos h$. On a alors

$$\sqrt{1-x^2}|\mathrm{P}(x)| = |\sin h\,\mathrm{P}(\cos h)| = \frac{1}{2}|\mathrm{Q}(\theta+h) - \mathrm{Q}(\theta-h)| \leqslant 1,$$

d'après l'hypothèse sur Q. Le polynôme vérifie donc l'hypothèse de la question 1. On en déduit que, pour tout $x \in [-1,1]$, on a $|P(x)| \leqslant n$. En particulier, pour tout $h \in \mathbb{R}$ tel que $\sin h \neq 0$, on a

$$\left|\frac{Q(\theta+h)-Q(\theta-h)}{2\sin h}\right| = |P(\cos h)| \leqslant n.$$

Sachant que

$$Q'(\theta) = \lim_{h\to 0} \frac{Q(\theta+h)-Q(\theta-h)}{2h} = \lim_{h\to 0} \frac{Q(\theta+h)-Q(\theta-h)}{2\sin h},$$

on en déduit, en faisant tendre h vers 0, que $|Q'(\theta)| \leqslant n$. Ceci étant vrai pour tout réel θ, on a l'inégalité voulue.

3. On utilise la question précédente, en posant $Q(\theta) = P(\cos\theta)$ pour tout $\theta \in \mathbb{R}$. Montrons que Q a la même forme que dans la question précédente. Cela résulte de la linéarisation de $\cos^n\theta$, pour $n \in \mathbb{N}$, à l'aide de la formule du binôme et de la formule de Moivre. On a, pour tout $n \in \mathbb{N}$,

$$\cos^n\theta = \left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^n = \frac{1}{2^n}\left(\sum_{k=0}^{n} C_n^k e^{i(n-2k)\theta}\right).$$

En regroupant les termes conjugués, on obtient

$$\cos^n\theta = \begin{cases} \dfrac{1}{2^{n-1}} \displaystyle\sum_{k=0}^{\frac{n-1}{2}} C_n^k \cos((n-2k)\theta) & \text{si } n \text{ est impair,} \\ \dfrac{1}{2^{n-1}} \displaystyle\sum_{k=0}^{\frac{n}{2}-1} C_n^k \cos((n-2k)\theta) + \dfrac{1}{2^n} & \text{si } n \text{ est pair.} \end{cases}$$

Chaque $\cos^k\theta$, pour $0 \leqslant k \leqslant n$, étant une combinaison linéaire à coefficients réels de $1, \cos\theta, \cos 2\theta, \ldots, \cos k\theta$, $P(\cos\theta)$ sera une combinaison linéaire de $1, \cos\theta, \cos 2\theta, \ldots, \cos n\theta$. Il existe donc des constantes réelles $\lambda_0, \lambda_1, \ldots, \lambda_n$ telles que, pour tout $\theta \in \mathbb{R}$, $Q(\theta) = \displaystyle\sum_{k=0}^{n} \lambda_k \cos k\theta$ et ainsi Q a la forme voulue.

On a, par hypothèse, pour tout $\theta \in \mathbb{R}$, $|Q(\theta)| = |P(\cos\theta)| \leqslant 1$. On en déduit, d'après la question 2 que, pour tout réel θ, $|Q'(\theta)| \leqslant n$. De l'égalité $Q'(\theta) = -\sin\theta P'(\cos\theta)$, on tire $|\sin\theta P'(\cos\theta)| \leqslant n$. En posant $x = \cos\theta$, on obtient alors, pour tout $x \in [-1,1]$,

$$|\sqrt{1-x^2}P'(x)| \leqslant n.$$

La fonction $\frac{1}{n}P'$ vérifie les hypothèses de la question 1. On en déduit que pour tout $x \in [-1,1]$, on a $\frac{1}{n}|P'(x)| \leqslant n$ et donc $|P'(x)| \leqslant n^2$. ◁

Par homogénéité, on en déduit pour tout polynôme trigonométrique Q *de degré inférieur ou égal à* n, $\sup_{\theta\in\mathbb{R}}|Q'(\theta)| \leqslant n \sup_{\theta\in\mathbb{R}}|Q(\theta)|$, *c'est-à-dire* $\|Q'\|_\infty \leqslant n\|Q\|_\infty$. *On a de même, pour* P *dans* $\mathbb{R}_n[X]$, $\|P'\| \leqslant n^2\|P\|$, *où* $\|P\| = \sup_{|x|\leqslant 1}|P(x)|$.

Les exercices qui suivent traitent du problème de la localisation des racines. Le premier démontre le très classique théorème de Gauss-Lucas et en donne quelques applications posées lors de différents oraux.

5.43. Théorème de Gauss-Lucas

1. Soit $P \in \mathbb{C}[X]$ non constant. Montrer que les racines de P' sont dans l'enveloppe convexe des racines de P.

2. Déterminer le plus grand entier $n \geqslant 2$ tel que les racines non nulles de $(X+1)^n - X^n - 1$ soient de module 1.

3. Soit P un polynôme non constant de $\mathbb{C}[X]$, Δ une droite du plan complexe, H_1 et H_2 les deux demi-plans ouverts limités par Δ. On suppose que P' a une racine dans H_1. Montrer que $P(H_1) = \mathbb{C}$.

(École normale supérieure)

▷ **Solution.**

1. On écrit $P = \lambda \prod_{k=1}^{r}(X-\lambda_k)^{n_k}$ où $r \geqslant 1$, $\lambda \in \mathbb{C}^*$, $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ sont les racines deux à deux distinctes de P et $n_1, \ldots, n_r$ leur ordre respectif. Comme

$$P' = \lambda \sum_{k=1}^{r} n_k (X-\lambda_k)^{n_k-1} \prod_{l\neq k}(X-\lambda_l)^{n_l},$$

on a classiquement la décomposition en éléments simples de P'/P :

$$\frac{P'}{P} = \sum_{k=1}^{r} \frac{n_k}{X-\lambda_k}.$$

Prenons une racine z de P'. Si z est l'une des racines λ_k, elle est évidemment dans l'enveloppe convexe des racines de P. Sinon, on peut écrire

$$0 = \frac{P'(z)}{P(z)} = \sum_{k=1}^{r} \frac{n_k}{z-\lambda_k} = \sum_{k=1}^{r} n_k \frac{\overline{z-\lambda_k}}{|z-\lambda_k|^2},$$

ce qui donne en passant au conjugué $\sum_{k=1}^{r} n_k \dfrac{z-\lambda_k}{|z-\lambda_k|^2} = 0$ puis

$$z = \frac{\sum_{k=1}^{r} \frac{n_k}{|z-\lambda_k|^2}\lambda_k}{\sum_{k=1}^{r} \frac{n_k}{|z-\lambda_k|^2}}.$$

Comme chaque $\frac{n_k}{|z-\lambda_k|^2}$ est strictement positif, cette formule exprime le fait que z est un barycentre à coefficients strictement positifs des λ_k. La racine z de P' est donc dans l'enveloppe convexe des racines de P.

2. Si $n = 2$, P a une seule racine 0. Soit $n \geqslant 3$ et $P = (X+1)^n - X^n - 1$. On a alors $P' = n(X+1)^{n-1} - nX^{n-1}$. Si z une racine de P', on remarque que z ne peut être nul et donc

$$\left(\frac{z+1}{z}\right)^{n-1} = 1.$$

Il existe donc $k \in [\![0, n-2]\!]$ tel que $\frac{z+1}{z} = e^{\frac{2ik\pi}{n-1}}$. En fait, $k \neq 0$ puisque $z+1 \neq z$. On en déduit que les $n-2$ racines de P' sont les nombres complexes

$$z_k = \frac{e^{\frac{-ik\pi}{n-1}}}{2i\sin\frac{k\pi}{n-1}} \quad (1 \leqslant k \leqslant n-2).$$

Si toutes les racines de P sont de module 1, les racines de P' sont nécessairement dans le disque unité, puisque ce sont des barycentres à coefficients positifs des racines de P. Le module de z_1 est $\frac{1}{2\sin\frac{\pi}{n-1}} > 0$ et pour $n \geqslant 8$, on a

$$2\sin\frac{\pi}{n-1} < 2\sin\frac{\pi}{6} = 1 \quad \text{et} \quad |z_1| > 1,$$

donc un entier $n \geqslant 8$ ne convient pas.

Regardons $P_0 = (X+1)^7 - X^7 - 1$. Visiblement, -1 et 0 sont racines de P_0 qui s'écrit

$$P_0 = X(X+1)(7X^4 + 14X^3 + 21X^2 + 14X + 7).$$

Le polynôme $Q = 7X^4 + 14X^3 + 21X^2 + 14X + 7$, qui est un polynôme réciproque, peut se mettre sous la forme $X^2R\left(X + \frac{1}{X}\right)$ où $R \in \mathbb{R}[X]$. En effet, en posant $Y = X + \frac{1}{X}$, on a

$$\begin{aligned} Q &= X^2\left(7\left(X^2 + \frac{1}{X^2}\right) + 14\left(X + \frac{1}{X}\right) + 21\right) \\ &= X^2\left(7Y^2 - 14 - 14Y + 21\right) \\ &= 7X^2(Y-1)^2. \end{aligned}$$

Un racine z de P_0 distincte de 0 et -1 doit vérifier $z + \dfrac{1}{z} - 1 = 0$, ce qui équivaut à $z^2 - z + 1 = 0$ et donc à $z = -j$ ou $z = -j^2$. L'ensemble des racines de P_0 est donc $\{0, -1, -j, -j^2\}$ qui est contenu dans le disque unité.

Conclusion. Le plus grand entier n pour lequel $(X+1)^n - X^n - 1$ a toutes ses racines non nulles de module 1 est $n_0 = 7$.

3. Supposons que $P(H_1) \neq \mathbb{C}$. Considérons $z \in \mathbb{C}$ tel que $z \notin P(H_1)$. Les racines du polynôme $P - z$ appartiennent donc à $\mathbb{C} \setminus H_1 = \overline{H_2}$, adhérence de H_2. Comme $\overline{H_2}$ est convexe, le résultat rappelé ci-dessus montre que les racines de P' sont encore dans $\overline{H_2}$. Cela contredit l'existence d'une racine de P' dans H_1. On a donc $P(H_1) = \mathbb{C}$. ◁

On pourra encore utiliser le théorème de Gauss-Lucas dans la seconde question de l'exercice suivant.

5.44. Polynôme à racines réelles

Soit $P \in \mathbb{C}[X]$. On suppose que P admet deux racines réelles distinctes et que P'' divise P.

1. Montrer que les racines de P sont toutes simples.

2. Montrer que les racines de P sont toutes réelles.

(École normale supérieure)

▷ **Solution.**

1. Notons $n \geqslant 2$ le degré de P. On peut supposer P unitaire sans perte de généralité et l'écrire sous la forme

$$P(X) = \prod_{i=1}^{p} (X - z_i) \prod_{j=1}^{q} (X - w_j)^{\alpha_j}$$

où les nombres complexes $z_1, \ldots, z_p, w_1, \ldots, w_q$ sont deux à deux distincts et où les exposants α_j sont tous supérieurs ou égaux à 2 (on a simplement factorisé P en isolant les racines simples). Le but est de montrer que $q = 0$. Chaque w_j est encore racine de P'' avec un ordre de multiplicité $\alpha_j - 2$. On peut donc écrire

$$P''(X) = R(X) \prod_{j=1}^{q} (X - w_j)^{\alpha_j - 2},$$

où R est un polynôme qui, par hypothèse, doit diviser $\prod_{i=1}^{p} (X - z_i)$.

Regardons ce que cela donne sur les degrés. On a $n = p + \sum_{j=1}^{q} \alpha_j$ et

$$n - 2 = \deg \mathrm{P}'' = r + \sum_{j=1}^{q} (\alpha_j - 2) = r + n - p - 2q,$$

où $r = \deg \mathrm{R} \leqslant p$. On en déduit que $(p - r) + 2q = 2$ ce qui laisse peu de possibilités (car $p - r \geqslant 0$). On veut montrer que $q = 0$ de sorte qu'il suffit d'exclure la solution $q = 1$, $p = r$. On aurait dans ce cas,

$$\begin{cases} \mathrm{P} = (\mathrm{X} - z_1) \dots (\mathrm{X} - z_p)(\mathrm{X} - w_1)^{\alpha_1} \\ \mathrm{P}'' = n(n-1)(\mathrm{X} - z_1) \dots (\mathrm{X} - z_p)(\mathrm{X} - w_1)^{\alpha_1 - 2}. \end{cases}$$

Il faut noter que cela serait parfaitement possible avec $p = 0$. Mais par hypothèse, P admet au moins deux racines distinctes et on a donc $p \geqslant 1$. Posons $\mathrm{S} = (\mathrm{X} - z_1) \dots (\mathrm{X} - z_p)$. On a $\mathrm{P} = (\mathrm{X} - w_1)^{\alpha_1} \mathrm{S}$ et en dérivant cela deux fois on obtient

$$\mathrm{P}'' = (\mathrm{X} - w_1)^{\alpha_1 - 2} [\alpha_1(\alpha_1 - 1)\mathrm{S} + 2\alpha_1(\mathrm{X} - w_1)\mathrm{S}' + (\mathrm{X} - w_1)^2 \mathrm{S}''].$$

Le polynôme entre crochets doit être égal à $\mathrm{R} = n(n-1)\mathrm{S}$. Mais sachant que $1 \leqslant \alpha_1 < n$ (car p n'est pas nul), on obtient une contradiction en évaluant en w_1, car on obtient alors que w_1 est racine de S.

Conclusion. Le cas $q = 1$ est impossible : on a $q = 0$ et $p = n$ de sorte que P est à racines simples dans $\mathbb{C}$.

2. Nous allons utiliser ici le théorème de Gauss-Lucas (voir l'exercice précédent) : si $\mathrm{Q} \in \mathbb{C}[\mathrm{X}]$, les racines de Q′ qui ne sont pas racines de Q sont des barycentres à coefficients strictement positifs des racines de Q ; en particulier, elles se trouvent dans l'enveloppe convexe de l'ensemble des racines de Q. Nous allons appliquer cela à P et P′. D'après la question précédente le polynôme P admet n racines simples $z_1, \dots, z_n$ et le polynôme P″ qui divise P admet donc $n - 2$ racines simples à prendre parmi les z_i. L'enveloppe convexe de $z_1, \dots, z_n$ est un polygone convexe K du plan complexe. Quitte à renuméroter les z_i on peut supposer que $z_1, \dots, z_k$ sont les sommets de ce polygone. Imaginons que $k \geqslant 3$ c'est-à-dire que K n'est pas un segment. Les racines de P′ sont alors nécessairement dans l'intérieur de K d'après le résultat que nous venons de rappeler. Et une nouvelle application de ce théorème montre qu'il en est de même des racines de P″. Mais il y a au plus $n - 3$ racines de P dans l'intérieur de K ce qui donne une contradiction. On a donc $k = 2$, c'est-à-dire que K est un segment. Comme par hypothèse P a deux racines réelles distinctes, le segment K est inclus dans $\mathbb{R}$ et on peut donc affirmer que toutes les racines de P sont réelles. ◁

5.45. Théorème d'Eneström-Kakeya

Soit $P = a_0 + a_1X + \cdots + a_nX^n \in \mathbb{R}[X]$ avec $a_k > 0$ pour tout k dans $[\![0, n]\!]$.

1. Montrer que si $a_0 \geqslant a_1 \geqslant \cdots \geqslant a_n$, alors les racines complexes de P sont toutes de module $\geqslant 1$.

2. Montrer, dans le cas général, que les racines complexes de P sont situées dans la couronne

$$\left\{ z \in \mathbb{C},\ \min_{0 \leqslant k \leqslant n-1} \frac{a_k}{a_{k+1}} \leqslant |z| \leqslant \max_{0 \leqslant k \leqslant n-1} \frac{a_k}{a_{k+1}} \right\}.$$

(École polytechnique)

▷ **Solution.**

1. Soit z une racine de P. On a

$$(1-z)P(z) = 0 = -a_n z^{n+1} + (a_n - a_{n-1})z^n + \cdots + (a_1 - a_0)z + a_0,$$

de sorte que $a_0 = (a_0 - a_1)z + \cdots + (a_{n-1} - a_n)z^n + a_n z^{n+1}$. Supposons par l'absurde que $|z| < 1$. En passant au module dans cette égalité et en majorant par inégalité triangulaire, on obtient

$$a_0 = |a_0| < (a_0 - a_1) + (a_1 - a_2) + \cdots + (a_{n-1} - a_n) + a_n = a_0.$$

Contradiction.

2. Posons $r = \min_{0 \leqslant k \leqslant n-1} \frac{a_k}{a_{k+1}}$ et $R = \max_{0 \leqslant k \leqslant n-1} \frac{a_k}{a_{k+1}}$ et considérons le polynôme Q défini par

$$Q(X) = P(rX) = a_0 + ra_1X + \cdots + r^n a_n X^n.$$

Alors Q satisfait les conditions de la question précédente. Si z est une racine de P, $\frac{z}{r}$ est une racine de Q. On a donc $|\frac{z}{r}| \geqslant 1$, soit $|z| \geqslant r$.

Pour la minoration, considérons

$$\widetilde{P} = X^n P\left(\frac{1}{X}\right) = a_0X^n + a_1X^{n-1} + \cdots + a_n.$$

Comme $\frac{a_{k+1}}{a_k} \geqslant \frac{1}{R}$, pour $0 \leqslant k \leqslant n-1$, d'après ce qu'on vient de démontrer, toute racine de $\widetilde{P}$ est de module $\geqslant \frac{1}{R}$. Si z est une racine de P, on a $z \neq 0$ puisque $a_0 > 0$ et $P(z) = z^n \widetilde{P}\left(\frac{1}{z}\right) = 0$. Ainsi, $\frac{1}{z}$ est une racine de $\widetilde{P}$, $\frac{1}{|z|} \geqslant \frac{1}{R}$ et finalement, $|z| \leqslant R$. ◁

5.46. Construction d'un polynôme satisfaisant des conditions sur le module de ses valeurs

Soit $P \in \mathbb{C}[X]$ non nul ayant une racine z_0 telle que $|z_0| < 1$. Montrer qu'il existe $Q \in \mathbb{C}[X]$ de degré inférieur ou égal à celui de P, tel que pour tout z de module 1 on ait $|Q(z)| = |P(z)|$ et pour tout z de module strictement inférieur à 1 on ait $|Q(z)| > |P(z)|$.

(École polytechnique)

▷ **Solution.**

On peut supposer P unitaire.

• Commençons par étudier le cas où $\deg P = 1$. On a alors $P = X - z_0$. Montrons que le polynôme $Q = 1 - \overline{z_0}X$ convient. L'hypothèse sur le degré est remplie et on a pour tout $z \in \mathbb{C}$,

$$\begin{aligned}|Q(z)|^2 - |P(z)|^2 &= (1 + |z|^2|z_0|^2 - z\overline{z_0} - \overline{z}z_0) - (|z|^2 + |z_0|^2 - z\overline{z_0} - \overline{z}z_0)\\ &= (1 - |z|^2)(1 - |z_0|^2).\end{aligned}$$

Le résultat en découle, car $1 - |z_0|^2 > 0$.

• Passons au cas général. Notons $z_0, z_1, \ldots, z_k$ les racines de P qui sont de module strictement inférieur à 1, comptées avec leur ordre de multiplicité. On peut écrire

$$P(X) = (X - z_0)\cdots(X - z_k)R(X),$$

où R est un polynôme qui ne s'annule pas dans le disque unité ouvert. Posons

$$Q(X) = (1 - \overline{z_0}X)\cdots(1 - \overline{z_k}X)R(X).$$

Il résulte du point précédent que si $|z| = 1$, alors $|Q(z)| = |P(z)|$ et que si $|z| < 1$, alors $|Q(z)| > |P(z)|$, car $|R(z)| > 0$. ◁

5.47. Inégalité de Landau

Pour $P = \sum_{i=0}^{n} a_i X^i \in \mathbb{C}[X]$, on pose $\|P\| = \left(\sum_{i=0}^{n} |a_i|^2\right)^{\frac{1}{2}}$.

1. Soit $R \in \mathbb{C}[X]$ et $z \in \mathbb{C}^*$. Montrer que

$$\|(1 - \overline{z}X)R\| = \|(X - z)R\|.$$

2. Soit $P \in \mathbb{C}[X]$ unitaire et de degré n. On note $M(P)$ le module du produit des racines hors du disque unité de $\mathbb{C}$ ($M(P) = 1$ s'il n'y en a pas). Montrer que $M(P) \leqslant \|P\|$.

(École polytechnique)

Solution.

1. Si $P \in \mathbb{C}[X]$, on a $\|P\|^2 = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} P(e^{i\theta})\overline{P(e^{i\theta})}d\theta$: c'est la formule de Parseval appliquée à la fonction 2π-périodique $\theta \longmapsto P(e^{i\theta})$. Cette constatation conduit à une preuve particulièrement courte de l'égalité demandée. En effet, on a

$$\|(X - z)R\|^2 = \frac{1}{2\pi}\int_0^{2\pi} (e^{i\theta} - z)R(e^{i\theta})(e^{-i\theta} - \overline{z})\overline{R(e^{i\theta})}d\theta$$

et

$$\|(1 - \overline{z}X)R\|^2 = \frac{1}{2\pi}\int_0^{2\pi} (1 - \overline{z}e^{i\theta})R(e^{i\theta})(1 - ze^{-i\theta})\overline{R(e^{i\theta})}d\theta$$

Il suffit de constater que $(e^{i\theta} - z)(e^{-i\theta} - \overline{z}) = (1 - \overline{z}e^{i\theta})(1 - ze^{-i\theta})$ pour tout θ. Notons qu'un calcul direct des quantités $\|(1 - \overline{z}X)R\|$ et $\|(X - z)R\|$ permet de conclure également.

2. On suppose que P possède des racines de module strictement supérieur à 1 : sinon, $M(P) = 1$ et l'inégalité est vérifiée car $\|P\| \geqslant 1$ puisque P est unitaire. Notons $z_1, \ldots, z_r$ les racines de P de module inférieur ou égal à 1 et $z_{r+1}, \ldots, z_n$ les racines de module strictement supérieur à 1. On a donc $M(P) = \displaystyle\prod_{i=r+1}^{n} |z_i|$.

La question précédente montre que si on remplace dans P un facteur $(X - z_i)$ par $(1 - \overline{z_i}X)$ la norme ne change pas. On va effectuer cette substitution pour toutes les racines $z_1, \ldots, z_r$. Posons

$$Q = \prod_{i=1}^{r}(1 - \overline{z_i}X)\prod_{j=r+1}^{n}(X - z_j).$$

On a $\|P\| = \|Q\|$. Mais le coefficient constant de Q est $\displaystyle\prod_{j=r+1}^{n} z_j$. On a donc $M(P) \leqslant \|Q\| = \|P\|$. ◁

5.48. Critère de Routh-Hurwitz pour le degré 3

Soit $P = X^3 + aX^2 + bX + c \in \mathbb{R}[X]$. Montrer que, pour que toutes les racines complexes de P aient une partie réelle strictement négative il faut et il suffit que $a > 0$, $b > 0$, $c > 0$ et $ab - c > 0$.

(**École normale supérieure**)

▷ **Solution.**

• P étant de degré impair, il possède au moins une racine réelle. Il existe donc $(\alpha, \beta, \gamma) \in \mathbb{R}^3$ tel que $P = (X+\alpha)(X^2+\beta X+\gamma)$. Les racines de P ont toutes une partie réelle strictement négative si $\alpha > 0$ et si les parties réelles des racines x_1 et x_2 du polynôme $Q = X^2+\beta X+\gamma$ sont strictement négatives.

Si le polynôme Q a deux racines réelles, c'est-à-dire si $\beta^2 \geqslant 4\gamma$, on doit avoir $x_1x_2 = \gamma > 0$ et $x_1 + x_2 = -\beta < 0$, soit $\beta > 0$. S'il possède deux racines complexes conjuguées, c'est-à-dire si $\beta^2 < 4\gamma$, on a alors $\mathrm{Re}(x_1) = \mathrm{Re}(x_2) = \dfrac{1}{2}(x_1+x_2) = \dfrac{-\beta}{2}$. Il faut donc $\beta > 0$; on remarque qu'on a, de plus, $\gamma > \dfrac{1}{4}\beta^2 > 0$. Dans tous les cas, il faut $\beta > 0$ et $\gamma > 0$.

Réciproquement, si les conditions $\beta > 0$ et $\gamma > 0$ sont réalisées, soit Q possède deux racines réelles et alors elles sont toutes deux négatives d'après ce qui précède, soit Q possède deux racines complexes conjuguées et leur partie réelle est négative puisque $\beta > 0$.

Les racines de P ont donc une partie réelle strictement négative si et seulement si $\alpha > 0, \beta > 0, \gamma > 0$.

• On remarque alors que $a = \alpha+\beta$, $b = \alpha\beta+\gamma$, $c = \alpha\gamma$. On en déduit que si $\alpha > 0$, $\beta > 0$ et $\gamma > 0$, alors $a > 0$, $b > 0$ et $c > 0$. Et on a de plus $ab - c = \beta(\alpha^2 + \alpha\beta + \gamma) > 0$.

Réciproquement, si on a $a > 0, b > 0, c > 0, ab - c > 0$, on obtient

$$\alpha+\beta > 0,\ \alpha\beta+\gamma > 0,\ \alpha\gamma > 0 \ \text{ et } \ \beta(\alpha^2+\alpha\beta+\gamma) > 0.$$

On en déduit que $\alpha^2+\alpha\beta+\gamma \geqslant \alpha\beta+\gamma > 0$; la dernière inégalité donne alors $\beta > 0$. D'autre part, α et γ sont de même signe ; s'ils étaient tous deux négatifs, on aurait $\alpha\beta+\gamma < 0$. On a donc $\alpha > 0, \beta > 0, \gamma > 0$, et d'après ce qui précède, les racines complexes du polynôme P ont toutes une partie réelle strictement négative.

Conclusion. Les racines complexes du polynôme $X^3 + aX^2 + bX + c$ ont une partie réelle strictement négative si et seulement si

$$\boxed{a > 0, b > 0, c > 0, ab - c > 0}. \quad \triangleleft$$

La règle de Descartes et le théorème de Sturm ramènent l'étude du nombre de racines sur un intervalle d'un polynôme à coefficients réels à l'évaluation d'une fonction simple — un nombre de changements de signe — en deux points. Ce sont des méthodes aisément programmables et présentes dans les logiciels de calcul formel (fonctions sturm, sturmseq, realroot de Maple).

5.49. Règle de Descartes

On considère un polynôme P à coefficients réels de degré n. Pour tout réel x, on considère la suite $(\mathrm{P}^{(k)}(x))_{k\in[\![0,n]\!]}$. On note $\mathrm{V}(x)$ le nombre de changements de signe stricts de cette suite : $\mathrm{V}(x)$ est le cardinal de

$$\{(i,j), 0\leqslant i<j\leqslant n,\ \mathrm{P}^{(i)}(x)\mathrm{P}^{(j)}(x)<0 \text{ et } \mathrm{P}^{(k)}(x)=0 \text{ si } i<k<j\}.$$

1. Soit a et b tels que $\mathrm{P}(a)\mathrm{P}(b)\neq 0$ et $\mu(a,b)$ le nombre de racines de P dans l'intervalle $[a,b]$, comptées avec leur ordre de multiplicité. Montrer que

$$\mu(a,b)\leqslant \mathrm{V}(a)-\mathrm{V}(b) \quad \text{et} \quad \mu(a,b)\equiv \mathrm{V}(a)-\mathrm{V}(b) \pmod 2.$$

2. On note $\mathrm{P}=a_0+a_1\mathrm{X}+\cdots+a_n\mathrm{X}^n$, $\nu(\mathrm{P})$ le nombre de changements de signe dans la suite $(a_0,\ldots,a_n)$ et $\mu(\mathrm{P})$ le nombre de racines de P dans $\mathbb{R}_+^*$ comptées avec leur ordre de multiplicité.

Montrer que $\mu(\mathrm{P})\leqslant\nu(\mathrm{P})$ et que $\mu(\mathrm{P})\equiv\nu(\mathrm{P})\pmod 2$.

(École normale supérieure)

Solution.

1. Il s'agit d'étudier comment varie V sur l'intervalle $[a,b]$.

Remarquons que la fonction V est constante sur chaque intervalle contenu dans $[a,b]$ sur lequel aucune fonction $\mathrm{P}^{(k)}$ $(0\leqslant k\leqslant n)$ ne s'annule.

Il suffit donc de voir comment varie $\mathrm{V}(x)$ quand x «traverse» en croissant une racine d'un des polynômes $\mathrm{P}^{(k)}$ $(0\leqslant k\leqslant n)$.

Le raisonnement qui suit est basé sur la remarque simple que voici. Supposons que c soit une racine d'un polynôme A, non nul. La fonction polynôme A^2 présente en c un minimum strict. Sa dérivée $2\mathrm{AA}'$ est négative, puis positive, sur un voisinage de c. Ainsi, pour $h\neq 0$ assez petit, $\mathrm{A}(c+h)\mathrm{A}'(c+h)$ a le signe de h.

• Supposons que c est une racine de P d'ordre m : c annule donc les polynômes $\mathrm{P},\mathrm{P}',\ldots,\mathrm{P}^{(m-1)}$, mais pas $\mathrm{P}^{(m)}$. On en déduit que, pour $h\neq 0$ assez petit et $0\leqslant i\leqslant m-1$, $\mathrm{P}^{(i)}(c+h)\mathrm{P}^{(i+1)}(c+h)$ a le signe de h et donc est négatif, puis positif. On en déduit qu'en traversant c, $\mathrm{V}(x)$ diminue de m, ordre de multiplicité de c.

• Supposons maintenant que c est une racine de $\mathrm{P}^{(k)}$ d'ordre m (on suppose $\mathrm{P}^{(k-1)}(c)\neq 0$). Le même raisonnement que précédemment montre qu'on a une diminution de $\mathrm{V}(x)$ de m, car les produits $\mathrm{P}^{(i)}(x)\mathrm{P}^{(i+1)}(x)$, pour $k\leqslant i\leqslant k+m-1$ de négatifs deviennent positifs quand x traverse c.

Mais une variation supplémentaire de $V(x)$ peut être due au changement de signe de $P^{(k-1)}(x)P^{(k)}(x)$. On a, par hypothèse, $P^{(k-1)}(c) \neq 0$ et donc $P^{(k-1)}$ garde un signe constant dans un voisinage de c. Quant à $P^{(k)}$, il change de signe en c si m est impair.

Si m est pair, $V(x)$ diminue donc de m ; si m est impair, $V(x)$ diminue donc de $m-1$ ou $m+1$ selon que $P^{(k-1)}(x)P^{(k)}(x)$ est positif ou négatif avant c. Dans tous les cas, $V(x)$ diminue ici d'un nombre pair.

• Quant on fait la somme de toutes les diminutions obtenues de la fonction V entre a et b, on trouve la somme des ordres de multiplicité des racines de P sur $]a,b[$, et donc sur $[a,b]$, car $P(a)P(b) \neq 0$, c'est-à-dire $\mu(a,b)$, à quoi s'ajoute un entier naturel pair. Il existe donc $N \in \mathbb{N}$ tel que $V(a) - V(b) = \mu(a,b) + 2N$. On en déduit que

$$\mu(a,b) \leqslant V(a) - V(b) \text{ et } \mu(a,b) \equiv V(a) - V(b) \pmod 2.$$

2. Quitte à diviser le polynôme P par X^k, où k est sa valuation, ce qui ne modifie ni $\mu(P)$, ni $\nu(P)$, on peut supposer que $a_0 \neq 0$.

En 0, on obtient $P^{(i)}(0) = i!\,a_i$, pour $0 \leqslant i \leqslant n$. On en déduit que $V(0) = \nu(P)$.

D'autre part, étant donné un polynôme quelconque A, pour x assez grand $A(x)$ garde un signe constant, celui de son coefficient dominant. On peut donc trouver $\alpha \in \mathbb{R}_+^*$ tel que, pour $x \geqslant \alpha$, $P(x), P'(x), \ldots, P^{(n)}(x)$ ont le signe de a_n. On en déduit que, pour $x \geqslant \alpha$, on a $V(x) = 0$.

Nous obtenons donc finalement $V(0) - V(\alpha) = \nu(P)$.

Nous voyons, d'autre part, que toutes les racines positives de P et de ses dérivées non nulles appartiennent à l'intervalle $]0,\alpha[$. De la première question, appliquée à l'intervalle $[0,\alpha]$, nous déduisons donc que

$$\mu(P) \leqslant \nu(P) \text{ et } \mu(P) \equiv \nu(P) \pmod 2. \quad \triangleleft$$

5.50. Théorème de Sturm

Soit $P \in \mathbb{R}[X]$. On pose $S_0 = P$, $S_1 = P'$, puis, aussi longtemps que c'est possible, $S_{i-1} = A_i S_i - S_{i+1}$, avec $\deg(S_{i+1}) < \deg(S_i)$.

Pour x réel, on note $V(x)$ le nombre de changements de signes (stricts) de la suite $S_0(x), \ldots, S_p(x)$, lorsque $S_{p+1} = 0$. Ainsi, $V(x)$ est le cardinal de l'ensemble

$$\{(i,j), 0 \leqslant i < j \leqslant p, S_i(x)S_j(x) < 0 \text{ et } S_k(x) = 0 \text{ si } i < k < j\}.$$

Soit $a < b$. On suppose que $P(a)P(b) \neq 0$. Montrer que le nombre de racines distinctes de P dans $[a,b]$ est égal à $V(a) - V(b)$.

(École normale supérieure)

Solution.

• S_{i+1} est l'opposé du reste dans la division euclidienne de S_{i-1} par S_i. Celle-ci est possible tant que S_i n'est pas nul. La suite $S_0, \dots, S_p, S_{p+1}$ est, donc au signe près la suite des restes obtenus en calculant le pgcd de P et P′ par l'algorithme d'Euclide. Par construction, on a donc, pour tout $i \in [\![0, p]\!]$,

$$\operatorname{pgcd}(S_i, S_{i+1}) = S_p.$$

On se ramène à un polynôme n'ayant que des racines simples en posant, pour tout $i \in [\![0, p]\!]$.

$$T_i = \frac{S_i}{S_p}.$$

Alors, pour tout $i \in [\![0, p]\!]$, les polynômes T_i et T_{i+1} sont premiers entre eux et n'ont donc pas de racine commune ; on note que $T_p = 1$.

Les racines de T_0 sont les racines de P, mais toutes les racines de T_0 sont simples. En effet, si α est racine de P d'ordre m, alors α est racine d'ordre $m-1$ de P′ et $(X-\alpha)^{m-1}$ est en facteur dans S_p.

• Notons que si $S_p(x) \neq 0$, $V(x)$ est égal au nombre $V_1(x)$ de changements de signes (stricts) de la suite $T_0(x), \dots, T_p(x)$. C'est vrai en particulier si $x = a$ ou b (car $P(x) \neq 0$ implique $S_p(x) \neq 0$). On a donc

$$V(a) - V(b) = V_1(a) - V_1(b).$$

• Il s'agit de démontrer que $V_1(a) - V_1(b)$ est égal au nombre de racines de P (ou de T_0) sur $[a, b]$.

On remarque d'abord que la fonction V_1 est constante sur tout intervalle contenu dans $[a, b]$ sur lequel aucune des fonctions T_i ne s'annule.

Comme les fonctions T_i n'ont qu'un nombre fini de racines, il suffit de démontrer que V_1 possède les deux propriété suivantes :

(*i*) $V_1(x)$ diminue de 1 quand x «traverse» une racine de T_0, c'est-à-dire une racine de P ;

(*ii*) $V_1(x)$ ne change pas lorsque x «traverse» un point α tel que $T_0(\alpha)$ est non nul et $T_i(\alpha) = 0$ pour au moins un indice $i \geqslant 1$.

• Supposons que α soit une racine de P. La fonction P^2 présente en α un minimum. Sa dérivée $2PP'$ est donc négative, puis positive. Ainsi, pour $h \neq 0$ assez petit $P(\alpha+h)P'(\alpha+h)$ a le signe de h et donc $T_0(\alpha+h)T_1(\alpha+h)$ a le signe de h (car $(S_p(\alpha+h))^2 > 0$ si $P(\alpha+h) \neq 0$). Ce qui montre qu'après la traversée de α, il y a un changement de signe de moins entre les termes d'indice 0 et 1.

• Supposons maintenant que α soit une racine de T_i, avec $i \geqslant 1$; on a nécessairement $i < p$. De l'égalité $S_{i-1} = A_i S_i - S_{i+1}$, on déduit $T_{i-1} = A_i T_i - T_{i+1}$, puis $T_{i-1}(\alpha) = -T_{i+1}(\alpha)$. Mais α n'est pas racine de T_{i-1}, ni de T_{i+1}. On a donc $T_{i-1}(\alpha)T_{i+1}(\alpha) < 0$. Par continuité, on

on déduit qu'on a encore $T_{i-1}(x)T_{i+1}(x) < 0$, pour x dans un voisinage de α. Le nombre de changements de signe entre les termes d'indices $i-1$ et $i+1$ ne change pas, il reste égal à 1 (quel que soit le signe de $T_i(x)$ sur ce voisinage de α).

Ceci achève la démonstration. ◁

Les exercices suivants traitent de fractions rationnelles. Le premier est une classique décomposition en éléments simples. Rappelons que si α est un pôle simple de la fraction rationnelle $F = \frac{P}{Q}$, la partie polaire de F relative à α est $\frac{P(\alpha)}{(X-\alpha)Q'(\alpha)}$.

5.51. Décomposition en éléments simples

Soit P un polynôme de $\mathbb{C}[X]$ n'admettant que des racines simples non nulles $x_1, \dots, x_n$.

Montrer que $\sum_{i=1}^{n} \frac{1}{x_i P'(x_i)} = -\frac{1}{P(0)}$. Que vaut $\sum_{i=1}^{n} \frac{1}{P'(x_i)}$?

(École polytechnique)

▷ **Solution.**

• On décompose en éléments simples $F(X) = \frac{1}{P(X)}$. Les pôles de F sont les x_i et ils sont tous simples par hypothèse. Puisque la partie entière de F est nulle on a

$$F(X) = \frac{1}{P(X)} = \sum_{k=1}^{n} \frac{1}{\prod_{k \neq i} (x_i - x_k)} \frac{1}{X - x_i} = \sum_{k=1}^{n} \frac{1}{P'(x_i)(X - x_i)}. \quad (*)$$

L'identité demandée s'obtient en évaluant cette égalité en 0.

• Pour calculer la somme $\sum_{i=1}^{n} \frac{1}{P'(x_i)}$, on multiplie $(*)$ par X, on évalue en x réel et on fait tendre x vers l'infini : $\sum_{i=1}^{n} \frac{1}{P'(x_i)}$ vaut 1 si $n = 1$ et 0 si $n > 1$. ◁

L'exercice qui suit offre, à partir d'une relation remarquable entre P et P', une preuve très simple d'une inégalité de Bernstein. Elle a été trouvée en 1973 par P.J. O'Hara.

5.52. Inégalité de Bernstein

1. Soit $n \in \mathbb{N}^*$ et $z_1, \ldots, z_n$ les racines de $X^n + 1$. Démontrer que pour tout polynôme P de $\mathbb{C}_n[X]$, on a

$$XP'(X) = \frac{n}{2}P(X) + \frac{2}{n}\sum_{k=1}^{n} \frac{z_k P(z_k X)}{(z_k - 1)^2}.$$

2. Si $P \in \mathbb{C}[X]$ on pose $\|P\| = \sup_{|z|=1} |P(z)|$. Montrer que pour tout $P \in \mathbb{C}_n[X]$ on a $\|P'\| \leqslant n\|P\|$.

(**École Polytechnique**)

▷ **Solution.**

1. Par linéarité, il suffit de vérifier l'identité pour $P = X^\ell$, lorsque ℓ est dans $[\![0, n]\!]$. Il s'agit donc de calculer

$$\alpha = \sum_{k=1}^{n} \frac{z_k z_k^\ell}{(z_k - 1)^2} = \sum_{k=1}^{n} \frac{z_k^{\ell+1}}{(z_k - 1)^2}$$

et de vérifier la relation $\ell = \frac{n}{2} + \frac{2}{n}\alpha$. Le carré au dénominateur fait penser à une dérivée. On considère donc $F = \dfrac{X^\ell}{X^n + 1}$ que l'on va dériver. Décomposons F en éléments simples dans $\mathbb{C}(X)$: il s'agit de trouver $\mu_1, \ldots, \mu_n$ dans $\mathbb{C}$ tels que

$$F = \sum_{k=1}^{n} \frac{\mu_k}{X - z_k} + E(F),$$

où E(F) désigne la partie entière de F qui vaut 0 si $\ell < n$ et 1 si $\ell = n$. Fixons $k \in [\![1, n]\!]$, on a alors

$$\mu_k = \left(X^\ell \frac{X - z_k}{X^n + 1}\right)(z_k) = \frac{z_k^\ell}{\prod_{j \neq k}(z_k - z_j)} = \frac{z_k^\ell}{T'(z_k)},$$

où $T = X^n + 1$. Ainsi, on obtient $\mu_k = \dfrac{z_k^\ell}{n z_k^{n-1}} = -\dfrac{z_k^{\ell+1}}{n}$ car $z_k^n = -1$. Il en résulte

$$F = -\frac{1}{n}\sum_{k=1}^{n} \frac{z_k^{\ell+1}}{(X - z_k)} \quad \text{et} \quad F' = \frac{1}{n}\sum_{k=1}^{n} \frac{z_k^{\ell+1}}{(X - z_k)^2}.$$

On a donc $F'(1) = \dfrac{\alpha}{n} = \dfrac{2\ell - n}{4}$ puisque $F' = \dfrac{\ell X^{\ell-1}(X^n + 1) - nX^{n-1}X^\ell}{(X^n + 1)^2}$.

Dans ces conditions, on obtient

$$\frac{n}{2}+\frac{2}{n}\alpha=\frac{n}{2}+\frac{2\ell-n}{2}=\ell.$$

C'est ce qu'on voulait.

2. Soit $P\in\mathbb{C}_n[X]$. Appliquons la formule précédente en un point z de module 1. Par inégalité triangulaire, il vient

$$|P'(z)|\leqslant\left(\frac{n}{2}+\frac{2}{n}\sum_{k=1}^{n}\frac{|z_k|}{|z_k-1|^2}\right)\|P\|$$

car les z_k sont tous de module 1. Il nous suffit donc de prouver que $\sum_{k=1}^{n}\frac{|z_k|}{|z_k-1|^2}\leqslant\frac{n^2}{4}$. Le point clé est que si $z=e^{i\theta}$ est un nombre complexe de module 1 distinct de 1, alors

$$\frac{z}{(z-1)^2}=\frac{e^{i\theta}}{(e^{i\theta}-1)^2}=\frac{1}{(e^{i\theta/2}-e^{-i\theta/2})^2}=\frac{-1}{4\sin^2\dfrac{\theta}{2}}$$

est un réel négatif. C'est en particulier le cas pour les z_k. On a donc

$$\sum_{k=1}^{n}\frac{|z_k|}{|z_k-1|^2}=-\sum_{k=1}^{n}\frac{z_k}{(z_k-1)^2}=\frac{n^2}{4}$$

d'après la question 1 appliquée avec le polynôme constant $P=1$. D'où le résultat. ◁

5.53. Inversion de la matrice de Hilbert

Soit $(a_1,\dots,a_n)$, $(b_1,\dots b_n)$ et $(c_1,\dots,c_n)$ dans $\mathbb{R}^n$. On suppose que les a_i sont deux à deux distincts de même que les b_j. On suppose de plus que $a_i+b_j\neq 0$ pour tout $(i,j)\in[\![1,n]\!]^2$.

1. Résoudre le système : $\forall j\in[\![1,n]\!]$, $\sum_{i=1}^{n}\frac{x_i}{a_i+b_j}=c_j$.

2. Soit $H=(h_{ij})\in\mathcal{M}_n(\mathbb{R})$ avec $h_{ij}=\frac{1}{i+j-1}$. Montrer que H est inversible et que H^{-1} est à coefficients entiers.

(**École normale supérieure**)

▷ **Solution.**

1. Pour $(x_1,\dots,x_n)\in\mathbb{R}^n$, on considère la fraction $F=\sum_{i=1}^{n}\frac{x_i}{a_i+X}$. Si l'on note P le polynôme $(X+a_1)\dots(X+a_n)$, alors il existe un polynôme Q de $\mathbb{R}_{n-1}[X]$ tel que $F=\frac{Q}{P}$ puisque $\deg F\leqslant -1$.

Le système considéré équivaut à : pour tout $j \in [\![1, n]\!]$,

$$F(b_j) = c_j, \text{ c'est-à-dire } Q(b_j) = P(b_j)c_j.$$

Les n réels $b_1, \dots b_n$ étant distincts, ces conditions déterminent un polynôme unique Q de $\mathbb{R}_{n-1}[X]$, qui s'exprime en fonction des polynômes interpolateurs de Lagrange relatifs au système $(b_1, \dots, b_n)$, c'est-à-dire les polynômes $L_j = \dfrac{\prod\limits_{k \neq j} (X - b_k)}{\prod\limits_{k \neq j} (b_j - b_k)}$, par $Q = \sum\limits_{j=1}^{n} P(b_j)c_j L_j$. La fraction rationnelle $F = \dfrac{Q}{P}$ est donc entièrement déterminée. On en déduit que le système considéré admet une solution unique. En effet, F se décompose de manière unique en $F = \sum\limits_{i=1}^{n} \dfrac{x_i}{a_i + X}$. On obtient, pour tout $i \in [\![1, n]\!]$.

$$x_i = \frac{Q(-a_i)}{P'(-a_i)} = \frac{1}{P'(-a_i)} \sum_{j=1}^{n} P(b_j)c_j L_j(-a_i).$$

On note que

$$P'(-a_i) = \prod_{k \neq i} (a_k - a_i),$$

et que

$$L_j(-a_i) = \frac{\prod\limits_{k \neq j} (-a_i - b_k)}{\prod\limits_{k \neq j} (b_j - b_k)} = \prod_{k=1}^{n} (a_i + b_k) \frac{1}{(a_i + b_j) \prod\limits_{k \neq j} (b_k - b_j)}.$$

On obtient finalement

$$x_i = \frac{\prod\limits_{k=1}^{n} (a_i + b_k)}{\prod\limits_{k \neq i} (a_k - a_i)} \sum_{j=1}^{n} c_j \frac{\prod\limits_{k \neq i} (a_k + b_j)}{\prod\limits_{k \neq j} (b_k - b_j)}.$$

2. Avec les notations de la question précédente, si on note A la matrice de $\mathcal{M}_n(\mathbb{R})$ dont le coefficient d'indice (i, j) est $\dfrac{1}{a_i + b_j}$ (matrice de Cauchy), le système devient $AX = C$, où $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ et $C = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}$.

Nous avons démontré que ce système a une solution unique pour tout $C \in \mathbb{R}^n$. La matrice A est donc inversible et l'expression des x_i en fonction des c_j donne les coefficients de A^{-1}.

Le coefficient d'indice (i,j) de A^{-1} est

$$a'_{i,j} = \frac{\prod_{k=1}^{n}(a_i+b_k)\prod_{k\neq i}(a_k+b_j)}{\prod_{k\neq i}(a_k-a_i)\prod_{k\neq j}(b_k-b_j)}.$$

La matrice H (matrice de Hilbert) est un cas particulier de matrice de Cauchy avec, pour tout $i \in [\![1,n]\!]$, $a_i = b_i = i - \frac{1}{2}$. Le coefficient d'indice (i,j) de H^{-1} est donc

$$h'_{i,j} = \frac{\prod_{k=1}^{n}(i+k-1)\prod_{k\neq i}(k+j-1)}{\prod_{k\neq i}(k-i)\prod_{k\neq j}(k-j)}.$$

On remarque que

$$\prod_{k=1}^{n}(i+k-1) = \frac{(i+n-1)!}{(i-1)!} \quad \text{et} \quad \prod_{k\neq i}(k-i) = (-1)^{i-1}(i-1)!(n-i)!.$$

On en déduit que

$$h'_{i,j} = (-1)^{i+j}\frac{(i+n-1)!(j+n-1)!}{((i-1)!)^2((j-1)!)^2(n-i)!(n-j)!(i+j-1)}.$$

L'égalité $\dfrac{1}{(i-1)!(j-1)!} = \dfrac{C_{i+j-2}^{i-1}}{(i+j-2)!}$ permet d'écrire

$$\begin{aligned} h'_{i,j} &= (-1)^{i+j}(C_{i+j-2}^{i-1})^2\frac{(i+n-1)!(j+n-1)!}{(n-i)!(n-j)!(i+j-2)!(i+j-1)!} \\ &= (-1)^{i+j}(i+n-1)(C_{i+j-2}^{i-1})^2C_{i+n-2}^{n-j}C_{j+n-1}^{n-i}. \end{aligned}$$

Cela montre que H^{-1} est à coefficients entiers. ◁

5.54. Automorphismes de K(X)

1. Déterminer les automorphismes de la K-algèbre K[X].
2. Déterminer les automorphismes de la K-algèbre K(X) des fractions rationnelles à une indéterminée.

(École polytechnique)

▷ **Solution.**

1. Soit Φ un automorphisme de la K-algèbre K[X]. Notons $P = \Phi(X)$. Alors $\Phi(X^n) = P^n$ pour tout $n \in \mathbb{N}$ et si $Q = \sum_{k\in\mathbb{N}} a_k X^k \in K[X]$,

$$\Phi(Q) = \sum_{k\in\mathbb{N}} a_k \Phi(X^k) = \sum_{k\in\mathbb{N}} a_k P^k = Q \circ P.$$

Pour que Φ soit surjective, il est nécessaire que P ne soit pas constant. Si $Q \neq 0$, $\deg(\Phi(Q)) = \deg P \times \deg Q$. La surjectivité de Φ impose $\deg P = 1$. Il existe donc $(a, b) \in K^* \times K$ tel que $\Phi(Q) = Q(aX + b)$ pour tout $Q \in K[X]$.

Réciproquement, si $(a, b) \in K^* \times K$, $Q \longmapsto Q(aX + b)$ est un automorphisme de la K-algèbre K[X] dont l'automorphisme réciproque est l'application $Q \longmapsto Q\left(\dfrac{X - b}{a}\right)$.

2. • Soit Φ un morphisme d'algèbres de K(X). Posons $F = \Phi(X)$. On démontre alors, comme dans la question 1, que, pour tout $P \in K[X]$, on a $\Phi(P) = P \circ F$. Si $G \in K(X)$ s'écrit $G = \dfrac{P}{Q}$, avec P et Q dans K[X], on obtient

$$\Phi(G) = \Phi\left(\frac{P}{Q}\right) = \frac{\Phi(P)}{\Phi(Q)} = \frac{P \circ F}{Q \circ F} = G \circ F.$$

Il est clair, réciproquement, que pour tout $F \in K(X)$, l'application

$$\Phi_F : G \in K(X) \longmapsto G \circ F \in K(X)$$

est un morphisme de K-algèbre. Reste à trouver à quelle condition sur F l'application Φ_F est un automorphisme.

• Si Φ_F est un automorphisme, il est surjectif et en particulier il existe $G \in K(X)$ tel que $\Phi(G) = G \circ F = X$. Soit A, B, P, Q quatre polynômes tels que $\dfrac{A}{B}$ et $\dfrac{P}{Q}$ soient des représentants irréductibles de F et G. On note p et q les degrés respectifs de P et Q, $a_0, \ldots, a_p$ et $b_0, \ldots, b_q$ les scalaires tels que $P = \sum_{j=0}^{p} a_j X^j$ et $Q = \sum_{k=0}^{q} b_k X^k$. La relation $G \circ F = X$ équivaut à $P \circ F = X(Q \circ F)$, c'est-à-dire à

$$\sum_{j=0}^{p} a_j F^j = X \sum_{k=0}^{q} b_k F^k, \text{ soit } \sum_{j=0}^{p} a_j \frac{A^j}{B^j} = X \sum_{k=0}^{q} b_k \frac{A^k}{B^k}.$$

Si on note $m = \max(p, q)$, on obtient

$$\sum_{j=0}^{p} a_j A^j B^{m-j} = X \sum_{k=0}^{q} b_k A^k B^{m-k}.$$

Tous les polynômes qui interviennent dans cette égalité sont divisibles par A, sauf ceux qui correspondent à $j = 0$ et $k = 0$. On en déduit que A divise $a_0 B^m - b_0 X B^m$. Le polynôme A étant premier avec B et donc avec B^m, on en déduit que A divise $a_0 - b_0 X$. Le couple (a_0, b_0) n'est pas égal à $(0,0)$ sinon $\frac{P}{Q}$ ne serait pas irréductible, donc le polynôme $a_0 - b_0 X$ est de degré $\leqslant 1$. On en déduit que $\deg(A) \leqslant 1$.

Intéressons-nous maintenant aux polynômes divisibles par B dans l'égalité précédente. Si $m = p = q$, B divise $(a_p - b_p X)A^p$ et donc $a_p - b_p X$, puisqu'il est premier avec A^p. Le polynôme $a_p - b_p X$ est de degré 1 car $b_p \neq 0$. On en déduit que $\deg(B) \leqslant 1$. Si $m = q > p$, on montre que B divise $b_q X$ et si $m = p > q$, B divise a_p. Dans tous les cas, la conclusion est la même : $\deg(P) \leqslant 1$.

Nous avons donc démontré qu'il existe $(a, b, c, d) \in K^4$ tel que

$$F = \frac{aX + b}{cX + d}.$$

Il est clair que F ne peut pas être constant ; il faut donc $ad - bc \neq 0$.

• Réciproquement, étant donnés quatre éléments a, b, c, d de K tels que $ad - bc \neq 0$, notons $\Phi_{a,b,c,d}$ le morphisme de K(X) défini par

$$\Phi_{a,b,c,d}(X) = \frac{aX + b}{cX + d}.$$

On montre facilement que $\Phi_{1,0,0,1} = \mathrm{Id}_{K(X)}$ et que si (a, b, c, d) et (a', b', c', d') sont deux quadruplets tels que $ad - bc \neq 0$ et $a'd' - b'c' \neq 0$, alors

$$\Phi_{a,b,c,d} \circ \Phi_{a',b',c',d'} = \Phi_{a'',b'',c'',d''},$$

où a'', b'', c'', d'' sont tels que $\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a'' & b'' \\ c'' & d'' \end{pmatrix}$.

La matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ est inversible. Si $\begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1}$, alors

$$\Phi_{a,b,c,d} \circ \Phi_{a',b',c',d'} = \Phi_{a',b',c',d'} \circ \Phi_{a,b,c,d} = \mathrm{Id}_{K(X)}.$$

Le morphisme $\Phi_{a,b,c,d}$ est inversible : c'est un automorphisme de K(X).

Conclusion. Les automorphismes d'algèbre de K(X) sont les applications

$$G \in K(X) \longmapsto G\left(\frac{aX + b}{cX + d}\right) \in K(X),$$

avec $(a, b, c, d) \in K^4$ et $ad - bc \neq 0$. De plus l'application qui à la matrice $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(K)$ associe $\Phi_{a,b,c,d}$ est un morphisme de groupes. Son noyau est le sous-groupe des matrices scalaires non nulles. ◁

Les nombres algébriques sur $\mathbb{Q}$ sont les racines des polynômes à coefficients entiers. On s'intéresse ici à l'approximation des nombres algébriques par des rationnels.

5.55. Approximation d'un irrationnel algébrique par des rationnels

Soit $P \in \mathbb{Z}[X]$ non nul et a une racine irrationnelle de P.

1. Théorème de Liouville : montrer qu'il existe des réels $c > 0$ et $r \geqslant 0$ tels que $\left|a - \frac{p}{q}\right| \geqslant \frac{c}{q^r}$ pour tout couple $(p, q) \in \mathbb{Z} \times \mathbb{N}^*$.

2. Montrer qu'il existe $d > 0$ tel que $\left|a - \frac{p}{q}\right| \leqslant \frac{d}{q^2}$ pour une infinité de couples $(p, q) \in \mathbb{Z} \times \mathbb{N}^*$.

3. Montrer que l'ordre de a comme racine de P est inférieur ou égal à $\frac{n}{2}$ où n est le degré de P.

(École normale supérieure)

▷ **Solution.**

1. On décompose P en produit de facteurs irréductibles dans $\mathbb{Q}[X]$ et on considère $Q \in \mathbb{Q}[X]$ irréductible, divisant P et tel que $Q(a) = 0$. On note $r \geqslant 2$ le degré de Q. Soit $p \in \mathbb{Z}$ et $q \in \mathbb{N}^*$. Comme $a \neq \frac{p}{q}$, en vertu du théorème des accroissements finis, il existe $d \in \left]\frac{p}{q}, a\right[$ tel que

$$Q\left(\frac{p}{q}\right) = Q\left(\frac{p}{q}\right) - Q(a) = Q'(d)\left(\frac{p}{q} - a\right).$$

En passant à la valeur absolue, on obtient $\left|Q\left(\frac{p}{q}\right)\right| = |Q'(d)|\left|a - \frac{p}{q}\right|$. Il existe $A \in \mathbb{Z}^*$ tel que $AQ \in \mathbb{Z}[X]$. Il s'ensuit que $Aq^r Q\left(\frac{p}{q}\right)$ est dans $\mathbb{Z}$ et même dans $\mathbb{Z}^*$, puisque Q n'a pas de racine rationnelle (en effet, il est irréductible sur $\mathbb{Q}$ et n'est pas de degré 1 puisque $a \notin \mathbb{Q}$ est racine). Par conséquent, on a $\left|Aq^r Q\left(\frac{p}{q}\right)\right| \geqslant 1$ et

$$|Q'(d)|\left|a - \frac{p}{q}\right| = \frac{1}{Aq^r}\left|Aq^r Q\left(\frac{p}{q}\right)\right| \geqslant \frac{1}{Aq^r}.$$

Cherchons à majorer $|Q'(d)|$: la fonction Q' est continue sur $[a-1, a+1]$, donc il existe $K > 0$ tel que, si $x \in [a-1, a+1]$, $|Q'(x)| \leqslant K$. Ainsi on a, si $\frac{p}{q} \in [a-1, a+1]$,

$$K\left|a - \frac{p}{q}\right| \geqslant \frac{1}{Aq^r} \quad \text{et} \quad \left|a - \frac{p}{q}\right| \geqslant \frac{c}{q^r}$$

en ayant posé $c = \dfrac{1}{\mathrm{KA}} > 0$. Quitte à changer c en 1 si $c > 1$, on peut supposer $c \leqslant 1$. L'inégalité $\left|a - \dfrac{p}{q}\right| \geqslant \dfrac{c}{q^r}$ reste alors vraie si $\left|a - \dfrac{p}{q}\right| > 1$.

Conclusion. On a trouvé $r \geqslant 2$ et $c > 0$ tel que si $(p, q) \in \mathbb{Z} \times \mathbb{N}^*$,

$$\boxed{\left|a - \frac{p}{q}\right| \geqslant \frac{c}{q^r}}.$$

2. Nous allons démontrer la propriété demandée avec $d = 1$.

• Montrons pour commencer le résultat suivant : étant donné N dans N^*, il existe $(p, q) \in \mathbb{Z}^2$ tel que $|qa - p| < \dfrac{1}{\mathrm{N}}$ et $0 < q \leqslant \mathrm{N}$. Pour cela, nous appliquons la méthode des tiroirs de Dirichlet. Considérons les $\mathrm{N} + 1$ nombres réels $a_k = ka - \mathrm{E}(ka)$ $(0 \leqslant k \leqslant \mathrm{N})$. Ils appartiennent à l'intervalle $[0, 1[$. Celui-ci est réunion des N «tiroirs» $\left[\dfrac{i}{\mathrm{N}}, \dfrac{i+1}{\mathrm{N}}\right[$, pour $0 \leqslant i \leqslant \mathrm{N} - 1$. Un de ces tiroirs doit contenir deux réels a_k. Il existe donc des entiers k_1 et k_2 tels que $0 \leqslant k_1 < k_2 \leqslant \mathrm{N}$ et $|a_{k_2} - a_{k_1}| < \dfrac{1}{\mathrm{N}}$. En posant $q = k_2 - k_1$ et $p = \mathrm{E}(k_2 a) - \mathrm{E}(k_1 a)$, on obtient $0 < q \leqslant \mathrm{N}$ et $|aq - p| < \dfrac{1}{\mathrm{N}}$.

• De ce résultat on déduit, en prenant un entier N strictement positif quelconque, que $\left|a - \dfrac{p}{q}\right| < \dfrac{1}{q\mathrm{N}} \leqslant \dfrac{1}{q^2}$, ce qui montre l'existence d'un couple (p, q) vérifiant l'inégalité voulue. Pour montrer qu'il en existe une infinité, on raisonne par l'absurde et on suppose qu'il n'en existe qu'un nombre fini $\dfrac{p_1}{q_1}, \ldots, \dfrac{p_m}{q_m}$. Puisque a est irrationnel, il existe un entier N strictement positif tel que, pour $1 \leqslant i \leqslant m$, on ait $\left|a - \dfrac{p_i}{q_i}\right| > \dfrac{1}{\mathrm{N}}$. Mais, d'après ce qui précède, il existe $(p, q) \in \mathbb{Z}^2$ tel que $|qa - p| < \dfrac{1}{\mathrm{N}}$ et $0 < q \leqslant \mathrm{N}$. On a alors $\left|a - \dfrac{p}{q}\right| \leqslant \dfrac{1}{q^2}$, mais aussi $\left|a - \dfrac{p}{q}\right| < \dfrac{1}{q\mathrm{N}} \leqslant \dfrac{1}{\mathrm{N}}$, ce qui montre que $\dfrac{p}{q}$ n'est pas un des $\dfrac{p_i}{q_i}$. D'où la contradiction cherchée.

3. Appelons d l'ordre de a comme racine de P et démontrons que $d \leqslant \dfrac{n}{2}$, par récurrence sur d. Considérons à nouveau $\mathrm{Q} \in \mathbb{Q}[\mathrm{X}]$, facteur irréductible de P, admettant a comme racine. Comme $a \notin \mathbb{Q}$, on a $\deg \mathrm{Q} = r \geqslant 2$.

• Si $d = 1$, alors $2 \leqslant \deg \mathrm{Q} \leqslant \deg \mathrm{P}$ et $n \geqslant 2d$.

• Supposons $d \geqslant 2$. Montrons que a est racine simple de Q. En effet, l'ensemble $\mathrm{I} = \{\mathrm{A} \in \mathrm{K}[\mathrm{X}], \mathrm{A}(a) = 0\}$ est un idéal de $\mathbb{Q}[\mathrm{X}]$. Il est donc principal et admet un générateur unitaire A non constant. Or $\mathrm{Q} \in \mathrm{I}$ donc A divise Q et $\mathrm{Q} \in \mathbb{Q}^* \mathrm{A}$, puisque Q est irréductible. Si a était une

racine double de Q, on aurait $Q'(a) = 0$ et $Q' \in I$. Alors Q diviserait Q', ce qui est exclu car $Q' \neq 0$ et $\deg Q' < \deg Q$.

Ainsi, a est une racine d'ordre $d-1$ du polynôme $P_1 = \dfrac{P}{Q}$. On a, par hypothèse de récurrence, $n - \deg Q = \deg P_1 \geqslant 2(d-1)$, d'où l'on déduit $n \geqslant 2(d-1) + \deg Q \geqslant 2d$. ◁

Le théorème démontré dans la première question de l'exercice permit à Liouville de donner les premiers exemples de nombres transcendants, c'est-à-dire non algébriques (en 1844). Supposons que a soit un réel irrationnel tel que, pour tout $k \in \mathbb{N}^*$, il existe $(p,q) \in \mathbb{Z} \times \mathbb{N}^*$, $q \geqslant 2$ tel que $\left|a - \dfrac{p}{q}\right| < \dfrac{1}{q^k}$. Alors, il ne peut exister c et r comme dans la question 1 de l'exercice. En effet, en prenant k assez grand, on arrive à une contradiction puisque $q \geqslant 2$. On montre ainsi que $a = \sum_{i=0}^{+\infty} \dfrac{1}{10^{i!}}$ est transcendant.

En 1863, Cantor, après avoir démontré que $\mathbb{R}$ n'est pas dénombrable, prouva l'existence d'une infinité de nombres transcendants en remarquant que l'ensemble des nombres algébriques est dénombrable. L'exercice suivant prouve la transcendance de e, résultat établi par Hermite en 1873.

5.56. Transcendance de e

1. Soit $P \in \mathbb{R}[X]$ et, pour $t \in \mathbb{R}$, $I(t) = \int_0^t e^{t-u} P(u)\,du$. Montrer que, si $\deg(P) = q$, $I(t) = e^t \sum_{i=0}^{q} P^{(i)}(0) - \sum_{i=0}^{q} P^{(i)}(t)$.

2. Supposons donnés des entiers relatifs $a_0, a_1, \ldots, a_n$ tels que $a_0 \neq 0$ et $a_0 + a_1 e + \cdots + a_n e^n = 0$. Soit $p \in \mathbb{N}$; on pose

$$P = X^{p-1}(X-1)^p(X-2)^p \cdots (X-n)^p$$

et

$$J = a_0 I(0) + a_1 I(1) + \cdots + a_n I(n).$$

Montrer que J est un entier. Montrer que $(p-1)!$ divise J et enfin que, pour tout entier premier p assez grand, $J \neq 0$.

3. Montrer qu'il existe $C \in \mathbb{R}$ tel que, pour tout $p \in \mathbb{N}$, $|J| \leqslant C^p$ et trouver une minoration de $|J|$. En déduire que e est transcendant.

(**École polytechnique**)

▷ **Solution.**

1. On définit $D(P) = \sum_{i=0}^{q} P^{(i)}$ et la fonction f par $f(u) = e^{-u} D(P)(u)$. On note que $D(P)' = D(P) - P$, car $P^{(q+1)} = 0$; on en déduit que, pour tout $u \in \mathbb{R}$, on a

$$f'(u) = -e^{-u}\mathrm{D}(\mathrm{P})(u) + e^{-u}\mathrm{D}(\mathrm{P})'(u) = -e^{-u}\mathrm{P}(u).$$

On obtient alors

$$\mathrm{I}(t) = e^t \int_0^t e^{-u}\mathrm{P}(u)\mathrm{d}u = e^t \left[-e^{-u}\mathrm{D}(\mathrm{P})(u)\right]_0^t = e^t\mathrm{D}(\mathrm{P})(0) - \mathrm{D}(\mathrm{P})(t),$$

c'est-à-dire

$$\mathrm{I}(t) = e^t \sum_{i=0}^{q} \mathrm{P}^{(i)}(0) - \sum_{i=0}^{q} \mathrm{P}^{(i)}(t).$$

2. • Avec les notations précédentes, on a, pour tout entier $k \in \mathbb{N}$,

$$\mathrm{I}(k) = e^k\mathrm{D}(\mathrm{P})(0) - \mathrm{D}(\mathrm{P})(k).$$

On en déduit que

$$\mathrm{J} = \left(\sum_{k=0}^{n} a_k e^k\right)\mathrm{D}(\mathrm{P})(0) - \sum_{k=0}^{n} a_k\mathrm{D}(\mathrm{P})(k) = -\sum_{k=0}^{n} a_k\mathrm{D}(\mathrm{P})(k).$$

Le polynôme P est à coefficients entiers, donc D(P) également ; les a_k étant entiers, on en déduit que J est un entier.

• Montrons, que pour $k \in [\![0, n]\!]$, D(P)(k) est divisible par $(p-1)!$. Le polynôme P s'écrit $\mathrm{X}^{p-1}\mathrm{Q}$, avec $\mathrm{Q} \in \mathbb{Z}[\mathrm{X}]$. Pour $i \in \mathbb{N}$, on calcule $\mathrm{P}^{(i)}$ par la formule de Leibniz. Les termes qui apportent une contribution non nulle à $\mathrm{P}^{(i)}(0)$ sont ceux pour lesquels on aura dérivé $p-1$ fois X^{p-1}. La dérivée $(p-1)$-ième de X^{p-1} étant $(p-1)!$, on en déduit que, pour tout entier i, $\mathrm{P}^{(i)}(0)$ est divisible par $(p-1)!$. Il en est donc de même de D(P)(0).

Pour $1 \leqslant k \leqslant n$, en écrivant $\mathrm{P} = (\mathrm{X}-k)^p\mathrm{R}$, avec $\mathrm{R} \in \mathbb{Z}[\mathrm{X}]$, on montre de la même manière que D(P)(k) est divisible par $p!$ et a fortiori par $(p-1)!$.

Finalement, chaque D(P)(k) étant divisible par $(p-1)!$, on conclut que J est divisible par $(p-1)!$.

• Montons que, si p est un nombre premier assez grand, J n'est pas divisible par $p!$. Avec les notations précédentes, on a, pour $i \geqslant 1$, $\mathrm{Q}^{(i)}(0)$ divisible par p, car il existe $\mathrm{Q}_1 \in \mathbb{Z}[\mathrm{X}]$ tel que $\mathrm{Q} = \mathrm{Q}_1^p$. On en déduit que, si $i \geqslant p$, alors $\mathrm{P}^{(i)}(0)$ est divisible par $p!$. Comme

$$\mathrm{P}^{(p-1)}(0) = (p-1)!\mathrm{Q}(0) = (p-1)!(-1)^{pn}(n!)^p,$$

il en résulte que

$$\mathrm{D}(\mathrm{P})(0) \equiv (p-1)!(-1)^{pn}(n!)^p \pmod{p!},$$

puis que

$$\mathrm{J} \equiv -a_0(p-1)!(-1)^{pn}(n!)^p \pmod{p!}.$$

Par hypothèse, a_0 est non nul. Si p est un entier premier supérieur à $|a_0|$ et à n, p ne divise ni a_0, ni $n!$ et donc $p!$ ne divise pas J. On en déduit qu'alors J n'est pas nul.

3. Pour majorer J, on majore $\mathrm{I}(k)$, pour $0 \leqslant k \leqslant n$, en revenant à l'expression initiale de $\mathrm{I}(t)$. On a, pour $k \in [\![0, n]\!]$,

$$|\mathrm{I}(k)| = \left| \int_0^k e^{t-u} \mathrm{P}(u) \mathrm{d}u \right| \leqslant \sup_{[0,k]} |\mathrm{P}| \int_0^k e^{k-u} \mathrm{d}u = (e^k - 1) \sup_{[0,k]} |\mathrm{P}|$$
$$\leqslant e^n \sup_{[0,n]} |\mathrm{P}|.$$

Or, pour tout $x \in [0, n]$, on a

$$|\mathrm{P}(x)| = x^{p-1} \left(|x-1||x-2| \cdots |x-n|\right)^p \leqslant n^{p-1} \left(n^n\right)^p.$$

On en déduit que, pour tout $k \in [\![0, n]\!]$,

$$|\mathrm{I}(k)| \leqslant e^n n^{p-1} \left(n^n\right)^p,$$

puis que, pour tout entier $p \in \mathbb{N}$,

$$|\mathrm{J}| \leqslant \sum_{k=0}^{n} |a_k| e^n n^{p-1} \left(n^n\right)^p.$$

Soit $a = \max\left(\sum_{k=0}^{n} |a_k| e^n, n \right)$. On a alors, pour tout $p \in \mathbb{N}$,

$$|\mathrm{J}| \leqslant a^p \left(n^n\right)^p \leqslant \left(a n^n\right)^p \quad \text{c'est-à-dire } |\mathrm{J}| \leqslant \mathrm{C}^p,$$

avec $\mathrm{C} = a n^n$.

Mais nous avons vu dans la question précédente que, si p est un entier premier assez grand, alors J est non nul et divisible par $(p-1)!$. On en déduit que $|\mathrm{J}| \geqslant (p-1)!$ et a fortiori, $\mathrm{C}^p \geqslant (p-1)!$. Or on a $\lim_{p \to +\infty} \dfrac{\mathrm{C}^p}{(p-1)!} = 0$. Pour p assez grand, on a donc $\mathrm{C}^p < (p-1)!$. D'où la contradiction cherchée. Il ne peut pas exister des entiers $a_0, a_1, \ldots, a_n$ appartenant à $\mathbb{Z}$ tels que $a_0 \neq 0$ et $a_0 + a_1 e + \cdots + a_n e^n = 0$.

On en déduit que e est transcendant. En effet, si e est algébrique sur $\mathbb{Q}$, il existe $\mathrm{P} \in \mathbb{Q}[\mathrm{X}]$, tel que $\mathrm{P}(e) = 0$. En multipliant P par un entier, on peut supposer que $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$. Il existe donc des entiers $a_0, a_1, \ldots, a_n$ appartenant à $\mathbb{Z}$ tels que $a_0 + a_1 e + \cdots + a_n e^n = 0$. On peut supposer que $a_0 \neq 0$, car si l'ordre de 0 comme racine de P est p, on divise P par X^p.

Les derniers exercices du chapitre traitent des polynômes à plusieurs variables. Le premier est une généralisation du résultat de l'exercice 5.24 aux polynômes à plusieurs variables.

5.57. Polynômes à plusieurs variables à valeurs entières

1. Soit $P \in \mathbb{R}[X]$ avec $\deg P \leqslant n$. Montrer que $P(\mathbb{Z}) \subset \mathbb{Z}$ si et seulement si $P(0), P(1), \ldots, P(n)$ sont dans $\mathbb{Z}$.

2. Soit $Q(X_1, \ldots, X_\ell)$ un polynôme à ℓ variables à coefficients dans $\mathbb{R}$. Donner une condition nécessaire et suffisante pour que l'on ait $P(\mathbb{Z}^\ell) \subset \mathbb{Z}$.

3. Montrer que pour tout $(m_1, \ldots, m_\ell) \in \mathbb{Z}^\ell$, $\prod_{k=1}^{\ell-1} k!$ divise $\prod_{1 \leqslant i < j \leqslant \ell} (m_j - m_i)$.

(École normale supérieure)

▷ **Solution.**

1. Il est clair que si $P(\mathbb{Z}) \subset \mathbb{Z}$, alors $P(0), \ldots, P(n)$ sont dans $\mathbb{Z}$. Pour démontrer la réciproque, nous raisonnerons par récurrence sur n.

• Si $n = 0$, le polynôme P est constant et cette constante est entière, puisque $P(0) \in \mathbb{Z}$.

• Supposons la propriété établie pour les polynômes de degré $\leqslant n-1$. Soit P un polynôme de degré $\leqslant n$ tel que $P(0), P(1), \ldots, P(n)$ soient dans $\mathbb{Z}$. Considérons le polynôme $Q = P(X+1) - P(X)$. Il est de degré $\leqslant n-1$ et, pour $0 \leqslant k \leqslant n-1$, on a $Q(k) = P(k+1) - P(k) \in \mathbb{Z}$. On en déduit que $Q(\mathbb{Z}) \subset \mathbb{Z}$. On remarque que, pour tout $k \in \mathbb{N}$, on a $P(k+1) = P(k) + Q(k)$ et $P(-k-1) = P(-k) - Q(-k-1)$. Sachant que $P(0) \in \mathbb{Z}$, une récurrence banale sur k permet de conclure que, pour tout $k \in \mathbb{N}$, $P(k)$ et $P(-k)$ sont dans $\mathbb{Z}$. On conclut $P(\mathbb{Z}) \subset \mathbb{Z}$, ce qui achève la démonstration[7].

2. Montrons que si $Q \in \mathbb{R}[X_1, \ldots, X_\ell]$ et si le degré de Q en chacune des variables est inférieur ou égal à n, alors $Q(\mathbb{Z}^\ell) \subset \mathbb{Z}$ si et seulement si on a, pour tout $(x_1, \ldots, x_\ell) \in [\![0, n]\!]^\ell$, $Q(x_1, \ldots, x_\ell) \in \mathbb{Z}$. Là encore une des deux implications est évidente. On démontre que si on a, pour tout $(x_1, \ldots, x_\ell) \in [\![0, n]\!]^\ell$, $Q(x_1, \ldots, x_\ell) \in \mathbb{Z}$, alors $Q(\mathbb{Z}^\ell) \subset \mathbb{Z}$, en raisonnant par récurrence sur $\ell \in \mathbb{N}^*$.

• Le cas $\ell = 1$ a été traité dans la question précédente.

• Supposons la propriété vérifiée au rang $\ell - 1 \geqslant 1$ et considérons un polynôme Q de ℓ variables et de degré $\leqslant n$ par rapport à chaque variable, tel que $Q(x_1, \ldots, x_\ell) \in \mathbb{Z}$ pour tout $(x_1, \ldots, x_\ell) \in [\![0, n]\!]^\ell$. Fixons x_ℓ dans $[\![0, n]\!]$ et considérons le polynôme $Q_1 \in \mathbb{R}[X_1, \ldots, X_{\ell-1}]$ défini par

$$Q_1(X_1, \ldots, X_{\ell-1}) = Q(X_1, \ldots, X_{\ell-1}, x_\ell).$$

Le polynôme Q_1 est de degré $\leqslant n$ par rapport à $X_1, \ldots, X_{\ell-1}$ et vérifie,

7. On a donné ici une preuve différente de celle que le lecteur trouvera dans l'exercice 5.24.

par hypothèse, pour tout $(x_1, \dots, x_{\ell-1}) \in [\![0, n]\!]^{\ell-1}$,

$$Q_1(x_1, \dots, x_{\ell-1}) = Q(x_1, \dots, x_{\ell-1}, x_\ell) \in \mathbb{Z}.$$

De l'hypothèse de récurrence, on déduit que $Q_1(\mathbb{Z}^{\ell-1}) \subset \mathbb{Z}$. On a donc, pour tout $(x_1, \dots, x_{\ell-1}) \in \mathbb{Z}^{\ell-1}$, $Q(x_1, \dots, x_{\ell-1}, x_\ell) = Q_1(x_1, \dots, x_{\ell-1})$ dans $\mathbb{Z}$, ceci pour tout $x_\ell \in [\![0, n]\!]$.

Fixons maintenant $(x_1, \dots, x_{\ell-1})$, élément quelconque de $\mathbb{Z}^{\ell-1}$ et considérons le polynôme $Q_2 \in \mathbb{R}[X_\ell]$ défini par

$$Q_2 = Q(x_1, \dots, x_{\ell-1}, X_\ell).$$

On a $\deg Q_2 \leqslant n$ et, d'après ce qui précède, $Q_2(x_\ell) \in \mathbb{Z}$ pour tout $x_\ell \in [\![0, n]\!]$. On en déduit que $Q(x_1, \dots, x_{\ell-1}, x_\ell) = Q_2(x_\ell) \in \mathbb{Z}$, pour tout $x_\ell \in \mathbb{Z}$, d'après la question 1. Cela montre que $Q(\mathbb{Z}^\ell) \subset \mathbb{Z}$ et termine la démonstration par récurrence.

3. Considérons le polynôme $Q \in \mathbb{R}[X_1, \dots, X_\ell]$ défini par

$$Q = \prod_{1 \leqslant i < j \leqslant \ell} \frac{X_j - X_i}{j - i}.$$

Montrons que $Q(\mathbb{Z}^\ell) \subset \mathbb{Z}$. Puisque Q est de degré $\ell - 1$ par rapport à chacune des variables, il suffit de vérifier, d'après la question 2 que $Q(x_1, \dots, x_\ell) \in \mathbb{Z}$ pour tout $(x_1, \dots, x_\ell) \in [\![0, \ell - 1]\!]^\ell$.

S'il existe deux indices distincts i et j tels que $x_i = x_j$, alors on a $Q(x_1, \dots, x_\ell) = 0 \in \mathbb{Z}$. Sinon, on obtient $\{x_1, \dots, x_\ell\} = \{0, 1, \dots, \ell - 1\}$ et donc $\{x_1 + 1, \dots, x_\ell + 1\} = \{1, 2, \dots, \ell\}$. Il existe $\sigma \in \mathcal{S}_n$ tel que, pour tout $i \in [\![1, \ell]\!]$, $x_i + 1 = \sigma(i)$. On en déduit que

$$Q(x_1, \dots, x_\ell) = \prod_{1 \leqslant i < j \leqslant \ell} \frac{\sigma(j) - \sigma(i)}{j - i} = \varepsilon(\sigma) \in \mathbb{Z}.$$

Calculons maintenant $N = \prod_{1 \leqslant i < j \leqslant \ell} (j - i)$. On obtient

$$N = \prod_{j=2}^{\ell} \left(\prod_{i=1}^{j-1} (j - i) \right) = \prod_{j=2}^{\ell} (j-1)! = \prod_{j=1}^{\ell-1} j!.$$

D'après ce qui précède, on a pour tout $(m_1, \dots, m_\ell) \in \mathbb{Z}^\ell$,

$$Q(m_1, \dots, m_\ell) = \frac{\prod_{1 \leqslant i < j \leqslant \ell} (m_j - m_i)}{N} \in \mathbb{Z}.$$

Autrement dit, $N = \prod_{j=1}^{\ell-1} j!$ divise $\prod_{1 \leqslant i < j \leqslant \ell} (m_j - m_i)$. ◁

Le lecteur trouvera dans l'exercice 1.8 du tome 2 d'algèbre une autre solution de cette dernière question, utilisant le déterminant de Vandermonde.

5.58. Applications de $\mathbb{R}^2$ dans $\mathbb{R}$ polynomiales par rapport à chaque variable

1. Soit $f : \mathbb{R}^2 \to \mathbb{R}$ telle que $y \mapsto f(x,y)$ est polynomiale pour tout $x \in \mathbb{R}$ et $x \mapsto f(x,y)$ est polynomiale pour tout $y \in \mathbb{R}$. Montrer que f est polynomiale.

2. Que se passe-t-il si on remplace le corps $\mathbb{R}$ par $\mathbb{Q}$?

(École normale supérieure)

▷ **Solution.**

1. Si f est polynomiale de degré n par rapport à y, pour tout $x \in \mathbb{R}$, l'application polynomiale $y \longmapsto f(x,y)$ est de degré inférieur ou égal à n. Cela incite à considérer, pour tout $d \in \mathbb{N}$, l'ensemble E_d des x pour lesquels l'application polynomiale $y \longmapsto f(x,y)$ est de degré $\leqslant d$. Comme $\mathbb{R} = \bigcup_{n\in\mathbb{N}} \mathrm{E}_d$, l'un des ensembles E_d est infini (sinon $\mathbb{R}$ serait dénombrable). Notons-le E_n. Pour tout $x \in \mathrm{E}_n$, l'application polynomiale $y \longmapsto f(x,y)$, qui est degré inférieure ou égale à n, est déterminée par les valeurs en $n+1$ points, disons $0, 1, \ldots, n$. En notant L_i $(0 \leqslant i \leqslant n)$ les polynômes interpolateurs élémentaires en ces $n+1$ valeurs, on obtient, pour $x \in \mathrm{E}_n$ et $y \in \mathbb{R}$,

$$f(x,y) = \sum_{i=0}^{n} f(x,i)\mathrm{L}_i(y).$$

Pour tout $i \in [\![0,n]\!]$, l'application $x \longmapsto f(x,i)$ est polynomiale. Ainsi l'application $g : (x,y) \longmapsto \sum_{i=0}^{n} f(x,i)\mathrm{L}_i(y)$ est polynomiale.

Soit $y \in \mathbb{R}$. L'application $x \longmapsto f(x,y)$ est polynomiale et pour tout $x \in \mathrm{E}_n$, $f(x,y) = g(x,y)$. Les applications polynomiales

$$x \longmapsto f(x,y) \quad \text{et} \quad x \longmapsto g(x,y)$$

qui coïncident pour une infinité de valeurs sont égales. On a donc, pour tout $(x,y) \in \mathbb{R}^2$, $f(x,y) = g(x,y)$ et $f = g$.

2. Si le corps de base est $\mathbb{Q}$ et plus généralement tout corps dénombrable, il est possible qu'aucun des ensembles E_d ne soit infini et la démonstration précédente tombe en défaut. Construisons un contre-exemple.

Le degré des applications $y \longmapsto f(x,y)$ ne doit pas être borné quand x décrit $\mathbb{Q}$; il en est de même des applications $y \longmapsto f(x,y)$. Ainsi f doit être une somme infinie de termes polynomiaux qui pour x ou y fixé sont presque tous nuls. Notons (r_i) une énumération de $\mathbb{Q}$ et pour tout

entier naturel i, L_i le polynôme interpolateur de degré i qui vaut 1 en r_i et 0 en $r_0, \ldots, r_{i-1}$. Posons, pour tout $(x, y) \in \mathbb{Q}^2$,

$$f(x,y) = \sum_{i=0}^{+\infty} L_i(x)L_i(y).$$

On note que $L_i(r_k) = 0$ si $k < i$. Ainsi la somme ne contient jamais qu'un nombre fini de termes et, pour tout $k \in \mathbb{N}$, $f(r_k, x) = \sum_{i=0}^{k} L_i(r_k)L_i(x)$. L'application $y \longmapsto f(x,y)$ est donc polynomiale pour tout $x \in \mathbb{Q}$. Il en est de même de l'application $x \longmapsto f(x,y)$ pour tout $y \in \mathbb{Q}$.

Mais f n'est pas polynomiale car pour tout $k \in \mathbb{N}$, $y \longmapsto f(r_k, x)$ est de degré k, donc l'ensemble des degrés des applications $y \longmapsto f(r_k, x)$ n'est pas borné. ◁

La démonstration du théorème de Bezout qui suit illustre les grandes différences entre l'arithmétique de $K[X]$ *et celle de* $K[X_1, \ldots, X_n]$ *pour* $n \geqslant 2$. *Ce dernier n'est pas un anneau principal (d'après l'exercice 3.9) ; l'identité de Bezout ne s'y applique pas : par exemple* X *et* Y *sont premiers entre eux, mais l'identité* $AX + BY = 1$ *n'est pas possible pour* $(A, B) \in K[X, Y]^2$. *Néanmoins, c'est un anneau intègre ; c'est même un anneau factoriel, c'est-à-dire que tout polynôme peut s'écrire comme produit de polynômes irréductibles (voir 3.10 pour la définition précise).*

5.59. Un théorème de Bezout

Soit P et Q deux polynômes de $K[X, Y]$, premiers entre eux.

1. Montrer qu'il existe $(A, B) \in K[X, Y]$ et $\Delta \in K[X]$, non nul, tels que $\Delta = AP + BQ$.

2. Montrer que l'ensemble des couples $(x, y) \in K^2$ solutions de $P(x, y) = Q(x, y) = 0$ est fini.

(École normale supérieure)

▷ **Solution.**

1. • On souhaiterait appliquer le théorème de Bezout à P et Q, mais comme cela est dit dans le préambule, on ne peut pas le faire directement. On va donc se placer sur un anneau où le théorème de Bezout s'applique. On peut voir P et Q comme des polynômes en Y à coefficients dans $K[X]$. Comme $K[X]$ est contenu dans le corps commutatif $L = K(X)$, P et Q se révèlent être des éléments de $L[Y]$, domaine sur lequel s'applique le théorème de Bezout.

• P et Q sont premiers entre eux dans K[X, Y]. Montons qu'ils restent premiers entre eux dans L(Y). Raisonnons par l'absurde et supposons qu'un irréductible D de L(Y) divise à la fois P et Q dans L(Y). Le polynôme D s'écrit

$$D = a_n(X)Y^n + a_{n-1}(X)Y^{n-1} + \cdots + a_1(X)Y + a_0(X),$$

avec $n \geqslant 1$ et les $a_i(X) \in K(X)$. Quitte à multiplier D par le produit des dénominateurs des fractions rationnelles $a_i(X)$ (ce produit est un élément de L non nul), on peut supposer que les $a_i(X)$ sont dans K[X] et $D \in K[X, Y]$.

Résumons : $D \in K[X, Y]$ divise P et Q dans $L[Y] = K(X)[Y]$. Il existe donc F_1 et F_2 dans L[Y] tels que $F_1D = P$ et $F_2D = Q$. Pour arriver à une contradiction, il suffit de montrer que F_1 et F_2 sont en fait dans $K[X][Y] = K[X, Y]$.

Pour cela, donnons quelques définitions (ce sont les mêmes que dans $\mathbb{Z}[X]$: on pourra se reporter à l'exercice 5.16 sur le critère d'Eisenstein). Pour tout $M \in K[X][Y]$ qui s'écrit $a_p(X)Y^p + \cdots + a_1(X)Y + a_0(X)$ (avec $p \geqslant 0$ et les $a_i(X)$ dans K[X]), on appelle *contenu* de M et on note $c_X(M)$ le pgcd des $a_i(X)$ pour i variant de 0 à p. On dit qu'un polynôme M est *primitif* si $c_X(M) = 1$. En divisant tous les coefficients de M par leur pgcd, on trouve un polynôme primitif : ainsi M s'écrit $M = c_X(M)\tilde{M}$ avec $\tilde{M} \in K[X][Y]$ primitif.

Nous allons nous appuyer sur le lemme suivant.

Lemme. (Lemme de Gauss) *Soit* P *et* Q *dans* K[X][Y].

(*i*) *Si* P *et* Q *sont primitifs, leur produit* PQ *l'est aussi.*

(*ii*) *On a* $c_X(PQ) = c_X(P) \times c_X(Q)$.

Démonstration.

(*i*) Supposons P et Q primitifs. Raisonnons par l'absurde et imaginons que PQ ne le soit pas. On écrit $P = \sum_{k=0}^{n} a_k(X)Y^k$, $Q = \sum_{k=0}^{m} b_k(X)Y^k$ et $PQ = \sum_{k=0}^{m+n} c_k(X)Y^k$. Puisque PQ n'est pas primitif, il existe $R \in K[X]$, irréductible, divisant tous les $c_k(X)$. Comme P et Q sont primitifs, il existe un plus petit entier k (resp. ℓ) tels que R ne divise pas $a_k(X)$ (resp. $b_\ell(X)$). Considérons alors

$$c_{k+\ell}(X) = a_0b_{k+\ell} + \cdots + a_{k-1}b_{\ell+1} + a_kb_\ell + a_{k+1}b_{\ell-1} + \cdots + a_{k+\ell}b_0.$$

Par définition de k, les k premiers termes sont divisibles par R puisque R divise a_i si $i < k$. De même, par définition de ℓ, les ℓ derniers termes sont divisibles par R. Comme $c_{k+\ell}$ est également divisible par R, il

s'ensuit que R divise le terme restant $a_k b_\ell$. Or, cela est impossible, en vertu du lemme d'Euclide, puisque ni a_k, ni b_ℓ ne sont divisibles par l'irréductible R.

On conclut donc que PQ est primitif.

(ii) On écrit $P = c_X(P)\widetilde{P}$, $Q = c_X(Q)\widetilde{Q}$. On a $PQ = c_X(P)c_X(Q)\widetilde{P}\widetilde{Q}$. D'après le 1, $\widetilde{P}\widetilde{Q}$ est primitif et donc le contenu de PQ est $c_X(P)c_X(Q)$. D'où l'égalité voulue. ◇

Comme $\widetilde{D}$ est également irréductible et divise P et Q dans L(Y), quitte à remplacer D par $\widetilde{D}$, on peut supposer D primitif. On peut multiplier $F_1 \in K(X)[Y]$ par un polynôme $R(X) \in K[X]$ afin de chasser les dénominateurs des coefficients de F_1 : $F_1 = \dfrac{F}{R(X)}$, et $F \in K[X][Y]$. On écrit ensuite $F = c_X(F)\widetilde{F}$ avec $\widetilde{F}$ primitif. Par conséquent, on obtient

$$D\frac{c_X(F)\widetilde{F}}{R(X)} = P \quad \text{ou encore} \quad Dc_X(F)\widetilde{F} = R(X)P.$$

Passons aux contenus : $c_X(F) = c_X(D)c_X(F) = c_X(Dc_X(F)\widetilde{F}) = Rc_X(P)$. Donc R divise $c_X(F)$ et finalement $F_1 = \dfrac{c_X(F)}{R}\widetilde{F} \in K[X][Y]$. Autrement dit, D divise P dans K[X, Y]. Il en est de même pour Q, par symétrie du problème. Comme P et Q sont premiers entre eux dans K[X, Y], nous aboutissons à une contradiction.

• Nous avons donc démontré que P et Q sont premiers entre eux dans K(X)[Y]. Il existe A′ et B′ dans K(X)[Y] tels que $1 = A'P + B'Q$. Soit $\Delta \in K[X]$ le produit des dénominateurs des coefficients de A et de B. On a alors

$$\Delta = AP + BQ \quad \text{avec} \quad A = \Delta A' \in K[X, Y] \text{ et } B = \Delta B' \in K[X, Y].$$

2. Soit maintenant $(x, y) \in K^2$ tel que $P(x, y) = Q(x, y) = 0$. En substituant dans l'égalité ci-dessus, il vient $\Delta(x) = 0$. Donc x est une racine de Δ. Il ne peut prendre qu'un nombre fini de valeurs puisque Δ est non nul. Par symétrie du problème en X et Y, y lui aussi ne peut prendre qu'un nombre fini de valeurs. Il en résulte que

$$V = \{(x, y) \in K^2,\ P(x, y) = Q(x, y) = 0\} \quad \text{est fini.}$$

Conclusion. Ce résultat s'interprète ainsi : l'intersection de deux courbes algébriques planes $P(x, y) = 0$ et $Q(x, y) = 0$ est finie si P et Q sont des polynômes de K[X, Y] premiers entre eux. ◁

Bezout a en fait montré que si P *et* Q *sont de degrés respectifs* m *et* n*, le nombre de points d'intersection des deux courbes, comptés avec leur ordre de multiplicité d'intersection et en tenant compte des points à l'infini, est exactement* mn *lorsque* K *est algébriquement clos.*

Chapitre 6

Espaces vectoriels. Algèbres

Au XVIII^{e} *siècle se développent la résolution des systèmes linéaires et la théorie des déterminants. Les raisonnements suggèrent rapidement le concept d'espace à n dimensions. Mais il fallait oser un langage géométrique, alors qu'une interprétation sensible dans le plan ou l'espace faisait défaut pour* $n > 3$.

De manière indépendante, Cayley en Angleterre et Grassman en Allemagne franchissent le pas vers 1843-1845 et parlent d'espace à n dimensions. Le point de vue de Caylcy est issu directement de la géométrie analytique : un vecteur d'un espace à n dimensions est un système de n réels ou n complexes. L'addition de deux vecteurs et la multiplication par un scalaire sont naturellement introduites par la généralisation de la dimension 3. *Pour parvenir vraiment à la notion d'espace vectoriel, il faut dégager le concept de sous-espace et de dimension d'un sous-espace. C'est ce que fera Grassman (professeur de lycée autodidacte en marge des milieux de la recherche) en cherchant à développer une analyse géométrique portant sur des calculs intrinsèques indépendants du choix des coordonnées. Grassman introduit le produit extérieur de deux vecteurs, la définition de l'indépendance linéaire, de la dimension d'un espace et démontre la relation fondamentale*

$$\dim \mathrm{V} + \dim \mathrm{W} = \dim(\mathrm{V} + \mathrm{W}) + \dim \mathrm{V} \cap \mathrm{W}.$$

Ces travaux eurent peu d'impact au début, mais ils furent repris par Henri Poincaré et Elie Cartan (notamment son «algèbre extérieure» en géométrie différentielle).

C'est en 1888 que Peano donnera la définition axiomatique d'un espace vectoriel réel. Jusqu'en 1930, le point de vue des matrices et des coordonnées prédomine par rapport au point de vue intrinsèque des espaces vectoriels.

Les premiers exercices portent sur les sous-espaces vectoriels.

6.1. Intersection de sous-espaces

Soit E un espace vectoriel de dimension n et $\mathrm{F}_1, \ldots, \mathrm{F}_k$ des sous-espaces de E. Montrer que si $\sum_{i=1}^{k} \dim \mathrm{F}_i > n(k-1)$ alors $\bigcap_{i=1}^{k} \mathrm{F}_i \neq \{0\}$.

(École normale supérieure)

▷ **Solution.**

• Il suffit d'observer que $\bigcap_{i=1}^{k} \mathrm{F}_i$ est isomorphe au noyau de l'application linéaire

$$\begin{array}{rccc} & \mathrm{F}_1 \times \cdots \times \mathrm{F}_k & \longrightarrow & \mathrm{E}^{k-1} \\ \psi : & (x_1, x_2, \ldots, x_k) & \longmapsto & (x_2 - x_1, x_3 - x_1, \ldots, x_k - x_1) \end{array}$$

En effet, $\psi(x_1, \ldots, x_k) = 0$ si et seulement si $(x_1, \ldots, x_k) = (x, x, \ldots, x)$ où x appartient à $\bigcap_{i=1}^{k} \mathrm{F}_i$. Le théorème du rang permet alors d'affirmer que

$$\dim(\mathrm{F}_1 \times \mathrm{F}_2 \times \cdots \times \mathrm{F}_k) = \sum_{i=1}^{k} \dim \mathrm{F}_i = \operatorname{rg} \psi + \dim \bigcap_{i=1}^{k} \mathrm{F}_i.$$

Comme $\operatorname{rg}(\psi) \leqslant \dim \mathrm{E}^{k-1} = n(k-1)$, on a $\dim \bigcap_{i=1}^{k} \mathrm{F}_i > 0$. ◁

On peut rédiger la solution à l'aide des espaces quotients. Prenons donc cet exercice comme prétexte pour fournir quelques rappels sur cette notion.

Soit E *un* K*-espace vectoriel,* F *un sous-espace de* E*. Comme* F *est un sous-groupe du groupe abélien* $(\mathrm{E}, +)$*, on peut considérer le groupe quotient* E/F*, défini à partir de la relation d'équivalence telle que, pour tout* $(x, y) \in \mathrm{E}^2$*,*

$$x \equiv y \pmod{\mathrm{F}} \iff y - x \in \mathrm{F}.$$

La loi de composition externe est compatible avec cette congruence, i.e. *si* $\lambda \in \mathrm{K}$ *et* $x \equiv y \pmod{\mathrm{F}}$*, alors* $\lambda x \equiv \lambda y \pmod{\mathrm{F}}$*, si bien que l'on peut munir* E/F *d'une loi de composition externe définie pour* $\lambda \in \mathrm{K}$ *et* $x \in \mathrm{E}$ *par* $\lambda \overline{x} = \overline{\lambda x}$ *(la classe de* λx *ne dépend pas du représentant* x *choisi dans* $\overline{x}$*). Dans ces conditions,* E/F *est un* K*-espace vectoriel (le lecteur est invité à faire les quelques vérifications d'usage) appelé quotient de* E *par* F*.*

Si E *est de dimension finie,* E/F *est aussi de dimension finie et l'on a* $\dim \mathrm{E/F} = \dim \mathrm{E} - \dim \mathrm{F}$*. En effet, c'est une conséquence du théorème du rang, puisque la surjection canonique* $s : x \in \mathrm{E} \longmapsto \overline{x} \in \mathrm{E/F}$ *est linéaire de noyau* F*.*

Revenons maintenant à l'exercice ci-dessus. Dire que $x \in \mathrm{F}_i$ *revient à dire que son image dans* $\mathrm{E/F}_i$ *par la surjection canonique est nulle. On considère donc*

$$\begin{array}{rccc} & \mathrm{E} & \longrightarrow & \mathrm{E/F}_1 \times \cdots \times \mathrm{E/F}_k \\ \Psi : & x & \longmapsto & (\overline{x}, \ldots, \overline{x}) \end{array}$$

Il s'agit juste de prouver que, sous l'hypothèse $\sum_{i=1}^{k} \dim \mathrm{F}_i > n(k-1)$, *cette application linéaire n'est pas injective. C'est vrai car* $\dim \mathrm{E} = n$ *et*

$$\dim\left(\mathrm{E}/\mathrm{F}_1 \times \cdots \times \mathrm{E}/\mathrm{F}_k\right) = nk - \sum_{i=1}^{k} \dim \mathrm{F}_i < n = \dim \mathrm{E}.$$

L'exercice suivant a trait à la notion de supplémentaire. En dimension finie, l'existence d'un supplémentaire à un sous-espace donné résulte du théorème de la base incomplète. Nous pourrons, dans certains exercices de ce chapitre, admettre qu'en dimension infinie, tout sous-espace vectoriel admet également un supplémentaire, résultat qui nécessite toutefois l'axiome du choix.

6.2. Supplémentaire commun

Soit K un corps infini, E un K-espace vectoriel de dimension finie.

1. Montrer qu'on ne peut avoir $\mathrm{E} = \mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_\mathrm{N}$, où les V_i sont des sous-espaces stricts de E. Que se passe-t-il si K est fini ?

2. Montrer que si $\mathrm{F}_1, \mathrm{F}_2, \ldots, \mathrm{F}_p$ sont des sous-espaces vectoriels de E de même dimension (finie), il existe G supplémentaire commun à tous les F_i.

(École normale supérieure)

▷ **Solution.**

1. Raisonnons par l'absurde et supposons que $\mathrm{E} = \mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_\mathrm{N}$, où les V_i sont des sous-espaces stricts de E. On a nécessairement $\mathrm{N} \geqslant 2$, sinon $\mathrm{E} = \mathrm{V}_1$. Quitte à retirer un certain nombre de sous-espaces, jusqu'à ce que le nombre N de sous-espaces dont la réunion est E soit minimal, on peut supposer qu'aucun sous-espace n'est contenu dans la réunion des $\mathrm{N}-1$ autres. Pour aboutir alors à une contradiction, nous proposons deux méthodes différentes.

La première utilise directement un argument de cardinalité. On peut trouver $x \in \mathrm{V}_\mathrm{N}$ tel que $x \notin \mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_{\mathrm{N}-1}$. D'autre part, on n'a pas $\mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_{\mathrm{N}-1} \subset \mathrm{V}_\mathrm{N}$, sinon on aurait $\mathrm{E} = \mathrm{V}_\mathrm{N}$. Il existe donc $y \in \mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_{\mathrm{N}-1}$ tel que $y \notin \mathrm{V}_\mathrm{N}$. Considérons, pour tout $\lambda \in \mathrm{K}$, le vecteur $y + \lambda x$. Il n'appartient pas à V_N, car sinon $y \in \mathrm{V}_\mathrm{N}$. Il appartient donc à $\mathrm{V}_1 \cup \mathrm{V}_2 \cup \cdots \cup \mathrm{V}_{\mathrm{N}-1}$ de sorte qu'il existe $i_\lambda \in [\![1, \mathrm{N}-1]\!]$ tel que $y + \lambda x \in \mathrm{V}_{i_\lambda}$. L'application $\varphi : \lambda \in \mathrm{K} \longmapsto i_\lambda \in [\![1, \mathrm{N}-1]\!]$ est injective. En effet, supposons qu'il existe λ et μ, dans K tels que $i_\lambda = i_\mu$. Les vecteurs $y + \lambda x$ et $y + \mu x$ sont dans V_{i_λ}, donc leur différence $(\lambda - \mu)x$

aussi. Comme $x \notin V_{i_\lambda}$ on a $\mu = \lambda$. On obtient donc une application injective de K, qui est un ensemble infini, dans $[\![1, N-1]\!]$ qui est un ensemble fini. C'est absurde : E ne peut pas être réunion d'une famille finie de sous-espace stricts.

Autre méthode. Quitte à remplacer chaque sous-espace V_i par un hyperplan le contenant, on peut supposer que les V_i sont des hyperplans. On considère alors, pour tout $i \in [\![1, N]\!]$, une forme linéaire φ_i sur E , telle que $V_i = \operatorname{Ker} \varphi_i$. On n'a pas $V_i \subset \bigcup_{\substack{1 \leqslant j \leqslant N \\ j \neq i}} V_j$. Il existe donc $x_i \in V_i \setminus \bigcup_{\substack{1 \leqslant j \leqslant N \\ j \neq i}} V_j$. Considérons, pour tout $i \in [\![1, N]\!]$, l'application $F_i : K^n \to K$ définie par $F_i(\lambda_1, \ldots, \lambda_N) = \varphi_i(\sum_{j=1}^{N} \lambda_j x_j)$. C'est une fonction polynomiale sur K^N. L'hypothèse $E = V_1 \cup V_2 \cup \cdots \cup V_N$ entraîne $\prod_{i=1}^{N} \varphi_i = 0$ et donc $\prod_{i=1}^{N} F_i = 0$. Le corps K étant infini, l'algèbre des fonctions polynomiales sur K^n est isomorphe à $K[X_1, \ldots, X_N]$, qui est un anneau intègre. Il existe donc un entier i_0 entre 1 et N tel que $F_{i_0} = 0$. Puisque $N \geqslant 2$, on peut considérer un indice $i \neq i_0$. On obtient alors $\varphi_{i_0}(x_i) = F_{i_0}(0, \ldots, 1, \ldots, 0) = 0$, ce qui contredit le fait que $x_i \notin V_{i_0}$.

Si le corps K est fini, on obtient $|K| \leqslant N - 1$ par la première démonstration. Ainsi E ne peut pas être réunion de $|K|$ sous-espaces stricts ou moins.

Le lecteur pourra montrer que $|K| + 1$ *sous-espaces stricts suffisent effectivement pour recouvrir* E.

2. Il existe des sous-espaces G de E tels que $F_i \cap G = \{0\}$ pour tout $i \in [\![1, p]\!]$, par exemple le sous-espace nul. Parmi ces sous-espaces, considérons-en un de dimension maximale que l'on note G. Soit r la dimension commune des F_i. Si $r + \dim G < \dim E$, les sous-espaces $F_i \oplus G$ sont des sous-espaces stricts de E. Il existe alors, d'après la première question, un vecteur x n'appartenant à aucun de ces espaces. Mais alors le sous-espace $G' = G \oplus Kx$ est en somme directe avec tous les F_i, contredisant ainsi la maximalité de G. On a donc $r + \dim G = \dim E$ et G est un supplémentaire commun à tous les F_i. ◁

On peut montrer, plus généralement, que dans le cas où K *est infini,* E *ne peut pas être réunion d'une famille* $(F_i)_{i \in I}$ *de sous-espaces stricts avec* $|I| < |K|$. *Par exemple si* $K = \mathbb{R}$, E *ne peut pas être réunion d'une famille dénombrable de sous-espaces stricts. Nous donnerons une preuve topologique de ce résultat dans le tome 3 d'analyse, à l'aide du théorème de Baire.*

La deuxième question se généralise alors directement avec la même démonstration : pour toute famille $(F_i)_{i \in I}$ *de sous-espaces de même dimension avec* $|I| < |K|$ *il existe un supplémentaire commun à tous les* F_i.

On rappelle que si V *est un espace vectoriel de dimension* n*, un drapeau de* V *est une suite croissante pour l'inclusion de* $n+1$ *sous-espaces vectoriels* $\mathrm{V}_0 \subset \mathrm{V}_1 \subset \cdots \subset \mathrm{V}_n$ *avec* $\dim \mathrm{V}_k = k$ *pour tout* k*. Cette notion se présente naturellement dans l'étude de la trigonalisation : un endomorphisme* u *de* V *est trigonalisable si et seulement s'il existe un drapeau stable par* u *(c'est-à-dire* $u(\mathrm{V}_k) \subset \mathrm{V}_k$ *pour tout* k*, avec les notations ci-dessus). Le groupe linéaire opère naturellement sur les drapeaux et le lecteur vérifiera facilement que cette opération est transitive. Le difficile exercice qui suit étudie le nombre d'orbites pour l'action du groupe linéaire sur les couples de drapeaux.*

6.3. Drapeaux

Soit V un K-espace vectoriel de dimension n, d et d' deux drapeaux, respectivement $\mathrm{V}_0 \subset \mathrm{V}_1 \subset \cdots \subset \mathrm{V}_n$ et $\mathrm{V}'_0 \subset \mathrm{V}'_1 \subset \cdots \subset \mathrm{V}'_n$.

1. On pose $\mathrm{A}_{ij} = \mathrm{V}'_{i-1} + \mathrm{V}_j$ pour $i \in [\![1, n+1]\!]$ et $j \in [\![0, n]\!]$ et

$$\sigma(i) = \min\{j,\ \mathrm{A}_{ij} = \mathrm{A}_{i+1,j}\} \quad \text{pour} \quad i \in [\![1, n]\!].$$

Montrer que σ est une bijection de $[\![1, n]\!]$ sur $[\![1, n]\!]$ (on pourra considérer l'application σ' définie comme σ, mais en échangeant les rôles de d et d').

2. Par définition, une base $(e_1, \ldots, e_n)$ est adaptée à d si, pour tout $k \in [\![1, n]\!]$, $\mathrm{Vect}(e_1, \ldots, e_k) = \mathrm{V}_k$. Trouver une base $(e_1, \ldots, e_n)$ adaptée à d et une permutation σ de $[\![1, n]\!]$ telle que $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$ soit adaptée à d'.

3. On fait opérer naturellement $\mathrm{GL}(\mathrm{V})$ sur les couples de drapeaux de V. Quel est le nombre d'orbites ?

(École normale supérieure)

▷ **Solution.**

1. • Notons, pour commencer, que, pour $i \in [\![1, n]\!]$, on a $\sigma(i) \in [\![1, n]\!]$. En effet, pour tout $i \in [\![1, n]\!]$, on a $\mathrm{A}_{i0} = \mathrm{V}'_{i-1} + \mathrm{V}_0 = \mathrm{V}'_{i-1}$ et donc $\mathrm{A}_{i0} \neq \mathrm{A}_{i+1,0}$. Explicitons la définition de $\sigma(i)$. L'espace $\mathrm{A}_{ij} = \mathrm{V}'_{i-1} + \mathrm{V}_j$ est toujours inclus dans $\mathrm{A}_{i+1,j} = \mathrm{V}'_i + \mathrm{V}_j$. On a l'inclusion inverse si $\mathrm{V}'_i \subset \mathrm{V}'_{i-1} + \mathrm{V}_j$. Ceci équivaut à $\mathrm{V}'_i \subset \mathrm{V}'_{i-1} + \mathrm{V}'_i \cap \mathrm{V}_j$, puisque si $x \in \mathrm{V}'_i$ s'écrit $y + z$, avec $(y, z) \in \mathrm{V}'_{i-1} \times \mathrm{V}_j$, alors $z = x - y \in \mathrm{V}'_i \cap \mathrm{V}_j$, et donc à $\mathrm{V}'_i = \mathrm{V}'_{i-1} + \mathrm{V}'_i \cap \mathrm{V}_j$, puisque l'inclusion inverse est toujours réalisée. On obtient les équivalences suivantes :

$$\mathrm{A}_{ij} = \mathrm{A}_{i+1,j} \iff \mathrm{V}'_i = \mathrm{V}'_{i-1} + \mathrm{V}'_i \cap \mathrm{V}_j \iff \mathrm{V}'_i \cap \mathrm{V}_j \not\subset \mathrm{V}'_{i-1},$$

la dernière équivalence résultant de $\dim(\mathrm{V}'_i) = \dim(\mathrm{V}'_{i-1}) + 1$. On peut donc caractériser $\sigma(i)$ par

$$\begin{cases} V'_i \cap V_{\sigma(i)-1} \subset V'_{i-1} \\ V'_i \cap V_{\sigma(i)} \not\subset V'_{i-1}, \end{cases}$$

puisque, si $V'_i \cap V_j \subset V'_{i-1}$, alors $V'_i \cap V_k \subset V'_i \cap V_j \subset V'_{i-1}$, pour tout $0 \leqslant k \leqslant j$.

• Pour montrer que σ est une permutation de $[\![1,n]\!]$, considérons, comme le suggère l'énoncé, l'application $\sigma' : [\![1,n]\!] \longmapsto [\![1,n]\!]$ définie comme σ, mais en échangeant les rôles de d et d' et montrons que σ et σ' sont réciproques l'une de l'autre. Par symétrie des rôles de d et d', il suffit de prouver que $\sigma' \circ \sigma = \mathrm{Id}_{[\![1,n]\!]}$. Soit $i \in [\![1,n]\!]$, $j = \sigma(i)$. Il faut prouver que $\sigma'(j) = i$, c'est-à-dire que

$$\begin{cases} V_j \cap V'_{i-1} \subset V_{j-1} \\ V_j \cap V'_i \not\subset V_{j-1}. \end{cases}$$

⋆ Si $V_j \cap V'_{i-1} \not\subset V_{j-1}$, alors $V_j = V_{j-1} + V_j \cap V'_{i-1}$. Considérons un élément x de $V'_i \cap V_j$. Il s'écrit $y + z$, avec $(y,z) \in V_{j-1} \times V_j \cap V'_{i-1}$. On a alors $y = x - z \in V'_i \cap V_{j-1} \subset V'_{i-1}$, puisque $\sigma(i) = j$; x est donc dans V'_{i-1}. On a donc $V'_i \cap V_j \subset V'_{i-1}$, ce qui est contradictoire avec $\sigma(i) = j$. On conclut que $V_j \cap V'_{i-1} \subset V_{j-1}$.

⋆ Si $V_j \cap V'_i \subset V_{j-1}$, on obtient

$$V_j \cap V'_i \subset V_{j-1} \cap V'_i \subset V'_{i-1}.$$

ce qui est faux. On a donc $V_j \cap V'_i \not\subset V_{j-1}$, ce qui termine la démonstration.

2. Construisons une base $(e_1, \ldots, e_n)$ adaptée à d telle que, pour la permutation σ précédemment définie, on ait $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$ adaptée à d'.

On a, pour $j \in [\![1,n]\!]$, $V_j = V_{j-1} + V_j \cap V'_{\sigma'(j)}$. Considérons, pour tout $j \in [\![1,n]\!]$, un vecteur e_j dans $V_j \cap V'_{\sigma'(j)} \setminus V_{j-1}$. La famille (e_1) est une base de V_1 et si on suppose que $(e_1, \ldots, e_{j-1})$ est une base de V_{j-1}, alors $(e_1, \ldots, e_j)$ est une base de V_j, puisque $\dim(V_j) = \dim(V_{j-1}) + 1$ et que $e_j \in V_j \setminus V_{j-1}$. On a donc montré que $(e_1, \ldots, e_n)$ est une base adaptée à d. On a, d'autre part, pour $i \in [\![1,n]\!]$,

$$e_{\sigma(i)} \in V_{\sigma(i)} \cap V'_{\sigma'\circ\sigma(i)} = V_{\sigma(i)} \cap V'_i \subset V'_i.$$

La famille $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$ est donc une base adaptée à d'.

3. Le groupe $\mathrm{GL}(V)$ agit naturellement sur les drapeaux : pour tout $f \in \mathrm{GL}(V)$ et tout drapeau $d = (V_0, \ldots V_n)$, $(f(V_0), \ldots f(V_n))$ est un drapeau (car f conserve la dimension et l'inclusion), noté $f(d)$. Si (d,d') est un couple de drapeaux, $(f(d), f(d'))$ est un couple de drapeaux. Nous allons montrer que les orbites de couples de drapeaux sous

cette action de groupe sont caractérisées par la valeur de la permutation σ définie dans la question 2.

• Commençons par démontrer que cette permutation est unique (mais pas la base $(e_1, \dots, e_n)$). Soit donc deux drapeaux d et d', une permutation τ de $[\![1, n]\!]$ et une base $(e_1, \dots e_n)$ de V telle que $(e_1, \dots, e_n)$ soit adaptée à d et $(e_{\tau(1)}, \dots, e_{\tau(n)})$ adaptée à d'. On a alors, pour tout $i \in [\![1, n]\!]$,

$$e_{\tau(i)} \in \mathrm{V}_{\tau(i)},\ e_{\tau(i)} \notin \mathrm{V}_{\tau(i)-1},\ e_{\tau(i)} \in \mathrm{V}'_i,\ e_{\tau(i)} \notin \mathrm{V}'_{i-1}.$$

On en déduit que $\mathrm{V}'_i \cap \mathrm{V}_{\tau(i)} \not\subset \mathrm{V}'_{i-1}$, car $e_{\tau(i)}$ appartient au premier ensemble mais pas au second. D'autre part on a

$$\begin{aligned}\mathrm{V}'_i \cap \mathrm{V}_{\tau(i)-1} &= \mathrm{Vect}(e_{\tau(1)}, \dots, e_{\tau(i-1)}, e_{\tau(i)}) \cap \mathrm{Vect}(e_1, \dots, e_{\tau(i)-1}) \\ &\subset \mathrm{Vect}(e_{\tau(1)}, e_{\tau(2)}, \dots, e_{\tau(i-1)}) \subset \mathrm{V}'_{i-1}.\end{aligned}$$

Il résulte de la question 1 que les inclusions

$$\begin{cases} \mathrm{V}'_i \cap \mathrm{V}_{\tau(i)-1} \subset \mathrm{V}'_{i-1} \\ \mathrm{V}'_i \cap \mathrm{V}_{\tau(i)} \not\subset \mathrm{V}'_{i-1} \end{cases}$$

permettent de conclure que $\tau(i) = \sigma(i)$. Ceci est vrai pour tout $i \in [\![1, n]\!]$. Il en résulte que $\tau = \sigma$, ce qui démontre l'unicité voulue.

• Dans la suite, fixons $d = (\mathrm{V}_0, \dots, \mathrm{V}_n)$ et $d' = (\mathrm{V}'_0, \dots, \mathrm{V}'_n)$ deux drapeaux de V, $\mathcal{B} = (e_1, \dots, e_n)$ une base de V, σ la permutation de $[\![1, n]\!]$ telle que $\mathcal{B}$ soit une base adaptée à d et que $(e_{\sigma(1)}, \dots, e_{\sigma(n)})$, que nous noterons désormais $\sigma(\mathcal{B})$, soit une base adaptée à d'.

⋆ Pour tout $f \in \mathrm{GL}(\mathrm{V})$, $\mathcal{B}' = (f(e_1), \dots, f(e_n))$ est une base de V ; pour tout $i \in [\![1, n]\!]$, $(f(e_1), \dots f(e_i))$ est une base de $f(\mathrm{V}_i)$ et $(f(e_{\sigma(1)}), \dots, f(e_{\sigma(i)}))$ est une base de $f(\mathrm{V}'_i)$. Ceci montre que $\mathcal{B}'$ est une base adaptée à $f(d)$ et $\sigma(\mathcal{B}')$ une base adaptée à $f(d')$. La même permutation σ qu'au couple (d, d') est donc associée au couple $(f(d), f(d'))$.

⋆ Réciproquement soit (d_1, d'_1) un couple de drapeaux pour lequel il existe une base $\mathcal{B}'$ de V, adaptée à d_1 et telle que $\sigma(\mathcal{B}')$ soit adaptée à d'_1. Notons $d_1 = (\mathrm{W}_0, \dots, \mathrm{W}_n)$, $d'_1 = (\mathrm{W}'_0, \dots, \mathrm{W}'_n)$ et $\mathcal{B}' = (\varepsilon_1, \dots, \varepsilon_n)$ et considérons $f \in \mathrm{GL}(\mathrm{V})$ défini par $f(e_i) = \varepsilon_i$, pour tout $i \in [\![1, n]\!]$.

On a alors, pour tout $i \in [\![1, n]\!]$,

$$\begin{aligned} f(\mathrm{V}_i) &= \mathrm{Vect}(f(e_1), \dots, f(e_n)) = \mathrm{Vect}(\varepsilon_1, \dots, \varepsilon_i) = \mathrm{W}_i\,. \\ f(\mathrm{V}'_i) &= \mathrm{Vect}(f(e_{\sigma(1)}), \dots, f(e_{\sigma(i)})) = \mathrm{Vect}(\varepsilon_{\sigma(1)}, \dots, \varepsilon_{\sigma(i)}) = \mathrm{W}'_i. \end{aligned}$$

On a donc $f(d, d') = (d_1, d'_1)$. Le couple de drapeaux (d_1, d'_1) appartient donc à l'orbite de (d, d').

On peut donc conclure que deux couples de drapeaux sont dans la même orbite si et seulement si la permutation σ qui leur est associée est

la même. L'ensemble des orbites peut être mis en bijection avec $\mathcal{S}_n$. Il y a donc $n!$ orbites. $\lhd$

Le lecteur trouvera dans l'exercice 7.18 une autre démonstration du résultat de la dernière question par une méthode matricielle utilisant la décomposition de Bruhat.

Nous commençons une série d'exercices assez généraux sur les applications linéaires. Le premier énoncé porte sur les lemmes de factorisation.

6.4. Lemmes de factorisation

Soit E, F et G deux K-espaces vectoriels de dimension finie.

1. Soit $f : \mathrm{E} \to \mathrm{F}$ et $g : \mathrm{E} \to \mathrm{G}$ des applications linéaires. Donner une condition nécessaire et suffisante sur f et g pour qu'il existe $h : \mathrm{F} \longrightarrow \mathrm{G}$ linéaire telle que $g = h \circ f$.

2. Soit $g : \mathrm{E} \longrightarrow \mathrm{G}$ et $h : \mathrm{F} \longrightarrow \mathrm{G}$ des applications linéaires. Donner une condition nécessaire et suffisante sur g et h pour qu'il existe $f : \mathrm{E} \longrightarrow \mathrm{F}$ linéaire telle que $g = h \circ f$.

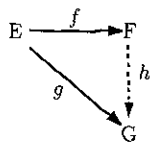
(École polytechnique)

▷ **Solution.**

1. • Condition nécessaire. Supposons qu'il existe $h : \mathrm{F} \longrightarrow \mathrm{G}$ linéaire tel que $g = h \circ f$. Alors pour tout $x \in \operatorname{Ker} f$, on a

$$g(x) = h(f(x)) = h(0) = 0.$$

On obtient donc $\operatorname{Ker} f \subset \operatorname{Ker} g$.



• Supposons réciproquement que $\operatorname{Ker} f \subset \operatorname{Ker} g$. Alors, pour tout $(x, x') \in \mathrm{E}^2$ tel que $f(x) = f(x')$, on a $x - x' \in \operatorname{Ker} f$, donc $x - x' \in \operatorname{Ker} g$ et $g(x) = g(x')$.

Pour $y \in \operatorname{Im} f$, on pose $h'(y) = g(x) \in \mathrm{G}$ où x est un antécédent quelconque de y par f. Cette définition est cohérente, car $g(x)$ ne dépend pas de l'antécédent x choisi, comme on vient de le voir.

Par construction, pour tout $x \in \mathrm{E}$, on a $h'(f(x)) = g(x)$. Ceci montre que $h' \circ f = g$.

Montrons que $h' : \operatorname{Ker} f \longrightarrow \mathrm{G}$ est linéaire. Soit $(y, y') \in (\operatorname{Im} f)^2$, $(\lambda, \mu) \in \mathrm{K}^2$ et $(x, x') \in \mathrm{E}^2$ tel que $f(x) = y$ et $f(x') = y'$. On obtient alors $f(\lambda x + \mu x') = \lambda y + \mu y'$ et $\lambda x + \mu x'$ est un antécédent de $\lambda y + \mu y'$ par f. On a par définition

$$h'(\lambda y + \mu y') = g(\lambda x + \mu x') = \lambda g(x) + \mu g(x') = \lambda h'(y) + \mu h'(y').$$

La linéarité est prouvée.

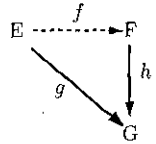
Toute application $h \in \mathcal{L}(\mathrm{F}, \mathrm{G})$ dont la restriction à $\operatorname{Im} f$ est égale à h' (et il en existe) répond alors à la question posée. On remarque que si f est surjective, h est unique.

Conclusion. Il existe $h : \mathrm{F} \longrightarrow \mathrm{G}$ linéaire telle que $g = h \circ f$ si, et seulement si, $\operatorname{Ker} f \subset \operatorname{Ker} g$.

2. • Dans cette seconde question, il est clairement nécessaire que l'on ait $\operatorname{Im} g \subset \operatorname{Im} h$.

• Supposons réciproquement que $\operatorname{Im} g \subset \operatorname{Im} h$. Considérons un supplémentaire F' de $\operatorname{Ker} h$ dans F. Alors h induit un isomorphisme $\tilde{h}$ de F' sur $\operatorname{Im} h$. Notons

$$k = \tilde{h}^{-1} : \operatorname{Im} h \longrightarrow \mathrm{F}'.$$



L'application $f = k \circ g$ est bien définie car $\operatorname{Im} g \subset \operatorname{Im} h$. Elle est linéaire de E dans F' ; nous la considérerons comme une application linéaire de E dans F. On obtient, $h(f(x)) = h(k(g(x))) = g(x)$, pour tout $x \in \mathrm{E}$, *i.e.* $g = h \circ f$.

Conclusion. Il existe $f : \mathrm{E} \longrightarrow \mathrm{F}$ linéaire telle que $g = h \circ f$ si, et seulement si, $\operatorname{Im} g \subset \operatorname{Im} h$. ◁

L'exercice précédent permet de répondre à la question suivante : E *étant un* K*-espace vectoriel,* F *un sous-espace de* E, $s : \mathrm{E} \to \mathrm{E/F}$ *la surjection canonique et* $u \in \mathcal{L}(\mathrm{E})$, *à quelle condition* u *induit-t-elle une application* $\overline{u}$ *de* $\mathcal{L}(\mathrm{E/F})$ *telle que* $\overline{u} \circ s = s \circ u$ *? Comme* s *est surjective, en cas d'existence,* $\overline{u}$ *est unique. D'après la première question de l'exercice,* $\overline{u}$ *existe si et seulement si* $\operatorname{Ker} s \subset \operatorname{Ker}(s \circ u)$. *Comme* $\operatorname{Ker} s = \mathrm{F}$, *cette condition signifie que pour tout* $x \in \mathrm{F}$, $u(x) \in \mathrm{F}$. *Autrement dit,* u *passe au quotient si et seulement si* F *est stable par* u.

Plus généralement, si E *et* F *sont deux espaces vectoriels,* u *un élément de* $\mathcal{L}(\mathrm{E}, \mathrm{F})$, E' *et* F' *deux sous-espaces vectoriels de* E *et* F *respectivement,* u *induit une application linéaire de* $\mathrm{E/E}'$ *dans* $\mathrm{F/F}'$ *si, et seulement si,* $u(\mathrm{E}') \subset \mathrm{F}'$.

6.5. Condition pour que $\operatorname{rg} g \leqslant \operatorname{rg} f$

Soit E, F deux K-espaces vectoriels non nuls de dimension finie, f et g dans $\mathcal{L}(\mathrm{E}, \mathrm{F})$. Montrer que $\operatorname{rg} g \leqslant \operatorname{rg} f$ si et seulement s'il existe $h \in \mathrm{GL}(\mathrm{F})$ et $k \in \mathcal{L}(\mathrm{E})$ tels que $h \circ g = f \circ k$.

(École polytechnique)

▷ **Solution.**

• Une des implications est aisée. En effet, s'il existe $h \in \mathrm{GL(F)}$ et $k \in \mathcal{L}(\mathrm{E})$ tels que $h \circ g = f \circ k$, on a $g = h^{-1} \circ f \circ k$ et donc

$$\mathrm{Im}\, g = \mathrm{Im}(h^{-1} \circ f \circ k) \subset \mathrm{Im}(h^{-1} \circ f) = h^{-1}(\mathrm{Im}\, f),$$

d'où l'on déduit, puisque $h^{-1} \in \mathrm{GL(F)}$,

$$\mathrm{rg}(g) \leqslant \dim(h^{-1}(\mathrm{Im}\, f)) = \dim(\mathrm{Im}\, f) = \mathrm{rg}\, f.$$

• Notons $p = \mathrm{rg}\, f$ et $q = \mathrm{rg}\, g$ et supposons réciproquement que l'on a $q \leqslant p$. On considère $(e_{p+1}, \ldots, e_n)$ une base de $\mathrm{Ker}\, f$ complétée en $(e_1, \ldots, e_n)$ base de E et $(\varepsilon_{q+1}, \ldots, \varepsilon_n)$ une base de $\mathrm{Ker}\, g$ complétée en $(\varepsilon_1, \ldots, \varepsilon_n)$ base de E. La famille $(g(\varepsilon_1), \ldots, g(\varepsilon_q))$ est alors une base de $\mathrm{Im}\, g$ que l'on peut compléter en $(g(\varepsilon_1), \ldots, g(\varepsilon_q), g_{q+1}, \ldots, g_m)$ base de F. De même, $(f(e_1), \ldots, f(e_p))$ est une base de $\mathrm{Im}\, f$ que l'on peut compléter en $(f(e_1), \ldots, f(e_p), f_{p+1}, \ldots, f_m)$ base de F. Soit k dans $\mathcal{L}(\mathrm{E})$ défini par $k(\varepsilon_1) = e_1, \ldots, k(\varepsilon_q) = e_q$, $k(\varepsilon_{q+1}) = \cdots = k(\varepsilon_n) = 0$ et h dans GL(F) qui transforme la base $(g(\varepsilon_1), \ldots, g(\varepsilon_q), g_{q+1}, \ldots, g_m)$ en la base $(f(e_1), \ldots, f(e_p), f_{p+1} \ldots, f_m)$.

On obtient alors, si $1 \leqslant i \leqslant q$, $f \circ k(\varepsilon_i) = f(e_i) = h(g(\varepsilon_i)) = h \circ g(\varepsilon_i)$ (car $i \leqslant q \leqslant p$), et si $i > q$, $f \circ k(\varepsilon_i) = 0 = h \circ g(\varepsilon_i)$. On a donc bien $h \circ g = f \circ k$. ◁

On peut donner une deuxième solution pour la réciproque utilisant un des lemmes de factorisation vus à l'exercice 6.4 : en effet, d'après le second lemme, il suffit de montrer que si $\mathrm{rg}\, g \leqslant \mathrm{rg}\, f$, *on peut trouver* $h \in \mathrm{GL(F)}$ *tel que* $\mathrm{Im}(h \circ g) \subset \mathrm{Im}\, f$. *Pour cela, considérons un sous-espace* G *de* $\mathrm{Im}\, f$ *de dimension* $\mathrm{rg}\, g$ *et un isomorphisme* h' *de* $\mathrm{Im}\, g$ *sur* G. *Si on note* F′ *et* G′ *des supplémentaires de* $\mathrm{Im}\, g$ *et* G *dans* F *respectivement, ces sous-espaces ont même dimension ; soit* h'' *un isomorphisme de* F′ *sur* G′. *L'endomorphisme* h *de* F *dont les restrictions à* $\mathrm{Im}\, g$ *et* F′ *sont respectivement* h' *et* h'' *est un isomorphisme de* F, *qui vérifie*

$$\mathrm{Im}\, h \circ g = h(\mathrm{Im}\, g) = \mathrm{G} \subset \mathrm{Im}\, f.$$

L'exercice 7.1 fournira une troisième solution, matricielle.

Il est bien connu du lecteur qu'un endomorphisme d'un espace E *qui stabilise toutes les droites vectorielles est une homothétie. Ce résultat intervient assez souvent ; aussi nous allons en rappeler la preuve.*

Lemme. *Soit* E *un* K*-espace vectoriel non nul, et* $u \in \mathcal{L}(\mathrm{E})$. *On suppose que pour* $x \in \mathrm{E}$, $u(x)$ *et* x *sont colinéaires. Alors* u *est une homothétie.*

Démonstration. *L'hypothèse signifie que tout vecteur non nul de* E *est un vecteur propre de* u. *Il suffit donc de montrer que* u *admet une unique valeur propre. Supposons qu'il existe deux valeurs propres distinctes* λ *et* λ' *de* u *et considérons deux vecteurs propres* x *et* x' *associés. On a*

alors $u(x) = \lambda x$ *et* $u(x') = \lambda' x'$. *On sait que* (x, x') *est une famille libre. Par hypothèse* $x + x'$ *est aussi un vecteur propre de* u *; il existe un scalaire* μ *tel que* $u(x+x') = \mu(x+x')$. *Mais, par linéarité nous obtenons aussi* $u(x + x') = u(x) + u(x') = \lambda x + \lambda' x'$. *Par unicité de l'écriture, il vient* $\lambda = \lambda' = \mu$, *ce qui est absurde.* $\diamondsuit$

Ce résultat est utile dans bien des situations. Par exemple, si E *est un* K*-espace vectoriel, on peut en déduire que le centre de* $\mathcal{L}(\mathrm{E})$ *est réduit aux homothéties. Les homothéties commutent évidemment avec tout endomorphisme. Réciproquement, supposons que* u *commute avec tout endomorphisme de* E. *Soit* $x \in \mathrm{E}$, H *un supplémentaire de* $\mathrm{K}x$ *et* p *la projection sur* $\mathrm{K}x$ *parallèlement à* H. *On a alors,*

$$u(x) = (u \circ p)(x) = (p \circ u)(x) = p(u(x)) \in \mathrm{K}x$$

et u *est une homothétie d'après le lemme.*

Les deux exercices suivants sont des variations sur le thème du lemme.

6.6. Endomorphismes stabilisant les sous-espaces de dimension k

Soit E un K-espace vectoriel de dimension finie n et $k \in [\![1, n-1]\!]$. Que peut-on dire d'un endomorphisme u de E laissant stables tous les sous-espaces de dimension k de E ?

(École normale supérieure)

▷ **Solution.**

On sait que pour $k = 1$, u est une homothétie (c'est le lemme ci-dessus). Soit $k \in [\![2, n-1]\!]$ et H un sous-espace de dimension $k-1$. On observe que H peut être obtenu comme intersection de deux sous-espaces de dimension k : comme $\dim \mathrm{H} \leqslant n-2$, on peut trouver un couple de vecteurs libres (e_1, e_2) tel que le plan $\mathrm{Vect}(e_1, e_2)$ soit en somme directe avec H. Posons $\mathrm{H}_1 = \mathrm{Vect}(\mathrm{H} \cup \{e_1\})$ et $\mathrm{H}_2 = \mathrm{Vect}(\mathrm{H} \cup \{e_2\})$. On a alors $\mathrm{H}_1 \cap \mathrm{H}_2 = \mathrm{H}$ et $\dim \mathrm{H}_1 = \dim \mathrm{H}_2 = k$.

Il en résulte que si $u \in \mathcal{L}(\mathrm{E})$ stabilise tous les sous-espaces de dimension k alors u stabilise aussi tous les sous-espaces de dimension $k-1$. Par une récurrence descendante finie, il en résulte que u stabilise toutes les droites vectorielles de E. On conclut alors que u est une homothétie. ◁

6.7. Exemple d'utilisation des espaces quotients

Soit E un K-espace vectoriel de dimension finie, $a \in \mathrm{E}$. Déterminer les éléments u de $\mathcal{L}(\mathrm{E})$ tel que pour tout $x \in \mathrm{E}$, la famille $(a, x, u(x))$ soit liée.

(École polytechnique)

▷ **Solution.**

• Si $a = 0$ ou $\dim \mathrm{E} \leqslant 2$, tout u convient. Supposons $a \neq 0$ et $n = \dim \mathrm{E} \geqslant 3$ et considérons u vérifiant la condition de l'énoncé. On pose $e_1 = a$ et on complète en une base $(e_1, \ldots, e_n)$ de E. Si $i \geqslant 2$, (e_i, a) est une famille libre et par hypothèse, $u(e_i) \in \mathrm{Vect}(e_i, e_1)$. Considérons le vecteur $u(e_1 + e_i)$. Comme la famille $(e_1 + e_i, e_1)$ est libre, on a

$$u(e_1 + e_i) \in \mathrm{Vect}(e_1 + e_i, e_1) = \mathrm{Vect}(e_1, e_i).$$

Puisque $u(e_1 + e_i) = u(e_1) + u(e_i)$ et que $u(e_i) \in \mathrm{Vect}(e_1, e_i)$, on en déduit que $u(e_1) \in \mathrm{Vect}(e_1, e_i)$. On obtient

$$u(e_1) \in \mathrm{Vect}(e_1, e_2) \cap \mathrm{Vect}(e_1, e_3) = \mathrm{K}e_1.$$

• On a donc $u(a) \in \mathrm{K}a$: la droite $\mathrm{K}a$ est stable par u, donc u induit un endomorphisme $\overline{u}$ de $\mathrm{E}/\mathrm{K}a$ défini par $\overline{u}(\overline{x}) = \overline{u(x)}$.

Soit $x \in \mathrm{E}$ tel que $\overline{x} \neq 0$. Alors, la famille (a, x) est libre et par hypothèse, $u(x) \in \mathrm{Vect}(a, x)$. Il s'ensuit que $\overline{u}(\overline{x}) = \overline{u(x)} \in \mathrm{Vect}(\overline{x})$. Comme cela vaut pour tout $\overline{x} \in \mathrm{E}/\mathrm{K}a$, $\overline{u}$ est une homothétie : il existe donc $\lambda \in \mathrm{K}$ tel que $\overline{u(x)} = \lambda\overline{x}$ pour tout $x \in \mathrm{E}$. Il s'ensuit que pour tout $x \in \mathrm{E}$, $u(x) - \lambda x \in \mathrm{K}a$. L'application $x \longmapsto u(x) - \lambda x$ étant linéaire, il existe φ forme linéaire de E telle que, pour tout $x \in \mathrm{E}$

$$\boxed{u(x) = \lambda x + \varphi(x)a}.$$

Réciproquement, une telle application répond bien au problème posé. ◁

Les deux exercices suivants concernent des endomorphismes nilpotents.

6.8. Majoration de l'indice de nilpotence

Soit E un K-espace vectoriel de dimension finie n et $u \in \mathcal{L}(\mathrm{E})$, nilpotent. Montrer que $u^n = 0$.

(École normale supérieure)

▷ **Solution.**

Voici une solution complètement élémentaire de cet exercice. Supposons par l'absurde que $u^n \neq 0$. appelons p l'indice de nilpotence de u ($p > n$). L'application u^{p-1} n'étant pas nulle, considérons un vecteur x de E tel que $u^{p-1}(x) \neq 0$. On va montrer que la famille $(x, u(x), \ldots, u^{p-1}(x))$ est libre ce qui fournit notre contradiction puisque cette famille est de cardinal $p > n$. Si la famille est liée, on peut écrire

une relation de liaison du type

$$\lambda_k u^k(x) + \lambda_{k+1} u_{k+1}(x) + \cdots + \lambda_{p-1} u^{p-1}(x) = 0$$

avec $\lambda_k \neq 0$. En appliquant u^{p-1-k} à cette égalité, il vient $\lambda_k u^{p-1}(x) = 0$, ce qui est impossible. L'hypothèse $u^n \neq 0$ est donc absurde. ◁

Bien entendu, il est encore plus rapide de dire que si $u^n \neq 0$, le polynôme minimal de u est X^p $(p > n)$, ce qui est impossible puisque le polynôme minimal est de degré inférieur à n (il divise le polynôme caractéristique par le théorème de Cayley-Hamilton). Enfin le résultat découle aussi directement du résultat de l'exercice 6.14.

Voici une petite application : si $n \geqslant 2$, où n désigne toujours la dimension de l'espace, un endomorphisme nilpotent u d'indice de nilpotence n n'est pas un carré. En effet, si $u = v^2$, alors v est aussi nilpotent et donc $v^n = 0$. Mais alors $u^{n-1} = v^{2n-2} = 0$ car $2n - 2 \geqslant n$ ce qui contredit le fait que u est d'indice n.

L'énoncé suivant est une généralisation du précédent.

6.9. Produit commutatif d'endomorphismes nilpotents

Soit E un K-espace vectoriel de dimension finie n et $u_1, \ldots, u_n$ des endomorphismes nilpotents de E qui commutent deux à deux. Que vaut $u_1 \circ u_2 \circ \cdots \circ u_n$?

(École polytechnique)

▷ **Solution.**

Si tous les u_k sont égaux à u_1, leur produit vaut u_1^n qui est nul, comme nous venons de le démontrer dans l'exercice 6.8. Il est donc raisonnable de penser qu'on a toujours $u_1 \circ u_2 \circ \cdots \circ u_n = 0$.

Nous savons que, lorsque deux endomorphismes commutent, le noyau et l'image de l'un sont stables par l'autre. Ainsi, pour $1 \leqslant k < n$, u_k laisse stable $F_{k+1} = \mathrm{Im}(u_{k+1} \circ \cdots \circ u_n)$. Or si F_{k+1} n'est pas réduit à $\{0\}$, la restriction de u_k à F_{k+1} ne peut être bijective : en effet, cette restriction est aussi nilpotente. Par conséquent, $u_k(F_{k+1})$ est contenu strictement dans F_{k+1} et

$$\dim F_k = \dim u_k(F_{k+1}) < \dim F_{k+1}.$$

Si l'un des F_i est nul c'est terminé et sinon, on a par une récurrence descendante finie que $\dim F_i < i$ pour tout $1 \leqslant i < n$. Ceci implique en particulier $\dim F_1 = \mathrm{rg}(u_1 \circ u_2 \circ \cdots \circ u_n) < 1$, c'est-à-dire

$$\boxed{u_1 \circ u_2 \circ \cdots \circ u_n = 0}. \quad \triangleleft$$

Voici maintenant une série d'exercices assez faciles, centrés sur le théorème du rang. Le premier présente quelques inégalités classiques d'usage courant.

6.10. Inégalité de Sylvester

Soit E un espace vectoriel de dimension finie n, a et b deux endomorphismes de E.

1. Comparer $\mathrm{rg}(a+b)$ à $\mathrm{rg}(a)+\mathrm{rg}(b)$ et $\mathrm{rg}(a)-\mathrm{rg}(b)$.

2. Prouver l'équivalence :

$$\mathrm{rg}(a+b)=\mathrm{rg}(a)+\mathrm{rg}(b) \iff (\mathrm{Im}\,a\cap\mathrm{Im}\,b=\{0\} \text{ et } \mathrm{Ker}\,a+\mathrm{Ker}\,b=\mathrm{E})$$

3. Montrer que $\mathrm{rg}(a)+\mathrm{rg}(b)-n\leqslant\mathrm{rg}(ab)\leqslant\min(\mathrm{rg}(a),\mathrm{rg}(b))$. C'est *l'inégalité de Sylvester*.

(École polytechnique)

▷ **Solution.**

1. De l'inclusion $\mathrm{Im}(a+b)\subset\mathrm{Im}\,a+\mathrm{Im}\,b$, on tire les inégalités

$$\mathrm{rg}(a+b)\leqslant\dim(\mathrm{Im}\,a+\mathrm{Im}\,b)\leqslant\dim(\mathrm{Im}\,a)+\dim(\mathrm{Im}\,b)=\mathrm{rg}(a)+\mathrm{rg}(b). \quad (\star)$$

En appliquant ce résultat à $a+b$ et $-b$, on obtient, compte tenu de l'égalité $\mathrm{rg}(b)=\mathrm{rg}(-b)$,

$$\mathrm{rg}(a)\leqslant\mathrm{rg}(a+b)+\mathrm{rg}(-b)\leqslant\mathrm{rg}(a+b)+\mathrm{rg}(b)$$

c'est-à-dire

$$\mathrm{rg}(a)-\mathrm{rg}(b)\leqslant\mathrm{rg}(a+b).$$

Par symétrie on a $\mathrm{rg}(b)-\mathrm{rg}(a)\leqslant\mathrm{rg}(a+b)$. On conclut que

$$\boxed{|\,\mathrm{rg}(a)-\mathrm{rg}(b)|\leqslant\mathrm{rg}(a+b)\leqslant\mathrm{rg}(a)+\mathrm{rg}(b)}.$$

2. • Supposons que $\mathrm{rg}(a+b)=\mathrm{rg}(a)+\mathrm{rg}(b)$. Toutes les inégalités de $(\star)$ sont alors des égalités. On a en particulier

$$\dim(\mathrm{Im}\,a+\mathrm{Im}\,b)=\dim\mathrm{Im}\,a+\dim\mathrm{Im}\,b.$$

Sachant que $\dim(\mathrm{Im}\,a\cap\mathrm{Im}\,b)=\dim\mathrm{Im}\,a+\dim\mathrm{Im}\,b-\dim(\mathrm{Im}\,a+\mathrm{Im}\,b)$, on en déduit que $\mathrm{Im}\,a\cap\mathrm{Im}\,b=\{0\}$.

Ceci étant réalisé, notons qu'on a ensuite $\mathrm{Ker}(a+b)=\mathrm{Ker}\,a\cap\mathrm{Ker}\,b$. En effet, on a toujours $\mathrm{Ker}\,a\cap\mathrm{Ker}\,b\subset\mathrm{Ker}(a+b)$ et si $x\in\mathrm{Ker}(a+b)$, on peut écrire

$$a(x)=-b(x)\in\mathrm{Im}\,a\cap\mathrm{Im}\,b=\{0\}$$

et $x \in \operatorname{Ker} a \cap \operatorname{Ker} b$. On obtient alors

$$\begin{aligned}\dim(\operatorname{Ker} a + \operatorname{Ker} b) &= \dim(\operatorname{Ker} a) + \dim(\operatorname{Ker} b) - \dim(\operatorname{Ker} a \cap \operatorname{Ker} b)\\ &= \dim(\operatorname{Ker} a) + \dim(\operatorname{Ker} b) - \dim(\operatorname{Ker}(a+b))\\ &= n - \operatorname{rg}(a) + n - \operatorname{rg}(b) - n + \operatorname{rg}(a+b) = n,\end{aligned}$$

puisque $\operatorname{rg}(a+b) = \operatorname{rg}(a) + \operatorname{rg}(b)$.

On a donc $\operatorname{Ker} a + \operatorname{Ker} b = \mathrm{E}$.

• Supposons que $\operatorname{Im} a \cap \operatorname{Im} b = \{0\}$ et $\operatorname{Ker} a + \operatorname{Ker} b = \mathrm{E}$. On a alors, comme il a été démontré précédemment, $\operatorname{Ker}(a+b) = \operatorname{Ker} a \cap \operatorname{Ker} b$. On en déduit

$$\begin{aligned}\operatorname{rg}(a+b) &= n - \dim(\operatorname{Ker}(a+b)) = n - \dim(\operatorname{Ker} a \cap \operatorname{Ker} b)\\ &= n - (\dim \operatorname{Ker} a + \dim \operatorname{Ker} b - \dim(\operatorname{Ker} a + \operatorname{Ker} b)\\ &= n - \dim(\operatorname{Ker} a) + n - \dim(\operatorname{Ker} b), \text{ car } \operatorname{Ker} a + \operatorname{Ker} b = \mathrm{E}\\ &= \operatorname{rg}(a) + \operatorname{rg}(b).\end{aligned}$$

3. De l'inclusion $\operatorname{Im}(ab) \subset \operatorname{Im} a$ résulte $\operatorname{rg}(ab) \leqslant \operatorname{rg} a$. Mais on a aussi $\operatorname{Im}(ab) = a(\operatorname{Im} b)$, d'où l'on déduit $\operatorname{rg}(ab) \leqslant \operatorname{rg}(b)$, car une application linéaire n'augmente pas la dimension des espaces vectoriels (cela résulte du théorème du rang). Finalement, on obtient

$$\operatorname{rg}(ab) \leqslant \min(\operatorname{rg}(a), \operatorname{rg}(b)).$$

Pour obtenir l'autre inégalité, considérons la restriction a' de a à $\operatorname{Im} b$. Elle vérifie $\operatorname{Im}(a') = \operatorname{Im}(ab)$ et $\operatorname{Ker} a' = \operatorname{Ker} a \cap \operatorname{Im} b$. Le théorème du rang donne

$$\begin{aligned}\dim \operatorname{Im}(ab) &= \dim(\operatorname{Im} b) - \dim(\operatorname{Ker} a \cap \operatorname{Im} b)\\ &\geqslant \dim(\operatorname{Im} b) - \dim(\operatorname{Ker} a) \geqslant \operatorname{rg}(b) - (n - \operatorname{rg}(a)).\end{aligned}$$

On obtient l'inégalité voulue :

$$\boxed{\operatorname{rg}(ab) \geqslant \operatorname{rg}(a) + \operatorname{rg}(b) - n}.$$

On peut obtenir d'une autre manière cette inégalité en utilisant un espace vectoriel quotient (cf. page 260 pour des rappels sur la question). Considérons la restriction u de b à $\operatorname{Ker}(ab)$. Son image est incluse dans $\operatorname{Ker} a$, car on a, pour tout $x \in \operatorname{Ker}(ab)$, $ab(x) = 0$ et donc $b(x) \in \operatorname{Ker} a$. On peut donc considérer u comme un élément de $\mathcal{L}(\operatorname{Ker}(ab), \operatorname{Ker} a)$. La restriction de u à $\operatorname{Ker} b$ est évidemment nulle, donc u induit une application linéaire $\overline{u} : \operatorname{Ker}(ab)/\operatorname{Ker} b \longrightarrow \operatorname{Ker} a$ définie par $\overline{u}(\overline{x}) = \overline{u(x)}$. L'application $\overline{u}$ est injective. En effet, pour tout $x \in \operatorname{Ker}(ab)$, on a

$\overline{u}(\overline{x}) = u(x) = b(x)$ et $\overline{u}(\overline{x}) = 0$ si et seulement si $x \in \operatorname{Ker} b$, c'est-à-dire si $\overline{x} = 0$. On en déduit que

$$\dim \operatorname{Ker} a \geqslant \dim \operatorname{Im} \overline{u} = \dim (\operatorname{Ker}(ab)/\operatorname{Ker} b) = \dim \operatorname{Ker}(ab) - \dim \operatorname{Ker} b,$$

c'est-à-dire

$$\dim \operatorname{Ker} a + \dim \operatorname{Ker} b \geqslant \dim \operatorname{Ker}(ab).$$

En utilisant le théorème du rang, on retrouve l'inégalité précédente. ◁

6.11. Pseudo-inverse

Soit E un espace vectoriel de dimension finie.

1. Soit f et g deux endomorphismes de E. On considère les trois conditions :

$$(i)\ f \circ g \circ f = f\ ; \quad (ii)\ g \circ f \circ g = g\ ; \quad (iii)\ \operatorname{rg}(f) = \operatorname{rg}(g).$$

Montrer que si deux de ces conditions sont réalisées, la troisième l'est aussi.

2. Soit f un endomorphisme de E. Montrer l'existence d'un endomorphisme g tel que (i), (ii) et (iii) soient vérifiées.

3. Soit f, g, h, $\tilde{f}$, $\tilde{g}$ des endomorphismes de E tels que

$$f \circ \tilde{f} \circ f = f \quad \text{et} \quad g \circ \tilde{g} \circ g = g.$$

Montrer qu'il existe un endomorphisme u de E tel que $f \circ u \circ g = h$ si, et seulement si, $f \circ \tilde{f} \circ h \circ \tilde{g} \circ g = h$.

(**École polytechnique**)

▷ **Solution.**

1. • Supposons (i) et (ii) vérifiées et montrons (iii). On a grâce à (i)

$$\operatorname{rg}(f) = \operatorname{rg}(f \circ g \circ f) \leqslant \operatorname{rg}(f \circ g) \leqslant \operatorname{rg} g$$

et de même, $\operatorname{rg}(g) = \operatorname{rg}(g \circ f \circ g) \leqslant \operatorname{rg}(g \circ f) \leqslant \operatorname{rg} f$. Par conséquent, on a $\operatorname{rg} f = \operatorname{rg} g$.

• Supposons (i) et (iii) vérifiées et montrons (ii). En composant la relation (i) à droite par g, on obtient $f \circ g = f \circ (g \circ f \circ g)$. Pour pouvoir simplifier par f à gauche, il suffit que la restriction de f à l'image de g soit injective. D'après (i), on a $\operatorname{rg}(f) \leqslant \operatorname{rg}(f \circ g)$. Comme $\operatorname{rg}(f \circ g) \leqslant \operatorname{rg} f$, il y a égalité. Mais le rang de $f \circ g$ est exactement le rang de la restriction de f à $\operatorname{Im} g$. D'où le résultat.

• La symétrie des rôles de f et g permet d'affirmer que (ii) et (iii) entraînent (i).

2. D'après la question 1, il suffit de construire g vérifiant (ii) et (iii). On remarque que si g convient, on a, pour $y \in \operatorname{Im} g$, $g \circ f(y) = y$. En particulier l'image de g doit être en somme directe avec le noyau de f et pour avoir (iii) il doit même s'agir d'un supplémentaire de $\operatorname{Ker} f$.

Soit F un supplémentaire de $\operatorname{Ker} f$. Alors $f' : x \in \mathrm{F} \longmapsto f(x) \in \operatorname{Im} f$ est un isomorphisme. Soit G un supplémentaire de $\operatorname{Im} f$. Il existe un endomorphisme g de E tel que $g_{|\operatorname{Im} f} = f'^{-1}$ et $g_{|\mathrm{G}} = 0$. On a alors l'égalité $\operatorname{Im} g = \operatorname{Im} f'^{-1} = \mathrm{F}$, qui entraine $\operatorname{rg} g = \dim \mathrm{F} = \operatorname{rg} f$: ainsi g vérifie (iii).

D'autre part, si $x \in \mathrm{E}$, $g(x)$ appartient à F. Il en résulte que

$$g \circ f \circ g(x) = g(f(g(x)) = g(f'(g(x))) = (g \circ f')(g(x)) = g(x).$$

On a donc $g \circ f \circ g = g$: g vérifie aussi (ii).

3. Supposons que $f \circ \tilde{f} \circ h \circ \tilde{g} \circ g = h$. Si on pose $u = \tilde{f} \circ h \circ \tilde{g}$, on obtient $f \circ u \circ g = h$.

Réciproquement, s'il existe $u \in \mathcal{L}(\mathrm{E})$ tel que $f \circ u \circ g = h$, on a

$$(f \circ \tilde{f} \circ f) \circ u \circ (g \circ \tilde{g} \circ g) = h$$

et donc $h = f \circ \tilde{f} \circ (f \circ u \circ g) \circ \tilde{g} \circ g = f \circ \tilde{f} \circ h \circ \tilde{g} \circ g$. ◁

6.12. Endomorphismes u tels que $\operatorname{Ker} u = \operatorname{Im} u$

Soit E un espace vectoriel (de dimension quelconque).

1. Soit u un endomorphisme de E tel que $\operatorname{Ker} u = \operatorname{Im} u$ et S un supplémentaire de $\operatorname{Im} u$: $\mathrm{E} = \mathrm{S} \oplus \operatorname{Im} u$.

a. Montrer que, pour tout $x \in \mathrm{E}$, il existe un unique couple $(y, z) \in \mathrm{S}^2$ tel que $x = y + u(z)$. On pose $z = v(x)$ et $y = w(x)$.

b. Montrer que v est linéaire et calculer $u \circ v + v \circ u$.

c. Montrer que w est linéaire et calculer $u \circ w + w \circ u$.

2. Soit $u \in \mathcal{L}(\mathrm{E})$ tel que $u^2 = 0$. On suppose qu'il existe v dans $\mathcal{L}(\mathrm{E})$ tel que $u \circ v + v \circ u = \mathrm{Id}_{\mathrm{E}}$. A-t-on nécessairement $\operatorname{Ker} u = \operatorname{Im} u$?

3. Soit $u \in \mathcal{L}(\mathrm{E})$ tel que $u^2 = 0$ et $u \neq 0$. On suppose qu'il existe $w \in \mathcal{L}(\mathrm{E})$ tel que $u \circ w + w \circ u = u$. A-t-on nécessairement $\operatorname{Ker} u = \operatorname{Im} u$?

(École polytechnique)

▷ **Solution.**

1.a. • Soit $x \in \mathrm{E}$. Comme $\mathrm{E} = \mathrm{S} \oplus \operatorname{Im} u$, il existe $(y, t) \in \mathrm{S} \times \operatorname{Im} u$ tel que $x = y + t$. Comme S est un supplémentaire de $\operatorname{Ker} u$, u induit un isomorphisme de S sur $\operatorname{Im} u$. Il existe donc $z \in \mathrm{S}$ tel que $t = u(z)$. On obtient donc $x = u + u(z)$, avec $(y, z) \in \mathrm{S}^2$.

• Cette écriture est unique. En effet, si $x = y' + u(z')$ avec $(y', z') \in \mathrm{S}^2$,

on a $y-y' = u(z'-z) \in \operatorname{Im} u \cap S$. D'où résultent $y-y' = 0$ et $z'-z \in \operatorname{Ker} u$. Il s'ensuit que $z'-z \in \operatorname{Im} u \cap S = \{0\}$. On a bien $y = y'$ et $z = z'$.

b. et c. • Montrons la linéarité de v et w. Soit x et x' deux vecteurs de E. Écrivons $x = y + u(z)$ et $x' = y' + u(z')$ avec $y = w(x)$, $y' = w(x')$, $z = v(x)$, $z' = v(x')$. Si $\lambda \in \mathrm{K}$, on obtient

$$x + \lambda x' = y + u(z) + \lambda(y' + u(z')) = \underbrace{y + \lambda y'}_{\in S} + u(\underbrace{z + \lambda z'}_{\in S}),$$

puis, par unicité de l'écriture,

$$w(x+\lambda x') = y+\lambda y' = w(x)+\lambda w(x'), \quad v(x+\lambda x') = z+\lambda z' = v(x)+\lambda v(x').$$

Les applications v et w sont donc linéaires.

• On écrit $x = y + u(z)$ avec $(y, z) \in S^2$ et on calcule $(u \circ v + v \circ u)(x)$. On obtient

$$\begin{aligned}(u \circ v + v \circ u)(x) &= u(z) + v(u(x)) = u(z) + v(u(y)), \text{ car } u(z) \in \operatorname{Ker} u,\\ &= u(z) + y = x \text{ , par définition de } v \ (y \in S).\end{aligned}$$

On conclut que $\boxed{u \circ v + v \circ u = \mathrm{Id}_E}$.

• On calcule de même $(u \circ w + w \circ u)(x)$ et on obtient

$$\begin{aligned}(u \circ w + w \circ u)(x) &= u(y) + w(u(x)) = u(y) + w(u(y))\\ &= u(y) + 0, \text{ par définition de } w,\\ &= u(x - u(z)) = u(x), \text{ car } u(z) \in \operatorname{Ker} u.\end{aligned}$$

On conclut que $\boxed{u \circ w + w \circ u = u}$.

2. La réponse est affirmative. De $u^2 = 0$, on déduit $\operatorname{Im} u \subset \operatorname{Ker} u$. Si $x \in \operatorname{Ker} u$, on peut écrire $x = u(v(x)) + v(u(x)) = u(v(x)) \in \operatorname{Im} u$. On a donc $\operatorname{Ker} u = \operatorname{Im} u$.

3. Par contre ici, la réponse est négative. Considérons $u \in \mathcal{L}(\mathbb{R}^3)$ défini par $u(e_1) = 0$, $u(e_2) = 0$ et $u(e_3) = e_2$, où (e_1, e_2, e_3) désigne la base canonique de $\mathbb{R}^3$. Alors $\operatorname{Im} u$ est strictement incluse dans $\operatorname{Ker} u$ et pourtant si $w = \dfrac{1}{2}\,\mathrm{Id}_{\mathbb{R}^3}$, on a bien $u \circ w + w \circ u = u$. ◁

6.13. Endomorphismes u tels que $\operatorname{Ker} u \oplus \operatorname{Im} u = \mathrm{E}$

Soit E un K-espace vectoriel de dimension finie et $u \in \mathcal{L}(\mathrm{E})$. Trouver une condition nécessaire et suffisante sur u pour qu'il existe $v \in \mathcal{L}(\mathrm{E})$ tel que $u \circ v = 0$ et $u + v$ inversible.

(École polytechnique)

▷ **Solution.**

• Analyse. Supposons que $v \in \mathcal{L}(\mathrm{E})$ réponde au problème posé. On

obtient $\operatorname{Im} v \subset \operatorname{Ker} u$ et donc $\operatorname{rg} v \leqslant n - \operatorname{rg} u$ et $n = \operatorname{rg}(u+v) \leqslant \operatorname{rg} u + \operatorname{rg} v$. On a donc $n = \operatorname{rg} u + \operatorname{rg} v$ et par suite $\operatorname{Im} v = \operatorname{Ker} u$, et $\operatorname{Im} u + \operatorname{Im} v = \mathrm{E}$ (car $\mathrm{E} = \operatorname{Im}(u+v) \subset \operatorname{Im} u + \operatorname{Im} v$) . Cette somme est directe et on conclut que

$$\boxed{\operatorname{Ker} u \oplus \operatorname{Im} u = \mathrm{E}}$$

• Réciproquement, supposons $\operatorname{Ker} u \oplus \operatorname{Im} u = \mathrm{E}$. Soit v le projecteur sur $\operatorname{Ker} u$ parallèlement à $\operatorname{Im} u$. On a évidemment $u \circ v = 0$. Montrons que de plus, $u + v \in \mathrm{GL}(\mathrm{E})$. Soit $x \in \mathrm{E}$ tel que $(u+v)(x) = 0$. On a alors $u(x) = -v(x) \in \operatorname{Ker} u \cap \operatorname{Im} u = \{0\}$. Donc $u(x) = v(x) = 0$ et $x \in \operatorname{Ker} u \cap \operatorname{Im} u = \{0\}$. On a donc $\operatorname{Ker}(u+v) = \{0\}$ et $u + v \in \mathrm{GL}(\mathrm{E})$.

Conclusion. Une condition nécessaire et suffisante d'existence de v est $\operatorname{Im} u \oplus \operatorname{Ker} u = \mathrm{E}$. ◁

L'exercice 6.15 donnera d'autres conditions équivalentes à celle-ci.

L'énoncé suivant établit des résultats importants qui sont à la base de la réduction de Jordan des endomorphismes nilpotents.

6.14. Décomposition de Fitting

Soit E un K-espace vectoriel de dimension finie n et u un endomorphisme de E.

1. Montrer que les suites $(\operatorname{Im} u^k)_{k\in\mathbb{N}}$ et $(\operatorname{Ker} u^k)_{k\in\mathbb{N}}$ sont d'abord strictement monotones pour l'inclusion puis constantes à partir d'un même rang $p \leqslant n$.

2. Montrer que la suite $(\operatorname{Ker} u^k)$ «s'essouffle», c'est-à-dire que la suite $(\dim \operatorname{Ker} u^{k+1} - \dim \operatorname{Ker} u^k)_{k\geqslant 0}$ est décroissante.

3. Démontrer que $\mathrm{E} = \operatorname{Ker} u^p \oplus \operatorname{Im} u^p$.

4. En déduire que toute matrice de $\mathcal{M}_n(\mathrm{K})$ est semblable à une matrice de la forme $\begin{pmatrix} \boxed{\mathrm{N}} & 0 \\ 0 & \boxed{\mathrm{C}} \end{pmatrix}$, où N est une matrice carrée nilpotente et C une matrice carrée inversible.

(École normale supérieure)

▷ **Solution.**

1. On a évidemment $\operatorname{Ker} u^k \subset \operatorname{Ker} u^{k+1}$ et $\operatorname{Im} u^{k+1} \subset \operatorname{Im} u^k$ pour tout entier naturel k. La suite $(\operatorname{Ker} u^k)_{k\in\mathbb{N}}$ est donc croissante pour l'inclusion et la suite $(\operatorname{Im} u^k)_{k\in\mathbb{N}}$ décroissante.

La suite d'entiers naturels $(\dim \operatorname{Ker} u^k)_{k\in\mathbb{N}}$ est croissante et majorée par n, donc est constante à partir d'un certain rang. La suite $(\operatorname{Ker} u^k)_{k\geqslant 0}$ est donc stationnaire. Soit p le plus petit entier tel que $\operatorname{Ker} u^p = \operatorname{Ker} u^{p+1}$. On montre que $\operatorname{Ker} u^{k+1} = \operatorname{Ker} u^k$, pour tout $k \geqslant p$. Soit $k \geqslant p$ et x dans $\operatorname{Ker} u^{k+1}$. Alors $u^{k-p}(x)$ est dans le noyau de u^{p+1}

et donc dans celui de u^p. Ainsi $u^p(u^{k-p}(x)) = u^k(x) = 0$ et $x \in \operatorname{Ker} u^k$. On a donc $\operatorname{Ker} u^{k+1} \subset \operatorname{Ker} u^k$ et donc $\operatorname{Ker} u^{k+1} = \operatorname{Ker} u^k$. La suite $(\operatorname{Ker} u^k)_{k\geqslant 0}$ est d'abord strictement croissante, puis constante à partir du rang p. En particulier, on a nécessairement $p \leqslant n$.

D'après le théorème du rang $\dim \mathrm{E} = \dim \operatorname{Ker} u^k + \dim \operatorname{Im} u^k$ pour tout k. Il en résulte que $\operatorname{rg} u^k - \operatorname{rg} u^{k+1} = \dim \operatorname{Ker} u^{k+1} - \dim \operatorname{Ker} u^k$ et donc que la suite des images est d'abord strictement décroissante, puis constante à partir du rang p.

2. La preuve la plus courte consiste à utiliser les espaces quotients (cf. p. 260 et 267 pour des rappels sur la question). On a, pour tout entier naturel k, $u(\operatorname{Ker} u^{k+2}) \subset \operatorname{Ker} u^{k+1}$ et $u(\operatorname{Ker} u^{k+1}) \subset \operatorname{Ker} u^k$. On en déduit que u induit une application linéaire $\overline{u}$ de $\operatorname{Ker} u^{k+2}/\operatorname{Ker} u^{k+1}$ dans $\operatorname{Ker} u^{k+1}/\operatorname{Ker} u^k$. Ces deux espaces vectoriels étant de dimensions respectives $\dim \operatorname{Ker} u^{k+2} - \dim \operatorname{Ker} u^{k+1}$ et $\dim \operatorname{Ker} u^{k+1} - \dim \operatorname{Ker} u^k$, il suffit pour conclure de montrer que $\overline{u}$ est injective. Soit $\overline{x} \in \operatorname{Ker} u^{k+2}/\operatorname{Ker} u^{k+1}$ tel que $\overline{u}(\overline{x}) = 0$. Cela signifie que $u(x) \in \operatorname{Ker} u^k$, donc que $x \in \operatorname{Ker} u^{k+1}$ et $\overline{x} = 0$.

Le lecteur qui veut éviter l'utilisation des quotients pourra rédiger la solution en introduisant un supplémentaire H de $\operatorname{Ker} u^{k+1}$ dans $\operatorname{Ker} u^{k+2}$ et montrer que la restriction de u à H est une injection de H sur un sous-espace de $\operatorname{Ker} u^{k+1}$ qui est en somme directe avec $\operatorname{Ker} u^k$.

3. D'après le théorème du rang, il suffit de montrer la somme est directe. Soit $y \in \operatorname{Ker} u^p \cap \operatorname{Im} u^p$ et $x \in \mathrm{E}$ tel que $y = u^p(x)$. On a alors $u^p(y) = u^{2p}(x) = 0$ donc $x \in \operatorname{Ker} u^{2p} = \operatorname{Ker} u^p$. Par conséquent, $y = u^p(x) = 0$. La somme est bien directe et

$$\boxed{\mathrm{E} = \operatorname{Ker} u^p \oplus \operatorname{Im} u^p}.$$

4. Soit $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$. On applique le résultat précédent à l'endomorphisme de K^n canoniquement associé à A. L'entier p étant défini à partir de cet endomorphisme comme dans la question 1, on a l'égalité $\mathrm{K}^n = \operatorname{Im} \mathrm{A}^p \oplus \operatorname{Ker} \mathrm{A}^p$. Les sous-espaces $\mathrm{F} = \operatorname{Ker} \mathrm{A}^p$ et $\mathrm{G} = \operatorname{Im} \mathrm{A}^p$ sont stables par A, car on a $\mathrm{A}(\operatorname{Ker} \mathrm{A}^p) \subset \operatorname{Ker} \mathrm{A}^{p-1} \subset \operatorname{Ker} \mathrm{A}^p$ et $\mathrm{A}(\operatorname{Im} \mathrm{A}^p) = \operatorname{Im} \mathrm{A}^{p+1} = \operatorname{Im} \mathrm{A}^p$. La matrice de A dans une base de K^n obtenue par juxtaposition d'une base de $\operatorname{Ker} \mathrm{A}^p$ et d'une base de $\operatorname{Im} \mathrm{A}^p$ est de la forme $\begin{pmatrix} \boxed{\mathrm{N}} & 0 \\ 0 & \boxed{\mathrm{C}} \end{pmatrix}$, où N et C sont des matrices carrées. La restriction de A à $\operatorname{Ker} \mathrm{A}^p$ est nilpotente d'indice p et la restriction de A à $\operatorname{Im} \mathrm{A}^p$ est surjective (car $\mathrm{A}(\operatorname{Im} \mathrm{A}^p) = \operatorname{Im} \mathrm{A}^p$), donc bijective. Ainsi, N est nilpotente et C inversible. ◁

L'exercice qui suit fournit diverses caractérisations des endomorphismes pour lesquels $p = 2$.

6.15. Endomorphismes tels que $\mathbf{E = \operatorname{Ker} u \oplus \operatorname{Im} u}$

Soit E de dimension finie sur K et $u \in \mathcal{L}(E)$.

1. Montrer qu'il y a équivalence entre :

(i) $\operatorname{Ker} u = \operatorname{Ker} u^2$;

(ii) $\operatorname{Im} u = \operatorname{Im} u^2$;

(iii) $E = \operatorname{Ker} u \oplus \operatorname{Im} u$.

2. Donner des exemples d'endomorphismes vérifiant ces conditions.

3. Le résultat subsiste-il en dimension infinie ?

(École polytechnique)

▷ **Solution.**

1. Notons que l'on a $\operatorname{Im} u^2 \subset \operatorname{Im} u$ et $\operatorname{Ker} u \subset \operatorname{Ker} u^2$, pour tout $u \in \mathcal{L}(E)$. En dimension finie, les propositions (i) et (ii) sont donc équivalentes puisque d'après le théorème du rang

$$\dim(\operatorname{Ker} u) + \dim(\operatorname{Im} u) = \dim(\operatorname{Ker} u^2) + \dim(\operatorname{Im} u^2) = \dim(E).$$

• Supposons (i) et montrons (iii). Soit $y \in \operatorname{Im} u \cap \operatorname{Ker} u$. Il existe $x \in E$ tel que $y = u(x)$. Comme $u(y) = 0$, on a $u^2(x) = 0$ et $x \in \operatorname{Ker} u^2 = \operatorname{Ker} u$. Ainsi $y = u(x) = 0$. Donc $\operatorname{Im} u$ et $\operatorname{Ker} u$ sont en somme directe. Comme le théorème du rang assure que $\dim \operatorname{Im} u + \dim \operatorname{Ker} u = \dim E$, on conclut que $E = \operatorname{Ker} u \oplus \operatorname{Im} u$.

• Supposons (iii) et montrons (ii). Soit $y \in \operatorname{Im} u$ et $x \in E$ tel que $u(x) = y$. Par hypothèse, x s'écrit $x = x' + x''$ avec $x' \in \operatorname{Im} u$ et $x'' \in \operatorname{Ker} u$. Il existe donc $z \in E$ tel que $x' = u(z)$ et $x = u(z) + x''$. D'où $y = u(u(z)) + u(x'') = u^2(z) \in \operatorname{Im} u^2$. On a donc $\operatorname{Im} u \subset \operatorname{Im} u^2$ et finalement (ii), puisque l'autre inclusion est toujours vérifiée.

Cela montre donc l'équivalence entre les trois propositions.

2. Tout projecteur, tout isomorphisme de E, tout endomorphisme diagonalisable vérifie ces conditions.

3. En dimension infinie, les trois propriétés ne sont pas équivalentes.

• Considérons $E = \mathbb{R}[X]$ et prenons comme endomorphisme u de E la dérivation. Comme u est surjective, $\operatorname{Im} u = \operatorname{Im} u^2 = E$: (ii) est vérifié. En revanche, $\operatorname{Ker} u = \mathbb{R}_0[X]$ et $\operatorname{Ker} u^2 = \mathbb{R}_1[X]$: (i) n'est pas vérifié. La propriété (iii) ne l'est pas non plus car $\operatorname{Ker} u \subset \operatorname{Im} u$ et $\operatorname{Ker} u \neq \{0\}$.

• Soit v l'endomorphisme $P \longmapsto XP$ de E. On a $\operatorname{Ker} v = \operatorname{Ker} v^2 = \{0\}$, $\operatorname{Im} v^2 = X^2\mathbb{R}[X] \neq \operatorname{Im} v = X\mathbb{R}[X]$; cette fois la condition (i) est remplie mais ni (ii), ni (iii) ne sont vérifiées.

• Par contre on peut remarquer qu'en dimension quelconque, (iii) est équivalent à (i) et (ii).

⋆ Supposons (i) et (ii). Comme précédemment, on montre que la somme $\operatorname{Ker} u + \operatorname{Im} u$ est directe. Montrons avec (ii) qu'elle fait E tout

entier. Si $x \in \mathrm{E}$, alors $u(x) \in \operatorname{Im} u = \operatorname{Im} u^2$ et il existe $y \in \mathrm{E}$ tel que $u(x) = u^2(y)$. On en déduit que $u(x - u(y)) = 0$ et $x - u(y) \in \operatorname{Ker} u$. On obtient finalement $x \in \operatorname{Im} u + \operatorname{Ker} u$ et $\mathrm{E} = \operatorname{Ker} u \oplus \operatorname{Im} u$.

⋆ Si (*iii*) est vérifié, on a déjà montré que $\operatorname{Im} u = \operatorname{Im} u^2$. Soit x dans $\operatorname{Ker} u^2$. On a $u(u(x)) = 0$, c'est-à-dire $u(x) \in \operatorname{Im} u \cap \operatorname{Ker} u = \{0\}$. On en déduit que $x \in \operatorname{Ker} u$. On a montré que $\operatorname{Ker} u^2 \subset \operatorname{Ker} u$ et donc $\operatorname{Ker} u^2 = \operatorname{Ker} u$. ◁

Dans l'énoncé suivant on détermine de manière très élémentaire les puissances d'un endomorphisme diagonalisable ayant deux valeurs propres.

6.16. Endomorphisme annulé par un polynôme de degré 2 à racines simples

Soit E un K-espace vectoriel de dimension finie et $f \in \mathcal{L}(\mathrm{E})$ vérifiant $(f - a\,\mathrm{Id})(f - b\,\mathrm{Id}) = 0$ où a et b sont deux éléments distincts de K.

1. Établir l'existence de λ et μ non nuls tels que $\lambda(f - a\,\mathrm{Id})$ et $\mu(f - b\,\mathrm{Id})$ soient des projecteurs.

2. Montrer que $\operatorname{Im}(f - b\,\mathrm{Id}) = \operatorname{Ker}(f - a\,\mathrm{Id})$.

3. Calculer f^n pour tout $n \in \mathbb{N}$.

4. Si $ab \neq 0$, montrer que $f \in \mathrm{GL}(\mathrm{E})$, et calculer f^n pour $n \in \mathbb{Z}$.

(**École polytechnique**)

▷ **Solution.**

1. On a par hypothèse $f^2 - (a+b)f + ab\,\mathrm{Id} = 0$. Si $\lambda \in \mathrm{K}^*$, il suffit, pour que $\lambda(f - a\,\mathrm{Id})$ soit un projecteur, que $[\lambda(f - a\,\mathrm{Id})]^2 = \lambda(f - a\,\mathrm{Id})$. C'est le cas dès que $\lambda(f^2 - 2af + a^2\,\mathrm{Id}) = f - a\,\mathrm{Id}$. En remplaçant f^2 par $(a+b)f - ab\,\mathrm{Id}$, on obtient

$$\lambda[(b-a)f - a(a-b)\,\mathrm{Id}] = f - a\,\mathrm{Id}\,.$$

Comme $a \neq b$, cette identité est vérifiée pour $\boxed{\lambda = \dfrac{1}{b-a}}$ et dans ce cas, $\lambda(f - a\,\mathrm{Id})$ est un projecteur. Par symétrie du problème, $\mu(f - b\,\mathrm{Id})$ est un projecteur pour $\boxed{\mu = \dfrac{1}{a-b}}$.

2. Comme $(f - a\,\mathrm{Id})(f - b\,\mathrm{Id}) = 0$, on a $\operatorname{Im}(f - b\,\mathrm{Id}) \subset \operatorname{Ker}(f - a\,\mathrm{Id})$. Réciproquement, si $x \in \operatorname{Ker}(f - a\,\mathrm{Id})$, on a $f(x) = ax$, d'où l'on déduit que $f(x) - bx = (a-b)x$ et donc $x = \dfrac{1}{b-a}(f(x) - bx) \in \operatorname{Im}(f - b\,\mathrm{Id})$. On conclut que $\operatorname{Im}(f - b\,\mathrm{Id}) = \operatorname{Ker}(f - a\,\mathrm{Id})$.

3. Écrivons le reste de la division euclidienne de X^n par $(X-a)(X-b)$ sous la forme $\alpha(X-a)+\beta(X-b)$ avec $(\alpha,\beta)\in K^2$ (c'est possible car $a\neq b$). On note Q le quotient. En substituant a, puis b à X, il vient

$$a^n=\beta(a-b) \quad \text{et} \quad b^n=\alpha(b-a).$$

En substituant f à X, on obtient

$$\begin{aligned} f^n &= Q(f)(f-a\,\mathrm{Id})(f-b\,\mathrm{Id})+\alpha(f-a\,\mathrm{Id})+\beta(f-b\,\mathrm{Id}) \\ &= \alpha(f-a\,\mathrm{Id})+\beta(f-b\,\mathrm{Id}) \end{aligned}$$

et donc

$$\boxed{f^n=\frac{1}{b-a}\left(b^n(f-a\,\mathrm{Id})-a^n(f-b\,\mathrm{Id})\right)}.$$

4. • Si $ab\neq 0$, l'égalité $f^2-(a+b)f+ab\,\mathrm{Id}=0$ devient

$$f\left[\frac{1}{ab}\left((a+b)\,\mathrm{Id}-f\right)\right]=\mathrm{Id}$$

et f est donc un isomorphisme de E.

• On pose $g=f^{-1}$. À partir de l'égalité $(f-a\,\mathrm{Id})(f-b\,\mathrm{Id})=0$, on obtient l'égalité $(g-a'\,\mathrm{Id})(g-b'\,\mathrm{Id})=0$, où $a'=\dfrac{1}{a}$ et $b'=\dfrac{1}{b}$. C'est une condition analogue à celle vérifiée par f. De la question précédente, on déduit que, pour $n\geqslant 1$,

$$\begin{aligned} g^n &= \frac{1}{b'-a'}\left(b'^n(g-a'\,\mathrm{Id})-a'^n(g-b'\,\mathrm{Id})\right) \\ &= f^{-1}\frac{1}{a-b}\left(b^{-n+1}(a\,\mathrm{Id}-f)-a^{-n+1}(b\,\mathrm{Id}-f)\right), \end{aligned}$$

puis, en multipliant par f

$$f^{1-n}=(f^{-1})^{n-1}=\frac{1}{b-a}\left(b^{1-n}(f-a\,\mathrm{Id})-a^{1-n}(f-b\,\mathrm{Id})\right).$$

On trouve la même formule pour les exposants négatifs que pour $n\geqslant 0$. Finalement, on obtient, pour $n\in\mathbb{Z}$,

$$\boxed{f^n=\frac{1}{b-a}\left(b^n(f-a\,\mathrm{Id})-a^n(f-b\,\mathrm{Id})\right)}. \quad \triangleleft$$

6.17. Équation linéaire dans $\mathcal{L}(E)$

Soit E un K-espace vectoriel.

1. Soit $(f,g) \in \mathcal{L}(E)^2$ tel que $f \circ g - g \circ f = \mathrm{Id}$. Vérifier que, pour tout $P \in K[X]$, on a $f \circ P(g) - P(g) \circ f = P'(g)$. Si K est un sous-corps de $\mathbb{C}$, montrer que $(g^n)_{n\geqslant 0}$ est une famille libre.

2. Donner, lorsque $K = \mathbb{R}$ et $E = \mathbb{R}[X]$, un exemple d'endomorphismes f et g de E vérifiant $f \circ g - g \circ f = \mathrm{Id}$.

3. On suppose que $K = \mathbb{Z}/p\mathbb{Z}$, où p est un nombre premier, et que E est de dimension p. Soit $(e_1, e_2, \ldots, e_p)$ une base de E et $g \in \mathcal{L}(E)$ tel que $g(e_i) = e_{i+1}$ si $1 \leqslant i \leqslant p-1$ et $g(e_p) = 0$.

Déterminer les endomorphismes f de E tels que $f \circ g - g \circ f = \mathrm{Id}$.

(École polytechnique)

▷ **Solution.**

1. Par linéarité, il suffit de prouver le résultat pour $P = X^n$. Procédons par récurrence sur $n \in \mathbb{N}$.

- C'est trivial si $n = 0$, c'est l'hypothèse pour $n = 1$.
- Supposons $n \geqslant 2$. D'après l'hypothèse de récurrence, on a

$$f \circ g^{n-1} - g^{n-1} \circ f = (n-1)g^{n-2}.$$

Composant à droite par g, on obtient

$$\begin{aligned}(n-1)g^{n-1} &= f \circ g^n - g^{n-1} \circ f \circ g = f \circ g^n - g^{n-1} \circ (g \circ f + \mathrm{Id}) \\ &= f \circ g^n - g^n \circ f - g^{n-1},\end{aligned}$$

c'est-à-dire $f \circ g^n - g^n \circ f = ng^{n-1}$.

Montrons en raisonnant par l'absurde que la famille $(g^n)_{n\geqslant 0}$ est libre. Supposons qu'est liée. Il existe donc $P \in K[X]$, non nul, tel que $P(g) = 0$. Choisissons P de degré minimal ; on a nécessairement $\deg P \geqslant 1$, car g n'est pas nul. Alors, il résulte de la formule que nous venons d'établir que

$$P'(g) = f \circ P(g) - P(g) \circ f = 0.$$

Or $P' \neq 0$ (on est sur un sous-corps de $\mathbb{C}$) et $\deg P' < \deg P$, ce qui contredit la minimalité de $\deg P$. Donc $(g^n)_{n\geqslant 0}$ est libre.

En particulier, si K *est un sous-corps de* $\mathbb{C}$ *(ou plus généralement un corps de caractéristique nulle), l'équation* $f \circ g - g \circ f = \mathrm{Id}$ *n'a pas de solution si* E *est de dimension finie. On peut aussi voir cela à l'aide de la trace : si* $\dim E = n$, $\mathrm{Tr}(\mathrm{Id}) = n$ *mais* $\mathrm{Tr}(f \circ g - g \circ f) = 0$.

2. Il s'agit de trouver f et g vérifiant

$$f \circ g = g \circ f + \mathrm{Id}\,.$$

Or si $P \in \mathbb{R}[X]$, on a $(XP)' = XP' + P$. Il suffit donc de prendre

$$f : P \longmapsto P' \quad \text{et} \quad g : X \longmapsto XP.$$

3. L'équation d'inconnue $f \in \mathcal{L}(E)$ est en fait une équation linéaire. Si f_0 est une solution particulière, on obtient les autres en additionnant à f_0 une solution de l'équation homogène $f \circ g - g \circ f = 0$. Or le sous-espace des solutions de l'équation homogène est $\mathcal{C}(g)$, le commutant de g. On va donc chercher une solution particulière, puis déterminer le commutant de g.

• Si f est solution, on obtient $g(f(e_p)) = f(g(e_p)) - e_p = -e_p$. C'est le cas par exemple si $f(e_p) = -e_{p-1}$. On obtient ensuite

$$g(f(e_{p-1})) = f(g(e_{p-1})) - e_{p-1} = f(e_p) - e_{p-1} = -2e_{p-1}.$$

On peut choisir $f(e_{p-1}) = -2e_{p-2}$. En continuant, on constate que

$$f(e_i) = -(p-i+1)e_{i-1} = (i-1)e_{i-1} \text{ pour } 2 \leqslant i \leqslant p \quad \text{et} \quad f(e_1) = 0$$

est un bon choix.

Procédons à la synthèse. Soit donc $f_0 \in \mathcal{L}(E)$ défini, pour $1 \leqslant i \leqslant p$, par $f_0(e_i) = (i-1)e_{i-1}$. Vérifions que f_0 convient. On obtient, pour tout $1 \leqslant i \leqslant p-1$,

$$(f_0 \circ g - g \circ f_0)(e_i) = f_0(e_{i+1}) - g((i-1)e_{i-1}) = ie_i - (i-1)e_i = e_i$$

et

$$(f_0 \circ g - g \circ f_0)(e_p) = -g(f_0(e_p)) = -g((p-1)e_{p-1}) = -(p-1)e_p = e_p,$$

car dans $\mathbb{Z}/p\mathbb{Z}$, $-(p-1) = 1$. On a donc $f_0 \circ g - g \circ f_0 = \mathrm{Id}$.

• Déterminons le commutant $\mathcal{C}(g)$ de g. Les initiés auront reconnu en g un endomorphisme cyclique. En effet, la famille $(e_1, g(e_1), \ldots, g^{p-1}(e_1))$, c'est-à-dire $(e_1, e_2, \ldots, e_p)$ est une base de E. Le commutant de g est alors confondu avec $K[g] = \{P(g), P \in K[X]\}$. Redémontrons ce résultat. On a déjà clairement $K[g] \subset \mathcal{C}(g)$. Inversement, montrons que si $f \in \mathcal{C}(g)$, il existe un unique $(\lambda_0, \ldots, \lambda_{p-1}) \in K^p$ tel que

$$f = \lambda_0 \mathrm{Id} + \lambda_1 g + \cdots + \lambda_{p-1} g^{p-1}. \tag{1}$$

En appliquant à e_1, on obtient

$$f(e_1) = \lambda_0 e_1 + \lambda_1 e_2 + \cdots + \lambda_{p-1} e_p. \tag{2}$$

Cette dernière égalité définit $(\lambda_0, \ldots, \lambda_p)$ de manière unique. Il s'agit de démontrer que l'égalité (1) est réalisée pour cette valeur de $(\lambda_0, \ldots, \lambda_p)$.

Il suffit de le vérifier pour les vecteurs de $e_1, g(e_1), \ldots, g_{p-1}(e_1))$ de E, c'est-à-dire de montrer que, pour $0 \leqslant i \leqslant p-1$,

$$\begin{aligned} f(g^i(e_1)) &= \lambda_0 g^i(e_1) + \lambda_1 g^{i+1}(e_1) + \cdots + \lambda_{p-1} g^{i+p-1}(e_1) \\ &= \lambda_0 g^i(e_1) + \cdots + \lambda_{p-1} g^i(e_p). \end{aligned}$$

Nous savons que c'est vrai pour $i = 0$; c'est l'égalité (2). En calculant l'image par g^i des deux membres de l'égalité (2) et en tenant compte du fait que f commute avec g donc avec g^i, on obtient

$$\begin{aligned} f(g^i(e_1)) = g^i(f(e_1)) &= g^i(\lambda_0 e_1 + \lambda_1 e_2 + \cdots + \lambda_{p-1} e_p) \\ &= \lambda_0 g^i(e_1) + \cdots + \lambda_{p-1} g^i(e_p). \end{aligned}$$

On obtient donc que f est dans $\mathrm{Vect}(\mathrm{Id}, g, \ldots, g^{p-1})$.

Nous avons démontré les inclusions

$$\mathcal{C}(g) \subset \mathrm{Vect}(\mathrm{Id}, g, \ldots, g^{p-1}) \subset \mathrm{K}[g] \subset \mathcal{C}(g),$$

qui sont donc des égalités. L'unicité de $(\lambda_0, \ldots, \lambda_p)$ pour tout $f \in \mathcal{C}(g)$ montre que $(\mathrm{Id}, g, \ldots, g^{p-1})$ est une base de $\mathcal{C}(g)$ qui est donc de dimension p.

Conclusion. L'ensemble des f qui conviennent est

$$f_0 + \mathrm{Vect}(\mathrm{Id}, g, \ldots, g^{p-1}). \lhd$$

Voici quelques exercices consacrés aux projecteurs. On rappelle que les projecteurs sont les idempotents de l'algèbre $\mathcal{L}(\mathrm{E})$, c'est-à-dire les endomorphismes p vérifiant $p^2 = p$. Leur importance provient des rapports étroits qu'ils entretiennent avec les décompositions de E *en somme directe de sous-espaces.*

6.18. Projecteurs

Soit p et q deux projecteurs d'un espace vectoriel E tels que $\mathrm{Im}\, p \subset \mathrm{Ker}\, q$ et $r = p + q - pq$. Montrer que r est un projecteur et trouver son image et son noyau.

(École polytechnique)

▷ **Solution.**

- Comme $\mathrm{Im}\, p \subset \mathrm{Ker}\, q$, on a $qp = 0$. On en déduit que

$$\begin{aligned} r^2 &= (p + q - pq)(p + q - pq) \\ &= p^2 + pq - p^2q + qp + q^2 - qpq - pqp - pq^2 + pqpq \\ &= p + pq - pq + q - pq = r. \end{aligned}$$

et r est bien un projecteur de E.

• On a $pr = p^2 = p$ de sorte que $\operatorname{Ker} r \subset \operatorname{Ker} p$. De même $qr = q^2 = q$ et $\operatorname{Ker} r \subset \operatorname{Ker} q$. Ainsi, $\operatorname{Ker} r \subset \operatorname{Ker} p \cap \operatorname{Ker} q$. L'autre inclusion étant évidente, on en déduit que

$$\boxed{\operatorname{Ker} r = \operatorname{Ker} p \cap \operatorname{Ker} q}.$$

• L'image de r est clairement incluse dans $\operatorname{Im} p + \operatorname{Im} q$. Réciproquement, soit $x \in \operatorname{Im} p + \operatorname{Im} q$ que l'on écrit $x = x_1 + x_2$ avec $x_1 \in \operatorname{Im} p$ et $x_2 \in \operatorname{Im} q$. On a $q(x) = q(x_1) + q(x_2) = q(x_2) = x_2$ car $\operatorname{Im} p \subset \operatorname{Ker} q$. Ainsi, on obtient

$$r(x) = p(x) + q(x) - pq(x) = x_1 + p(x_2) + x_2 - p(x_2) = x_1 + x_2 = x.$$

Il en résulte que $x \in \operatorname{Im} r$, et on a finalement, la somme étant directe car $\operatorname{Im} p \subset \operatorname{Ker} q$,

$$\boxed{\operatorname{Im} r = \operatorname{Im} p \oplus \operatorname{Im} q}. \quad \triangleleft$$

Soit p un projecteur d'un K*-espace vectoriel de dimension finie. Si* K *est de caractéristique nulle la trace de p est égale*[1] *au rang de p. En fait, on a de manière générale,* $\operatorname{Tr} p = (\operatorname{rg} p)1_{\mathrm{K}}$ *où* 1_{K} *est l'élément unité de* K. *Pour s'en convaincre, il suffit d'écrire la matrice de p dans une base obtenue par réunion d'une base de* $\operatorname{Im} p$ *et une base de* $\operatorname{Ker} p$.

6.19. Une somme de projecteurs

Soit E un K-espace vectoriel de dimension finie n, K sous-corps de $\mathbb{C}$. On se donne n endomorphismes non nuls de E, $p_1, p_2, \ldots, p_n$ tels que, pour $1 \leqslant i, j \leqslant n$, $p_i \circ p_j = \delta_{i,j} p_i$.

1. Montrer que pour $1 \leqslant i \leqslant n$, $\operatorname{rg} p_i = 1$.

2. Montrer que les sous espaces $\operatorname{Im} p_i$, pour $1 \leqslant i \leqslant n$, sont en somme directe.

(École polytechnique)

▷ **Solution.**

1. Pour tout i, on a $p_i^2 = p_i$, donc p_i est un projecteur. Considérons la somme $p = p_1 + \cdots + p_n$. D'après l'hypothèse, on a

$$p^2 = \sum_{1 \leqslant i,j \leqslant n} p_i \circ p_j = \sum_{i=1}^{n} p_i^2 = \sum_{i=1}^{n} p_i = p$$

1. La trace de p est un élément de K et le rang de p est dans $\mathbb{N}$. Mais si K est de caractéristique nulle, on peut identifier $\mathbb{Q}$ à un sous-corps de K, ce qui est implicitement fait ici.

et p est aussi un projecteur. Comme le rang d'un projecteur n'est autre que sa trace, on a $\operatorname{rg} p = \operatorname{Tr} p = \sum_{i=1}^{n} \operatorname{Tr} p_i = \sum_{i=1}^{n} \operatorname{rg} p_i \geqslant n$, puisque les p_i sont non nuls. Donc nécessairement, $\operatorname{rg} p = \sum_{i=1}^{n} \operatorname{rg} p_i = n$, ce qui signifie que $p = \mathrm{Id}_E$ et pour chaque i, $\operatorname{rg} p_i = 1$.

2. Montrons que les sous-espaces $\operatorname{Im} p_i$ sont en somme directe. On a

$$E = \mathrm{Id}_E(E) = (p_1 + \cdots + p_n)(E) \subset p_1(E) + \cdots + p_n(E)$$

et donc $E = \sum_{i=1}^{n} \operatorname{Im} p_i$. Chaque $\operatorname{Im} p_i$ est une droite et comme on a de plus $\dim E = \sum_{i=1}^{n} \dim \operatorname{Im} p_i$, les images sont bien en somme directe ; c'est une conséquence du lemme suivant.

Lemme. *Soit* E *un sous-espace de dimension finie,* $(F_1, \ldots, F_p)$ *une famille de sous-espaces de* E. *Alors on a l'équivalence*

$$F_1 + \cdots + F_p \text{ est directe} \iff \dim(F_1 + \cdots + F_p) = \sum_{k=1}^{p} \dim F_k .$$

Démonstration. On introduit l'application linéaire

$$f : \begin{array}{ccc} F_1 \times F_2 \times \cdots \times F_p & \longrightarrow & F_1 + F_2 + \cdots + F_p \\ (x_1, x_2, \ldots, x_p) & \longmapsto & x_1 + x_2 + \cdots + x_p \end{array}$$

Elle est surjective. Dire que la somme $F_1 + \cdots + F_p$ est directe, c'est dire que f est injective ou encore que $\dim \operatorname{Ker} f = 0$, ou encore, d'après le théorème du rang que

$$\dim(F_1 + \cdots + F_p) = \dim(F_1 \times \cdots \times F_p) = \dim F_1 + \cdots + \dim F_p . \diamondsuit$$

Si dans l'exercice suivant, il est question de projecteurs de $\mathbb{C}[X]$, *on s'intéresse surtout à la recherche de sous-espaces stables par certains endomorphismes.*

6.20. Endomorphismes de $\mathbb{C}[X]$

Pour $Q \in \mathbb{C}[X]$, non nul, soit $\pi_Q : \mathbb{C}[X] \longmapsto \mathbb{C}[X]$ l'application qui à P fait correspondre le reste de la division euclidienne de P par Q.

1. Montrer que, pour tout polynôme Q non nul, π_Q est un projecteur. Déterminer son image et son noyau.

2. Montrer que si Q_1 et Q_2 sont deux polynômes non nuls, on a, pour tout $P \in \mathbb{C}[X]$, $\pi_{Q_1Q_2}(Q_1P) = Q_1\pi_{Q_2}(P)$.

On fixe $Q \in \mathbb{C}[X]$, non nul, et on considère $S_Q : \mathbb{C}[X] \longmapsto \mathbb{C}[X]$ application définie par $S_Q(P) = \pi_Q(XP)$.

On dit qu'un sous-espace M de $\mathbb{C}[X]$ est stable si $S_Q(M) \subset M$, qu'il est invariant si $S_Q(M) = M$.

3. Montrer que si $Q = Q_1Q_2$, alors $Q_1 \operatorname{Im} \pi_{Q_2}$ est stable. À quelle condition est-il invariant ?

4. Soit N un sous-espace stable et $M = N + Q\,\mathbb{C}[X]$. Montrer que M est de la forme $Q_1\,\mathbb{C}[X]$.

5. Soit N un sous-espace invariant. Montrer qu'il existe Q_1 et Q_2 tels que $Q = Q_1Q_2$ et $N = Q_1 \operatorname{Im} \pi_{Q_2}$.

(École polytechnique)

▷ **Solution.**

1. Montrons que π_Q est linéaire. Notons n le degré de Q. Considérons P et P′ dans $\mathbb{C}[X]$, λ et λ' dans $\mathbb{C}$. Par définition, il existe P_1 et P'_1 dans $\mathbb{C}[X]$ tels que

$$\begin{cases} P = QP_1 + \pi_Q(P) \\ P' = QP'_1 + \pi_Q(P'). \end{cases}$$

On en déduit que $\lambda P + \lambda' P' = Q(\lambda P_1 + \lambda' P'_1) + (\lambda\pi_Q(P) + \lambda'\pi_Q(P'))$. Les polynômes $\pi_Q(P)$ et $\pi_Q(P')$ appartiennent à $\mathbb{C}_{n-1}[X]$. Il en est de même de $\lambda\pi_Q(P) + \lambda'\pi_Q(P')$. De l'unicité de la division euclidienne de $\lambda P + \lambda' P'$ par Q, il résulte que

$$\pi_Q(\lambda P + \lambda' P') = \lambda\pi_Q(P) + \lambda'\pi_Q(P'),$$

ce qui montre la linéarité de π_Q.

Pour tout $P \in \mathbb{C}[X]$, on a $\pi_Q(P) \in \mathbb{C}_{n-1}[X]$. La division euclidienne de $\pi_Q(P)$ par Q s'écrit donc $\pi_Q(P) = Q \times 0 + \pi_Q(P)$. On en déduit que $\pi_Q(\pi_Q(P)) = \pi_Q(P)$. On conclut que $\pi_Q \circ \pi_Q = \pi_Q$: π_Q est un projecteur.

On a clairement $\pi_Q(P) = 0$ si et seulement si Q divise P. Autrement dit, on a $\operatorname{Ker} \pi_Q = Q\,\mathbb{C}[X]$. Il est évident que $\operatorname{Im} \pi_Q = \mathbb{C}_{n-1}[X]$.

2. On effectue la division euclidienne de P par Q_2. Il existe $P_1 \in \mathbb{C}[X]$ tel que $P = Q_2P_1 + \pi_{Q_2}(P)$. On en déduit $Q_1P = Q_1Q_2P_1 + Q_1\pi_{Q_2}(P)$. Le polynôme $\pi_{Q_2}(P)$ est de degré strictement inférieur à celui de Q_2. On en déduit que le degré de $Q_1\pi_{Q_2}(P)$ est degré strictement inférieur au degré de Q_1Q_2. On a donc affaire à la division euclidienne de Q_1P par Q_1Q_2. On en déduit l'égalité $\pi_{Q_1Q_2}(Q_1P) = Q_1\pi_{Q_2}(P)$.

3. Soit $Q_1\pi_{Q_2}(P)$ un élément quelconque de $Q_1 \operatorname{Im}\pi_{Q_2}$. On a, d'après la question 2,

$$S_Q(Q_1\pi_{Q_2}(P)) = \pi_Q(XQ_1\pi_{Q_2}(P)) = Q_1\pi_{Q_2}(X\pi_{Q2}(P)).$$

On obtient $S_Q(Q_1 \operatorname{Im}\pi_{Q_2}) = Q_1\pi_{Q_2}(X \operatorname{Im}\pi_{Q_2}) \subset Q_1 \operatorname{Im}\pi_{Q_2} : Q_1 \operatorname{Im}\pi_{Q_2}$ est stable. De plus, il est invariant si $\pi_{Q_2}(X \operatorname{Im}\pi_{Q_2}) = \operatorname{Im}\pi_{Q_2}$.

Si le polynôme Q_2 est premier avec X, on obtient, pour tout $P \in \mathbb{C}[X]$,

$$\pi_{Q_2}(X\pi_{Q2}(P)) = 0 \Longrightarrow Q_2 \text{ divise } X\pi_{Q2}(P) \Longrightarrow Q_2 \text{ divise } \pi_{Q2}(P)$$
$$\Longrightarrow \pi_{Q2}(P) = 0.$$

La restriction de π_{Q_2} à $X \operatorname{Im}\pi_{Q_2}$ est injective. On a donc

$$\dim \pi_{Q_2}(X \operatorname{Im}\pi_{Q_2}) = \dim(X \operatorname{Im}\pi_{Q_2}) = \dim(\operatorname{Im}\pi_{Q_2}),$$

ce qui avec l'inclusion déjà démontrée, permet de conclure que

$$\pi_{Q_2}(X \operatorname{Im}\pi_{Q_2}) = \operatorname{Im}\pi_{Q_2} : Q_1 \operatorname{Im}\pi_{Q_2} \text{ est invariant.}$$

Par contre si X divise Q_2, on écrit $Q_2 = XQ_2'$ et on obtient, d'après la question 2,

$$\pi_{Q_2}(X\pi_{Q_2}(P)) = \pi_{XQ_2'}(X\pi_{Q_2}(P)) = X\pi_{Q_2'}(\pi_{Q_2}(P)).$$

Ce polynôme est toujours divisible par X. On n'obtient pas ainsi $\operatorname{Im}\pi_{Q_2}$ en entier. Le sous-espace vectoriel $Q_1 \operatorname{Im}\pi_{Q_2}$ n'est pas invariant.

4. L'anneau $\mathbb{C}[X]$ étant principal, pour démontrer que M s'écrit $Q_1\mathbb{C}[X]$, il suffit de démontrer que c'est un idéal de $\mathbb{C}[X]$. C'est déjà clairement un sous-groupe additif, puisqu'un sous-espace vectoriel. En effet, c'est la somme de deux sous-espaces vectoriels de $\mathbb{C}[X]$.

Il reste à montrer que si $P \in M$ et $R \in \mathbb{C}[X]$, alors $PR \in M$. Le sous-espace M étant somme de N et de $Q\mathbb{C}[X]$, il suffit de le démontrer pour $P \in N$ et pour $P \in Q\mathbb{C}[X]$. Or, il est clair que si $P \in Q\mathbb{C}[X]$, alors $PR \in Q\mathbb{C}[X]$. On prend donc $P \in N$ et on démontre que $PR \in M$. Par linéarité, on peut se limiter à $R = X^k$, pour $k \geqslant 1$, et même à $R = X$.

Supposons en effet démontré que, pour tout $P \in N$, on a $XP \in M$. Un récurrence simple sur k conduit à $X^kP \in M$ pour tout $P \in N$. C'est démontré pour $k = 1$, et si on suppose que c'est vrai pour $k \geqslant 1$, alors, pour $P \in N$, il existe $P_1 \in N$ et $P_2 \in \mathbb{C}[X]$ tels que $X^kP = P_1 + QP_2$. On en déduit que $X^{k+1}P = XP_1 + QXP_2$. Le cas $k = 1$ donne $XP_1 \in M$ et d'autre part, on a $QXP_2 \in Q\mathbb{C}[X] \subset M$. On conclut que $X^{k+1}P$ est dans M.

Soit donc $P \in N$. Il existe $P' \in \mathbb{C}[X]$ tel que

$$PX = QP' + \pi_Q(PX) = QP' + S_Q(P).$$

Le polynôme QP' est dans $Q\,\mathbb{C}[X]$ donc dans M et par hypothèse, N est stable, donc $S_Q(P)$ est dans N. On en déduit que XP appartient à M.

5. Le sous-espace N est invariant, donc stable. On définit M et Q_1 comme dans la question 4. On remarque que Q est dans M. On en déduit que Q_1 divise Q : il existe $Q_2 \in \mathbb{C}[X]$ tel que $Q = Q_1Q_2$.

Soit $P \in N$. Puisque N est invariant par S_Q, il existe $P' \in N$ tel que $P = S_Q(P')$. Le polynôme P' appartient à N donc à $Q_1\,\mathbb{C}[X]$. Il s'écrit donc $P' = Q_1P''$. On a donc

$$P = S_Q(Q_1P'') = \pi_Q(Q_1XP'') = \pi_{Q_1Q_2}(Q_1XP'') = Q_1\pi_{Q_2}(XP''),$$

d'après la question 2. Ceci démontre que $N \subset Q_1 \operatorname{Im}\pi_{Q_2}$. Soit N' le sous-espace vectoriel de $\operatorname{Im}\pi_{Q_2}$ tel que $N = Q_1N'$.

De l'égalité $Q_1\,\mathbb{C}[X] = N + Q\,\mathbb{C}[X]$, on déduit que $\mathbb{C}[X] = N' + Q_2\,\mathbb{C}[X]$. Le sous-espace N' étant inclus dans $\operatorname{Im}\pi_{Q_2} = C_{d-1}[X]$, où d est le degré de Q_2, on a $N' \cap Q_2\,\mathbb{C}[X] = \{0\}$. On a donc $\mathbb{C}[X] = N' \oplus Q_2\,\mathbb{C}[X]$. Mais on a aussi $\mathbb{C}[X] = \operatorname{Im}Q_2 \oplus Q_2\,\mathbb{C}[X]$ (π_{Q_2} est la projection sur $\operatorname{Im}Q_2$, parallèlement à $Q_2\,\mathbb{C}[X]$), et puisque $N' \subset \operatorname{Im}\pi_{Q_2}$, on conclut que $N' = \operatorname{Im}\pi_{Q_2}$, c'est-à-dire que $N = Q_1 \operatorname{Im}\pi_{Q_2}$.

On a démontré que tout sous-espace N invariant s'écrit $Q_1 \operatorname{Im}\pi_{Q_2}$, avec $Q_1Q_2 = Q$. La question 3 permet de préciser que Q_2 est premier avec X. ◁

L'égalité entre le rang et la trace d'un projecteur rappelée plus haut est utilisée dans l'exercice suivant, pour déterminer la dimension de l'espace des points invariants sous l'action d'un sous-groupe fini de GL(E).

6.21. Formule de Burnside

1. Soit E un $\mathbb{C}$-espace vectoriel de dimension finie $n \geqslant 1$, G un sous-groupe fini de GL(E). On pose

$$E^G = \{x \in E,\ \forall g \in G,\ g(x) = x\}.$$

Montrer que

$$\dim E^G = \frac{1}{|G|} \sum_{g \in G} \operatorname{Tr}(g).$$

2. Soit G un sous-groupe de $\mathcal{S}_n$. Pour $g \in G$, on note $F(g)$ le nombre de points fixes de g. Soit r le nombre d'orbites pour l'opération de G sur $[\![1, n]\!]$. Déduire de ce qui précède que

$$r = \frac{1}{|G|} \sum_{g \in G} F(g).$$

3. Donner une preuve directe de la formule de la question 2.

(École normale supérieure)

▷ **Solution.**

1. Posons $u = \dfrac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} g$. On voit que, si $g \in \mathrm{G}$, on a $gu = ug = u$. En effet, on peut écrire par exemple,

$$gu = g \frac{1}{|\mathrm{G}|} \sum_{g' \in \mathrm{G}} g' = \frac{1}{|\mathrm{G}|} \sum_{g' \in \mathrm{G}} gg' = u$$

car, g étant fixé, l'application $g' \longmapsto gg'$ est une bijection de G sur G. On en déduit que

$$u^2 = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} gu = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} u = u,$$

c'est-à-dire que u est un projecteur.

Si $x \in \mathrm{E}^{\mathrm{G}}$, il est clair que $u(x) = x$. Mais inversement, si $x \in \operatorname{Im} u$, on a $u(x) = x$ et la relation $gu = u$ implique que pour tout $g \in \mathrm{G}$,

$$g(x) = g(u(x)) = u(x) = x,$$

ce qui montre que $x \in \mathrm{E}^{\mathrm{G}}$. Finalement E^{G} est l'image du projecteur u. En caractéristique nulle, ce qui est le cas ici, la trace d'un projecteur est égale à son rang. D'où le résultat.

2. On réalise G comme sous-groupe de $\mathrm{GL}_n(\mathbb{C})$ à l'aide des matrices de permutation : à tout $g \in \mathrm{G}$, on associe la matrice P_g de terme général $\delta_{i,g(j)}$, où δ désigne le symbole de Kronecker. On vérifie que, pour $(g, g') \in \mathrm{G}^2$, on a $\mathrm{P}_g \circ \mathrm{P}_{g'} = \mathrm{P}_{gg'}$. L'application $g \longmapsto \mathrm{P}_g$ est donc un morphisme de groupes, injectif, de G dans $\mathrm{GL}_n(\mathbb{C})$. On en déduit que $\mathrm{G}' = \{\mathrm{P}_g,\ g \in \mathrm{G}\}$ est un sous-groupe de $\mathrm{GL}_n(\mathbb{C})$ isomorphe à G. Le lien avec la question précédente apparaît déjà puisque $\mathrm{F}(g)$, le nombre de points fixes de la permutation g, n'est autre que la trace de la matrice P_g. Il ne reste plus qu'à voir pourquoi la dimension du sous-espace $\mathrm{E}^{\mathrm{G}'}$ est égale au nombre r d'orbites.

Notons $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{C}^n$, $\Omega_1, \ldots, \Omega_r$ les différentes orbites de $[\![1, n]\!]$ sous l'opération de G et posons $\mathrm{F}_k = \operatorname{Vect}(e_i)_{i \in \Omega_k}$ pour tout $k \in [\![1, r]\!]$. On remarque que, pour tout $g \in \mathrm{G}$, on a $\mathrm{P}_g(e_j) = e_{g(j)}$. Les sous-espaces vectoriels F_k sont donc stables par les matrices de G'. De plus, on a $\mathbb{C}^n = \mathrm{F}_1 \oplus \cdots \oplus \mathrm{F}_r$. Soit $\mathrm{X} \in \mathbb{C}^n$ que l'on décompose en $\mathrm{X} = \mathrm{X}_1 + \cdots + \mathrm{X}_r$ avec $\mathrm{X}_i \in \mathrm{F}_i$, pour $1 \leqslant i \leqslant r$. On obtient, pour $g \in \mathrm{G}$, $\mathrm{P}_g\mathrm{X} = \sum_{i=1}^{r} \mathrm{P}_g\mathrm{X}_i$. Comme $\mathrm{P}_g\mathrm{X}_i \in \mathrm{F}_i$, on a, par unicité de la décomposition, $\mathrm{P}_g(\mathrm{X}) = \mathrm{X}$ si et seulement si $\mathrm{P}_g\mathrm{X}_i = \mathrm{X}_i$ pour tout $i \in [\![1, r]\!]$. On en déduit que

$$\mathrm{E}^{\mathrm{G}'} = \bigoplus_{i=1}^{r} (\mathrm{E}^{\mathrm{G}'} \cap \mathrm{F}_i).$$

Or, il est aisé de voir que $\mathrm{E}^{\mathrm{G}'} \cap \mathrm{F}_i$ est la droite vectorielle engendrée par le vecteur $\sum_{k\in\Omega_i} e_k$. En effet si $x \in \mathrm{F}_i$ s'écrit $x = \sum_{k\in\Omega_i} \lambda_k e_k$, où les λ_k sont dans $\mathbb{C}$, on obtient, pour tout $g \in \mathrm{G}$, $\mathrm{P}_g(x) = \sum_{k\in\Omega_i} \lambda_k e_{g(k)} = \sum_{k\in\Omega_i} \lambda_{g^{-1}(k)} e_k$.

On en déduit que $x \in \mathrm{E}^{\mathrm{G}'} \cap \mathrm{F}_i$ si et seulement si, pour tout $k \in \Omega_i$ et tout $g \in \mathrm{G}$, on a $\lambda_{g^{-1}(k)} = \lambda_k$. Le groupe G agissant transitivement sur Ω_i, par définition, il faut que tous les λ_k soient égaux, ce qui conduit au résultat annoncé. On a donc, pour tout $k \in \Omega_i$, $\dim(\mathrm{E}^{\mathrm{G}'} \cap \mathrm{F}_i) = 1$ et par conséquent $\dim \mathrm{E}^{\mathrm{G}'} = r$.

3. L'idée est simplement de calculer de deux manières différentes le cardinal de l'ensemble $\mathrm{A} = \{(g,x) \in \mathrm{G} \times [\![1,n]\!],\ g(x) = x\}$. Si on somme d'abord selon G, on obtient $|\mathrm{A}| = \sum_{g\in\mathrm{G}} \mathrm{F}(g)$. Si on somme selon les éléments de $[\![1,n]\!]$, on a $|\mathrm{A}| = \sum_{k=1}^{n} |\mathrm{G}_k|$ où $\mathrm{G}_k = \{g \in \mathrm{G},\ g(k) = k\}$ est le stabilisateur de k. Or, si on note $\Omega_1, \ldots, \Omega_r$ les orbites de $[\![1,n]\!]$ pour l'action de G, on a

$$\sum_{1\leqslant k\leqslant n} |\mathrm{G}_k| = \sum_{i=1}^{r} \sum_{k\in\Omega_i} |\mathrm{G}_k| = \sum_{i=1}^{r} \sum_{k\in\Omega_i} \frac{|\mathrm{G}|}{|\Omega_i|} = r|\mathrm{G}|.$$

On retrouve la formule en identifiant les résultats des deux calculs. ◁

Le résultat de cette dernière question se généralise à un groupe fini quelconque G *opérant sur un ensemble fini* X*, la preuve étant la même. Cette formule de Burnside*[2] *est un résultat important en combinatoire (méthode de Pólya) qui intervient lorsqu'on cherche à dénombrer les configurations d'un ensemble modulo l'action d'un groupe : par exemple, de combien de manières différentes peut-on peindre un cube avec d couleurs, à rotation près ?*

Dans l'étude d'un endomorphisme u on essaye, autant que faire se peut, de découper l'espace en une somme directe de sous-espaces stables pour se ramener à l'étude (que l'on espère plus simple) des restrictions de u à ces sous-espaces. C'est typiquement l'objet de la réduction avec les sous-espaces propres ou les sous-espaces caractéristiques. Dans l'étude des représentations linéaires d'un groupe fini G*, on rencontre la même démarche, qui conduit à la notion de représentations irréductibles*[3]*. C'est le lemme de l'exercice suivant qui permet cela : on se donne un*

2. Bien qu'usuellement connue sous ce nom, elle n'est pas due à Burnside mais à Frobenius. Il y a comme cela un certain nombre de théorèmes injustement baptisés...

3. Pour une introduction à la théorie des représentations des groupes le lecteur pourra consulter SERRE (J.-P.), *Représentations linéaires des groupes finis*, Hermann, 1978 ou résoudre le problème posé aux ENS Lyon-Cachan en 1997 qui concerne le début de la théorie : représentations irréductibles, orthogonalité des caractères...

groupe fini de GL(E) *qui stabilise un sous-espace* F *de* E *et on cherche à montrer l'existence d'un supplémentaire stable. Ce sera l'occasion de voir, comme annoncé plus haut, le rapport entre les projecteurs et les décompositions en somme directe.*

6.22. Théorème de Maschke

Soit E un $\mathbb{C}$-espace vectoriel de dimension finie $n \geqslant 1$, G un sous-groupe fini de GL(E) et F un sous-espace vectoriel de E stable par tous les éléments de G. Montrer que F admet un supplémentaire stable par tous les éléments de G.

(École normale supérieure)

▷ **Solution.**

• Remarquons qu'il y a une correspondance bijective entre les supplémentaires de F et les projecteurs dont l'image est F, un supplémentaire de F étant le noyau d'un tel projecteur. La question posée équivaut donc à trouver un projecteur p d'image F tel que $\operatorname{Im} p$ (par hypothèse) et $\operatorname{Ker} p$ soient stables par tous les éléments de G.

Soit p un projecteur de E et $u \in \mathcal{L}(\mathrm{E})$. Si $\operatorname{Ker} p$ et $\operatorname{Im} p$ sont stables par u, alors u commute avec p sur $\operatorname{Ker} p$ et sur $\operatorname{Im} p$, donc par linéarité sur $\mathrm{E} = \operatorname{Im} p \oplus \operatorname{Ker} p$. Réciproquement, si u et p commutent, alors $\operatorname{Ker} p$ et $\operatorname{Im} p = \operatorname{Ker}(p - \mathrm{Id})$ sont des sous-espaces stables par u (en effet, ce sont des sous-espaces propres de p).

Il nous faut donc trouver un projecteur p qui commute avec tous les éléments de G ; son noyau fournira alors un supplémentaire de F stable par tous les éléments de G.

Partons d'un projecteur quelconque q dont l'image est F. L'idée fondamentale est de moyenner les conjugués de q par les éléments de G. Posons

$$p = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} g \circ q \circ g^{-1}.$$

• Montrons que p commute avec tout élément de G. Si $g_0 \in \mathrm{G}$, on a

$$g_0 \circ p = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} g_0 \circ g \circ q \circ g^{-1} = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} (g_0 \circ g) \circ q \circ g^{-1} \circ g_0^{-1} \circ g_0$$

$$= \left(\frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} (g_0 \circ g) \circ q \circ (g_0 \circ g)^{-1} \right) g_0 = p \circ g_0,$$

car lorsque g décrit le groupe G, $g_0 \circ g$ également puisque la translation à gauche $g \longmapsto g_0 \circ g$ est une bijection de G.

• Pour finir, il n'y a plus qu'à vérifier que p est un projecteur d'image F.

Si $x \in F$, puisque F est stable par tous les éléments de G, on obtient, pour tout $g \in G$,

$$g \circ q \circ g^{-1}(x) = g(g^{-1}(x)) = x.$$

Il en résulte que $p(x) = x$. Tout vecteur de F est fixe par p.

Si maintenant x est un vecteur quelconque dans E, on a $p(x) \in F$ car $g \circ q \circ g^{-1}(x) \in F$ pour tout $g \in G$. On en déduit que $p(p(x)) = p(x)$. On conclut que p est bien un projecteur d'image F. $\lhd$

La preuve ci-dessus reste valide en caractéristique p (premier) à condition que p ne divise pas le cardinal de G. Sur le corps des réels (resp. des complexes) on peut également utiliser un argument de moyenne avec le produit scalaire : on munit E d'une structure euclidienne (resp. hermitienne) quelconque et on pose, pour $(x, y) \in E^2$.

$$\langle x, y \rangle_G = \frac{1}{|G|} \sum_{g \in G} \langle g(x), g(y) \rangle.$$

On vérifie que $\langle\ ,\ \rangle_G$ est encore un produit scalaire sur E et que tous les éléments de G sont orthogonaux (resp. unitaires) pour ce produit scalaire. L'orthogonal de F au sens de $\langle\ ,\ \rangle_G$ fournit alors un supplémentaire stable. Cette démarche a aussi été proposée en exercice d'oral à l'École polytechnique.

Nous allons maintenant commencer une série d'exercices plus abstraits, consacrés à l'étude de l'algèbre $\mathcal{L}(E)$ où E est un K-espace vectoriel de dimension finie. Nous rappelons au lecteur que le centre de $\mathcal{L}(E)$ est réduit aux homothéties (on trouvera une démonstration de ce fait p. 269). L'exercice suivant est consacré aux automorphismes de l'algèbre $\mathcal{L}(E)$.

6.23. Automorphismes de la K-algèbre $\mathcal{L}(E)$

Soit E un K-espace vectoriel de dimension finie $n \geqslant 1$. Montrer que tout automorphisme de l'algèbre $\mathcal{L}(E)$ est de la forme

$$u \longmapsto \tau \circ u \circ \tau^{-1} \quad \text{où } \tau \in GL(E).$$

(École polytechnique)

▷ **Solution.**

Les conjugaisons $\varphi_\tau : u \longmapsto \tau \circ u \circ \tau^{-1}$, où $\tau \in \mathrm{GL(E)}$, sont clairement des automorphismes d'algèbre.

Réciproquement, soit φ un automorphisme d'algèbre de $\mathcal{L}(\mathrm{E})$. Choisissons une base $(e_1, \dots, e_n)$ de E et notons pour $1 \leqslant i, j \leqslant n$, u_{ij} l'endomorphisme de $\mathcal{L}(\mathrm{E})$ défini par $u_{ij}(e_k) = \delta_{jk} e_i$, pour tout $k \in [\![1, n]\!]$. La matrice de u_{ij} dans la base canonique de K^n est la matrice E_{ij} de la base canonique de $\mathcal{M}_n(\mathrm{K})$. Les u_{ij} forment donc une base de $\mathcal{L}(\mathrm{E})$. Posons alors $v_{ij} = \varphi(u_{ij})$.

On rappelle que pour tout (i, j, k, l), $u_{ij} \circ u_{kl} = \delta_{jk} u_{il}$. Comme φ est un automorphisme d'algèbre, les v_{ij} vérifient les mêmes relations. En particulier, on a, pour tout i, $v_{ii}^2 = v_{ii}$, de sorte que v_{ii} est un projecteur.

Si φ est de la forme φ_τ, on vérifie que les v_{ij} sont définis à partir de la base $(\tau(e_1), \dots, \tau(e_n))$, exactement comme les u_{ij} le sont à partir de $(e_1, \dots, e_n)$. Cette remarque nous donne la méthode pour construire à partir de φ la base $(\tau(e_1), \dots, \tau(e_n))$ et l'application τ.

Choisissons un vecteur non nul e'_1 dans l'image de v_{11} et posons pour tout $i \in [\![1, n]\!]$, $e'_i = v_{i1}(e'_1)$. On observe alors que :

- e'_i est dans l'image de v_{ii}, car $v_{ii}(e'_i) = v_{ii} \circ v_{i1}(e'_1) = v_{i1}(e'_1) = e'_i$;
- e'_i est non nul, car $v_{1i}(e'_i) = v_{1i} \circ v_{i1}(e'_1) = v_{11}(e'_1) = e'_1 \neq 0$;
- $(e'_1, \dots, e'_n)$ est une base de E. En effet, on a $v_{jj}(e'_i) = v_{jj} \circ v_{i1}(e'_1) = \delta_{ij} v_{j1}(e'_1) = \delta_{ij} e'_i$. Ainsi, si $\sum_{i=1}^{n} \lambda_i e'_i = 0$, en appliquant v_{jj} à cette relation, il vient $\lambda_j = 0$.

Notons alors τ l'unique isomorphisme de E qui envoie e_i sur e'_i pour tout i. On va montrer que $v_{ij} = \tau \circ u_{ij} \circ \tau^{-1}$ pour tout couple (i, j). En effet, on a, pour tout $k \in [\![1, n]\!]$,

$$(\tau \circ u_{ij} \circ \tau^{-1})(e'_k) = (\tau \circ u_{ij})(e_k) = \tau(\delta_{jk} e_i) = \delta_{jk} e'_i$$

et par ailleurs,

$$v_{ij}(e'_k) = v_{ij} \circ v_{k1}(e'_1) = \delta_{jk} v_{i1}(e'_1) = \delta_{jk} e'_i.$$

Il en résulte que les automorphismes φ et φ_τ, qui coïncident sur la base (u_{ij}) de $\mathcal{L}(\mathrm{E})$ sont égaux. ◁

Les idéaux de $\mathcal{L}(\mathrm{E})$ ont fait l'objet de plusieurs exercices posés aux concours. Comme $\mathcal{L}(\mathrm{E})$ est une algèbre non commutative (dès que $\dim \mathrm{E} > 1$), on est amené à distinguer les idéaux à droite et les idéaux à gauche. Rappelons les définitions : un idéal à gauche (resp. à droite) est un sous-groupe additif tel que pour tout $a \in \mathcal{L}(\mathrm{E})$ et $u \in \mathrm{I}$, $au \in \mathrm{I}$ (resp. $ua \in \mathrm{I}$). On parle d'idéal bilatère pour un idéal à droite et à gauche. Le premier exercice fait montrer que $\mathcal{L}(\mathrm{E})$ est une algèbre simple, c'est-à-dire sans idéal bilatère non trivial.

6.24. Simplicité de $\mathcal{L}(E)$

1. Soit E un espace vectoriel de dimension finie. Montrer que les seuls idéaux bilatères de $\mathcal{L}(E)$ sont $\{0\}$ et E. Cela reste-t-il vrai en dimension infinie ?

2. Soit p une semi-norme sur $M_n(\mathbb{C})$ vérifiant $p(AB) \leqslant p(A)p(B)$ pour tout $(A, B) \in M_n(\mathbb{C})^2$. Montrer que p est nulle ou que p est une norme.

(École normale supérieure)

▷ **Solution.**

1. On se ramène à la recherche des idéaux bilatères de $\mathcal{M}_n(K)$. Soit I un idéal bilatère non nul de $\mathcal{M}_n(K)$ et $M = (m_{ij})_{1\leqslant i,j\leqslant n} \in I$, non nulle. Notons $(E_{ij})_{1\leqslant i,j\leqslant n}$ la base canonique de $\mathcal{M}_n(K)$ et $(i_0, j_0) \in [\![1, n]\!]^2$ tel que $m_{i_0 j_0} \neq 0$. Nous savons que $E_{ij}E_{kl} = \delta_{jk}E_{il}$ pour tout $(i, j, k. l)$ dans $[\![1, n]\!]^4$. Ainsi, l'élément

$$E_{ii_0}ME_{j_0 j} = \sum_{1\leqslant k,l\leqslant n} m_{kl}E_{ii_0}E_{kl}E_{j_0 j} = \sum_{k=1}^{n} m_{kj_0}E_{ii_0}E_{kj} = m_{i_0 j_0}E_{ij}$$

est dans I, et comme $m_{i_0 j_0} \neq 0$, E_{ij} est également dans I. Ceci est vrai pour tout (i, j) de $[\![1, n]\!]^2$. Comme les E_{ij} engendrent $\mathcal{M}_n(K)$, on obtient $I = \mathcal{M}_n(K)$.

En revanche, si E est de dimension infinie, il existe des idéaux bilatères non triviaux, par exemple l'idéal des endomorphismes de rang fini.

2. Considérons l'ensemble N des matrices A telles que $p(A) = 0$. La définition d'une semi-norme implique que N est un sous-espace vectoriel de $M_n(\mathbb{C})$. L'inégalité vérifiée par p montre que c'est un idéal bilatère de $M_n(\mathbb{C})$. Comme les seuls idéaux bilatères sont $\{0\}$ et $M_n(\mathbb{C})$, on a le résultat. ◁

Les deux exercices qui suivent sont consacrés à la description des idéaux à gauche et à droite.

6.25. Idéaux à gauche de $\mathcal{L}(E)$

Soit E un K-espace vectoriel de dimension finie. On note, pour tout F sous-espace de E,

$$g(F) = \{u \in \mathcal{L}(E), F \subset \operatorname{Ker} u\},$$

et si H est une partie de $\mathcal{L}(E)$, on pose $f(H) = \bigcap_{u\in H} \operatorname{Ker} u$.

1. Vérifier que, pour tout sous-espace vectoriel F de E, $g(F)$ est un idéal à gauche de $\mathcal{L}(E)$ et que l'on a $(f \circ g)(F) = F$.

Soit I un idéal à gauche de $\mathcal{L}(E)$.

2. Soit $u \in I$, $v \in \mathcal{L}(E)$ tel que $\operatorname{Ker} u \subset \operatorname{Ker} v$. Montrer que $v \in I$.

3. Si p et q sont deux projecteurs appartenant à I, montrer qu'il existe un projecteur r dans I tel que $\operatorname{Ker} r = \operatorname{Ker} p \cap \operatorname{Ker} q$.

4. Prouver qu'il existe un projecteur p appartenant à I tel que $\operatorname{Ker} p = f(I)$, puis que $I = \mathcal{L}(E)p$.

5. Étudier $(g \circ f)(I)$ et conclure.

(École polytechnique)

▷ **Solution.**

1. Pour u et v dans $g(F)$ et $a \in \mathcal{L}(E)$, on a $\operatorname{Ker} u \subset \operatorname{Ker} au$ et $F \subset \operatorname{Ker} u \cap \operatorname{Ker} v \subset \operatorname{Ker}(u+v)$. Les endomorphismes au et $u+v$ sont donc dans $g(F)$. Il en résulte que $g(F)$ est un idéal à gauche de $\mathcal{L}(E)$.

• On a clairement $F \subset \bigcap_{u \in g(F)} \operatorname{Ker} u$. Si p est la projection sur un supplémentaire G de F, parallèlement à F, on a $\operatorname{Ker} p = F$ et $p \in g(F)$. On en déduit que $\bigcap_{u \in g(F)} \operatorname{Ker} u \subset F$ et finalement $\bigcap_{u \in g(F)} \operatorname{Ker} u = F$, c'est-à-dire

$$\boxed{(f \circ g)(F) = F}.$$

2. Cela résulte directement du lemme de factorisation (cf. la première question de l'exercice 6.4) : de l'inclusion $\operatorname{Ker} u \subset \operatorname{Ker} v$, on déduit qu'il existe $a \in \mathcal{L}(E)$ tel que $v = au$. D'où il résulte que v est dans I.

3. Soit p et q deux projecteurs appartenant à I. On cherche un projecteur r appartenant à I tel que $\operatorname{Ker} r = \operatorname{Ker} p \cap \operatorname{Ker} q$. Soit X (resp. Y) un supplémentaire de $\operatorname{Ker} p \cap \operatorname{Ker} q$ dans $\operatorname{Ker} p$ (resp $\operatorname{Ker} q$) et Z un supplémentaire dans E de $\operatorname{Ker} p + \operatorname{Ker} q$. On obtient donc $E = Z \oplus X \oplus Y \oplus (\operatorname{Ker} p \cap \operatorname{Ker} q)$. Le projecteur r_1 sur $Y \oplus Z$ parallèlement à $X \oplus (\operatorname{Ker} p \cap \operatorname{Ker} q) = \operatorname{Ker} p$ a pour noyau $\operatorname{Ker} p$. Il est donc dans I d'après la question précédente. De même, le projecteur r_2 sur X et de noyau $\operatorname{Ker} q \oplus Z$ est dans I, toujours d'après la question précédente. On pose $r = r_1 + r_2$. Alors r est le projecteur sur $Z \oplus X \oplus Y$ de noyau $\operatorname{Ker} p \cap \operatorname{Ker} q$ et il est dans I.

4. Soit p un projecteur de I de rang maximal ou, si l'on préfère, dont la dimension du noyau est minimale et u un élément quelconque de I. On peut trouver un projecteur q tel que $\operatorname{Ker} q = \operatorname{Ker} u$. Celui-ci est dans I, d'après la question 2. De la question précédente, on déduit qu'il existe $r \in I$ tel que $\operatorname{Ker} r = \operatorname{Ker} p \cap \operatorname{Ker} q$. Comme $\dim \operatorname{Ker} r \geqslant \dim \operatorname{Ker} p$ par choix de p, on en déduit que $\operatorname{Ker} p = \operatorname{Ker} r \subset \operatorname{Ker} q = \operatorname{Ker} u$. Par conséquent, on obtient

$$\operatorname{Ker} p \subset \bigcap_{u \in I} \operatorname{Ker} u \subset \operatorname{Ker} p \quad \text{et} \quad \boxed{\operatorname{Ker} p = f(I)}.$$

On a déjà l'inclusion $\mathcal{L}(E)p \subset I$ puisque $p \in I$. Réciproquement, si $u \in I$, comme $\operatorname{Ker} p \subset \operatorname{Ker} u$, la question 2 montre que $u \in \mathcal{L}(E)p$. On conclut que $I = \mathcal{L}(E)p$.

5. On a $I \subset g(f(I))$ car tous les éléments de l'idéal I s'annulent sur le sous-espace $f(I)$. Mais inversement, si u vérifie $f(I) = \operatorname{Ker} p \subset \operatorname{Ker} u$, on a $u \in I$, toujours d'après la question 2. Ainsi, $g(f(I)) = I$ et $g \circ f$ est l'identité.

Conclusion. L'application g établit une bijection, strictement décroissante pour l'inclusion, entre les idéaux à gauche et les sous-espaces vectoriels de E. Sa bijection réciproque est f. De plus, tout idéal à gauche I est engendré par un projecteur (quelconque) de noyau le sous-espace $f(I)$. ◁

On va obtenir une description géométrique similaire des idéaux à droite, les noyaux étant remplacés par les images.

6.26. Idéaux à droite de $\mathcal{L}(E)$

Soit E un K-espace vectoriel de dimension finie. Si I est un idéal à droite de $\mathcal{L}(E)$ et F un sous-espace de E, on pose

$$f(I) = \sum_{u \in I} \operatorname{Im} u \quad \text{et} \quad g(F) = \{u \in \mathcal{L}(E), \operatorname{Im} u \subset F\}$$

1. Montrer que pour tout sous-espace F, $g(F)$ est un idéal à droite de $\mathcal{L}(E)$ et que $f \circ g(F) = F$.

Soit I un idéal à droite.

2. Soit $u \in I$, $v \in \mathcal{L}(E)$ tel que $\operatorname{Im} v \subset \operatorname{Im} u$. Montrer que $v \in I$.

3. Soit u et v deux éléments de I. Montrer qu'il existe un projecteur $p \in I$ tel que $\operatorname{Im} p = \operatorname{Im} u + \operatorname{Im} v$.

4. En déduire qu'il existe un projecteur p appartenant à I tel que $\operatorname{Im} p = f(I)$, puis que $I = p\mathcal{L}(E)$.

5. Conclure.

(École polytechnique)

▷ **Solution.**

1. On a clairement $0 \in g(F)$. Si u et v sont dans $g(F)$ et $a \in \mathcal{L}(E)$, on obtient

$$\operatorname{Im}(u+v) \subset \operatorname{Im} u + \operatorname{Im} v \subset F + F = F \quad \text{et} \quad u + v \in g(F),$$
$$\operatorname{Im}(u \circ a) \subset \operatorname{Im} u \subset F \quad \text{et} \quad u \circ a \in g(F).$$

Ainsi, $g(F)$ est un idéal à droite de $\mathcal{L}(E)$.

L'inclusion $f(g(F)) \subset F$ est évidente. Si p est un projecteur sur F, parallèlement à un supplémentaire de F, on a $\operatorname{Im} p = F$ et $p \in g(F)$. On en déduit que $F \subset f(g(F))$ et finalement

$$\boxed{f(g(F)) = F}$$

2. Cela résulte directement du second lemme de factorisation (cf. la question 2 de l'exercice 6.4) : de l'inclusion $\operatorname{Im} v \subset \operatorname{Im} u$, on déduit l'existence de $a \in \mathcal{L}(E)$ tel que $v = ua$, de sorte que $v \in I$.

3. Soit S un supplémentaire de $\operatorname{Im} u \cap \operatorname{Im} v$ dans $\operatorname{Im} u$. On montre facilement l'égalité $S \oplus \operatorname{Im} v = \operatorname{Im} u + \operatorname{Im} v$. Soit S' un supplémentaire de cette somme dans E, de sorte que $E = S' \oplus S \oplus \operatorname{Im} v$. Comme $v \in I$, le projecteur q_1 sur $\operatorname{Im} v$ parallèlement à $S \oplus S'$ est dans I d'après la question précédente (car $\operatorname{Im} q_1 = \operatorname{Im} v$). Comme $u \in I$, le projecteur q_2 sur S parallèlement à $\operatorname{Im} v \oplus S'$ est aussi dans I, toujours par la question précédente (puisque $\operatorname{Im} q_2 \subset \operatorname{Im} u$). On en déduit que $p = q_1 + q_2$, qui est le projecteur sur $\operatorname{Im} u + \operatorname{Im} v$ parallèlement à S', est aussi dans I.

4. On choisit un élément v de I de rang maximal et on pose $F = \operatorname{Im} v$, puis on considère un projecteur p sur F. D'après la question 2, p est aussi dans I. La question précédente montre que tout élément de $u \in I$ a son image incluse dans F (sinon on pourrait trouver dans I un élément dont l'image $\operatorname{Im} u + F$ contiendrait F strictement ; son rang serait supérieur à celui de v). On a donc $f(I) = F = \operatorname{Im} p$.

On a clairement $p\mathcal{L}(E) \subset I$. Réciproquement, si $u \in I$, on sait que $\operatorname{Im} u \subset \operatorname{Im} p$ et d'après la question 2, il existe $a \in \mathcal{L}(E)$ tel que $u = pa$ et donc $u \in p\mathcal{L}(E)$. On conclut que $I = p\mathcal{L}(E)$.

5. On a clairement $I \subset g(f(I))$. Inversement soit $u \in g(f(I))$. On a alors $\operatorname{Im} u \subset f(I) = \operatorname{Im} p$ de sorte que $u \in I$, toujours d'après la question 2.

Conclusion. L'application f établit une bijection, strictement croissante pour l'inclusion, entre les idéaux à droite et les sous-espaces vectoriels de E. Sa bijection réciproque est g. De plus, tout idéal à droite I est engendré par un projecteur (quelconque) d'image le sous-espace $f(I)$. ◁

Le thème d'étude suivant de ce chapitre est la dualité. Rappelons que le dual d'un K*-espace vectoriel* E *est le* K*-espace vectoriel* $E^* = \mathcal{L}(E, K)$ *des formes linéaires de* E *dans* K*. On a* $\dim E^* = \dim E$ *en dimension finie, sans qu'il existe d'isomorphisme canonique entre* E *et son dual. Soit* $(e_1, \dots, e_n)$ *une base de* E*. La famille* $(e_1^*, \dots, e_n^*)$ *d'éléments de* E^* *définie par* $e_i^*(e_j) = \delta_{ij}$ *est une base de* E^* *appelée* base duale *de* $(e_1, \dots, e_n)$.

Les formes linéaires permettent de caractériser analytiquement les hyperplans de E *: tout hyperplan est le noyau d'une forme linéaire non nulle, forme qui est unique à un scalaire multiplicatif près.*

Plus généralement, si F *est une partie de* E *et* G *une partie de* E^**, on note*

$$F^\perp = \{u \in E^*, \forall x \in F, u(x) = 0\} \quad et \quad G^\circ = \{x \in E, \forall u \in G, u(x) = 0\}.$$

On appelle $F^\perp$ *l'orthogonal de* F *et* G° *l'orthogonal de* G*. On montre que* $F^\perp$ *et* G° *sont des sous-espaces vectoriels de* E^* *et* E *respectivement et que* $F^\perp = (\mathrm{Vect}\, F)^\perp$ *et* $G^\circ = (\mathrm{Vect}\, G)^\circ$*. On a, de plus,* $F \subset (F^\perp)^\circ$ *et* $G \subset (G^\circ)^\perp$*.*

En dimension finie, si F *et* G *sont des sous-espaces vectoriels de* E *et* E^* *respectivement, on a les propriétés suivantes :* $\dim F^\perp = \mathrm{codim}\, F$ *et* $\dim G^\circ = \mathrm{codim}\, G$*. d'où on déduit facilement* $(F^\perp)^\circ = F$ *et* $(G^\circ)^\perp = G$*.*

Le premier exercice ci-après, montre que l'égalité $(G^\circ)^\perp = G$ *subsiste en dimension infinie pour les sous-espaces* G *de dimension finie.*

6.27. Orthogonalité duale en dimension quelconque

Soit $f_1, \ldots, f_p, g$ des formes linéaires sur un espace vectoriel E de dimension quelconque. On suppose que $\bigcap_{i=1}^{p} \mathrm{Ker}\, f_i \subset \mathrm{Ker}\, g$. Montrer que g appartient à $\mathrm{Vect}(f_1, \ldots, f_p)$.

(**École normale supérieure**)

▷ **Solution.**

On peut supposer g non nulle, sans quoi le résultat est évident.

• En dimension finie, l'exercice est une une simple application des propriétés de l'orthogonalité duale. Si on note F le sous-espace de E^* engendré par $(f_1, \ldots, f_p)$ et D la droite de E^* engendrée par g, on a par hypothèse

$$\bigcap_{i=1}^{p} \mathrm{Ker}\, f_i = F^\circ \subset D^\circ = \mathrm{Ker}\, g.$$

Il en résulte, en prenant l'orthogonal que $(D^\circ)^\perp \subset (F^\circ)^\perp$. Sachant qu'en dimension finie, on a $(G^\circ)^\perp = G$ pour tout sous-espace vectoriel G de E^*, on obtient $D \subset F$, ce qui est le résultat demandé.

• En dimension infinie, l'égalité $(G^\circ)^\perp = G$ n'est malheureusement plus vraie en général. On a *a priori* seulement l'inclusion $G \subset (G^\circ)^\perp$. L'exercice demande de prouver qu'il y a égalité lorsque G est de dimension finie. On raisonne par récurrence sur p. On pose $H = \mathrm{Ker}\, g$: c'est un hyperplan de E puisqu'on a supposé $g \neq 0$.

⋆ Traitons le cas $p = 1$. On a $H_1 = \operatorname{Ker} f_1 \subset H$. La forme linéaire f_1 n'est donc pas nulle et H_1 est un hyperplan de E. Par suite, on a $H_1 = H$ et le cours permet d'affirmer que les formes linéaires g et f_1 sont proportionnelles.

⋆ Supposons le résultat vrai au rang p et considérons des formes linéaires $f_1, \dots, f_{p+1}$ telles que $\bigcap_{i=1}^{p+1} \operatorname{Ker} f_i \subset H$. On peut supposer que $f_{p+1} \neq 0$, sans quoi l'hypothèse de récurrence s'applique directement. Notons alors $\tilde{g}$, (resp. $\tilde{f}_i$ pour $1 \leqslant i \leqslant p$) la restriction de g (resp. de f_i) à l'hyperplan $\operatorname{Ker} f_{p+1}$. On a clairement $\bigcap_{i=1}^{p} \operatorname{Ker} \tilde{f}_i \subset \operatorname{Ker} \tilde{g}$. Par hypothèse de récurrence, il existe des scalaires $(\lambda_1, \dots \lambda_p)$ tels que $\tilde{g} = \sum_{i=1}^{p} \lambda_i \tilde{f}_i$. La forme linéaire $g - \sum_{i=1}^{p} \lambda_i f_i$ est alors nulle sur $\operatorname{Ker} f_{p+1}$; elle est donc proportionnelle à f_{p+1}, ce qui permet de conclure. ◁

La solution de l'exercice suivant exploite les propriétés de l'orthogonalité duale. Nous donnerons une seconde solution par récurrence, dans le chapitre sur les déterminants du tome 2 d'algèbre (exercice 1.16).

6.28. Familles libres d'applications

1. Soit K un corps commutatif et $f_1, \dots, f_n$ des applications de K dans K. Montrer que la famille $(f_1, \dots, f_n)$ est libre dans $\mathcal{F}(K, K)$ si et seulement s'il existe $(x_1, \dots, x_n) \in K^n$ tel que la matrice $(f_i(x_j))_{1 \leqslant i,j \leqslant n}$ soit inversible.

2. Déterminer les applications $f : \mathbb{R} \longrightarrow \mathbb{R}$, dérivables, dont les translatées (*i.e.* les applications $f_a : x \longmapsto f(x + a)$ pour $a \in \mathbb{R}$) engendrent un sous-espace vectoriel de dimension finie de $\mathcal{F}(\mathbb{R}, \mathbb{R})$.

(École polytechnique)

▷ **Solution.**

1. • Si la famille $(f_1, \dots, f_n)$ est liée, les lignes de la matrice $(f_i(x_j))_{1 \leqslant i,j \leqslant n}$ sont liées, quel que soit le choix des scalaires $x_1, \dots, x_n$. La contraposée fournit donc une des deux implications.

• Supposons que la famille $B = (f_1, \dots, f_n)$ est libre et notons F le sous-espace de dimension n qu'elle engendre. Pour tout $a \in K$, l'application d'évaluation en a, $e_a : F \to K$, qui à $f \in F$ associe $f(a) \in K$, est une forme linéaire sur F. L'ensemble A des formes linéaires e_a, pour $a \in K$, constitue une partie génératrice de F^*. En effet, si $f \in A^\circ$, on a $f(a) = e_a(f) = 0$ pour tout $a \in K$, *i.e.* $f = 0$. Ainsi on a $A^\circ = \{0\}$

et $\mathrm{Vect}\,\mathrm{A} = ((\mathrm{Vect}\,\mathrm{A})^\circ)^\perp = (\mathrm{A}^\circ)^\perp = 0^\perp = \mathrm{F}$, puisqu'on est en dimension finie. On peut donc choisir des scalaires $x_1, \ldots, x_n$ tels que les formes linéaires $e_{x_1}, \ldots, e_{x_n}$ constituent une base de F^*. Montrons que le n-uplet $(x_1, \ldots, x_n)$ répond à la question.

Considérons la matrice $\mathrm{M} = (f_i(x_j))_{1 \leqslant i,j \leqslant n}$ et montrons que les lignes $\mathrm{L}_1, \ldots, \mathrm{L}_n$ de M forment une famille libre. Soit $(\lambda_1, \ldots, \lambda_n) \in \mathrm{K}^n$ tel que $\sum_{i=1}^{n} \lambda_i \mathrm{L}_i = 0$. On a alors, pour tout $j \in [\![1, n]\!]$, $\sum_{i=1}^{n} \lambda_i f_i(x_j) = 0$, c'est-à-dire $e_{x_j}\left(\sum_{i=1}^{n} \lambda_i f_i\right) = 0$. La famille $e_{x_1}, \ldots, e_{x_n}$ étant une base de F^*, on a donc $\sum_{i=1}^{n} \lambda_i f_i \in (\mathrm{F}^*)^\circ = \{0\}$. La famille $(f_1, \ldots, f_n)$ étant une base de F, on en déduit que $\lambda_1 = \cdots = \lambda_n = 0$. La matrice M est donc inversible.

2. Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$, dérivable, dont les translatées engendrent un $\mathbb{R}$-espace vectoriel F de dimension finie n. On considère des réels $a_1, \ldots, a_n$ tels que la famille $(f_{a_1}, \ldots, f_{a_n})$ soit une base de F. D'après la question 1, il existe des réels $x_1, \ldots, x_n$ tels que la matrice $\mathrm{M} = (f_{a_i}(x_j))_{1 \leqslant i,j \leqslant n}$ soit inversible. La fonction f étant dérivable, les fonctions f_{a_i} le sont également. Ainsi tout élément de F est dérivable.

Soit g un élément quelconque de F. Montrons que g' est encore dans F. Il est clair que, pour tout $a \in \mathbb{R}$, la fonction g_a est dans F. En effet, on a $g_a \in \mathrm{Vect}(f_{a_1+a}, \ldots, f_{a_n+a}) \subset \mathrm{F}$. Il existe donc des réels $\lambda_1(a), \ldots, \lambda_n(a)$ tels que $g_a = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}$. Montrons que les fonctions λ_i sont dérivables. On a, pour $1 \leqslant j \leqslant n$,

$$g(a + x_j) = g_a(x_j) = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}(x_j).$$

Matriciellement, cela s'écrit $\begin{pmatrix} g(a+x_1) \\ \vdots \\ g(a+x_n) \end{pmatrix} = {}^t\mathrm{M} \begin{pmatrix} \lambda_1(a) \\ \vdots \\ \lambda_n(a) \end{pmatrix}$. On en déduit que $\begin{pmatrix} \lambda_1(a) \\ \vdots \\ \lambda_n(a) \end{pmatrix} = {}^t\mathrm{M}^{-1} \begin{pmatrix} g(a+x_1) \\ \vdots \\ g(a+x_n) \end{pmatrix}$. Les coefficients de la matrice ${}^t\mathrm{M}^{-1}$ étant indépendants de a, on en déduit que les fonctions $\lambda_1, \ldots, \lambda_n$ sont des combinaisons linéaires des fonctions $g_{x_1}, \ldots, g_{x_n}$. Elles sont dérivables comme ces dernières.

On a, pour tout réel x, $g(x + a) = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}(x)$. En dérivant par

et $\mathrm{Vect}\,\mathrm{A} = ((\mathrm{Vect}\,\mathrm{A})^\circ)^\perp = (\mathrm{A}^\circ)^\perp = 0^\perp = \mathrm{F}$, puisqu'on est en dimension finie. On peut donc choisir des scalaires $x_1, \ldots, x_n$ tels que les formes linéaires $e_{x_1}, \ldots, e_{x_n}$ constituent une base de F^*. Montrons que le n-uplet $(x_1, \ldots, x_n)$ répond à la question.

Considérons la matrice $\mathrm{M} = (f_i(x_j))_{1 \leqslant i,j \leqslant n}$ et montrons que les lignes $\mathrm{L}_1, \ldots, \mathrm{L}_n$ de M forment une famille libre. Soit $(\lambda_1, \ldots, \lambda_n) \in \mathrm{K}^n$ tel que $\sum_{i=1}^{n} \lambda_i \mathrm{L}_i = 0$. On a alors, pour tout $j \in [\![1, n]\!]$, $\sum_{i=1}^{n} \lambda_i f_i(x_j) = 0$, c'est-à-dire $e_{x_j}\left(\sum_{i=1}^{n} \lambda_i f_i\right) = 0$. La famille $e_{x_1}, \ldots, e_{x_n}$ étant une base de F^*, on a donc $\sum_{i=1}^{n} \lambda_i f_i \in (\mathrm{F}^*)^\circ = \{0\}$. La famille $(f_1, \ldots, f_n)$ étant une base de F, on en déduit que $\lambda_1 = \cdots = \lambda_n = 0$. La matrice M est donc inversible.

2. Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$, dérivable, dont les translatées engendrent un $\mathbb{R}$-espace vectoriel F de dimension finie n. On considère des réels $a_1, \ldots, a_n$ tels que la famille $(f_{a_1}, \ldots, f_{a_n})$ soit une base de F. D'après la question 1, il existe des réels $x_1, \ldots, x_n$ tels que la matrice $\mathrm{M} = (f_{a_i}(x_j))_{1 \leqslant i,j \leqslant n}$ soit inversible. La fonction f étant dérivable, les fonctions f_{a_i} le sont également. Ainsi tout élément de F est dérivable.

Soit g un élément quelconque de F. Montrons que g' est encore dans F. Il est clair que, pour tout $a \in \mathbb{R}$, la fonction g_a est dans F. En effet, on a $g_a \in \mathrm{Vect}(f_{a_1+a}, \ldots, f_{a_n+a}) \subset \mathrm{F}$. Il existe donc des réels $\lambda_1(a), \ldots, \lambda_n(a)$ tels que $g_a = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}$. Montrons que les fonctions λ_i sont dérivables. On a, pour $1 \leqslant j \leqslant n$,

$$g(a + x_j) = g_a(x_j) = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}(x_j).$$

Matriciellement, cela s'écrit $\begin{pmatrix} g(a+x_1) \\ \vdots \\ g(a+x_n) \end{pmatrix} = {}^t\mathrm{M} \begin{pmatrix} \lambda_1(a) \\ \vdots \\ \lambda_n(a) \end{pmatrix}$. On en déduit que $\begin{pmatrix} \lambda_1(a) \\ \vdots \\ \lambda_n(a) \end{pmatrix} = {}^t\mathrm{M}^{-1} \begin{pmatrix} g(a+x_1) \\ \vdots \\ g(a+x_n) \end{pmatrix}$. Les coefficients de la matrice ${}^t\mathrm{M}^{-1}$ étant indépendants de a, on en déduit que les fonctions $\lambda_1, \ldots, \lambda_n$ sont des combinaisons linéaires des fonctions $g_{x_1}, \ldots, g_{x_n}$. Elles sont dérivables comme ces dernières.

On a, pour tout réel x, $g(x + a) = \sum_{i=1}^{n} \lambda_i(a) f_{a_i}(x)$. En dérivant par

rapport à a, on obtient $g'(x+a) = \sum_{i=1}^{n} \lambda_i'(a) f_{a_i}(x)$. On obtient en particulier, en prenant $a = 0$, $g' = \sum_{i=1}^{n} \lambda_i'(0) f_{a_i} \in \mathrm{F}$. On en déduit, de manière immédiate, que tout élément g de F est $\mathcal{C}^\infty$ et que, pour tout entier naturel k, on a $g^{(k)} \in \mathrm{F}$. C'est vrai en particulier pour la fonction f. L'espace vectoriel F étant de dimension finie n, il existe un entier p $(1 \leqslant p \leqslant n)$ tel que $f^{(p)} \in \mathrm{Vect}(f, f', \ldots, f^{(p-1)})$. La fonction f est solution d'une équation linéaire homogène à coefficients constants d'ordre p.

Réciproquement, si f est solution d'une équation linéaire homogène à coefficients constants d'ordre p, il est clair que, pour tout $a \in \mathbb{R}$, la fonction f_a est solution de la même équation différentielle. L'ensemble des solutions de cette équation différentielle étant un espace vectoriel de dimension p, les fonctions f_a engendrent un espace vectoriel de dimension finie inférieure ou égale à p.

Conclusion. Les translatés de f engendrent un espace vectoriel de dimension finie si, et seulement si, f est solution d'une équation linéaire homogène à coefficients constants. ◁

Le lecteur trouvera d'autres exercices sur la dualité dans le chapitre 7 sur les matrices.

Les deux exercices qui suivent concernent la notion de famille positivement génératrice dans un espace vectoriel réel.

6.29. Familles positivement génératrices

Soit un espace vectoriel réel E de dimension $n \geqslant 1$ et une famille $\mathcal{F} = (e_1, e_1, \ldots, e_p)$ de vecteurs de E *positivement génératrice*, c'est-à-dire telle que pour tout $x \in \mathrm{E}$, il existe $(\lambda_1, \ldots, \lambda_p) \in (\mathbb{R}_+^*)^p$ tel que $x = \lambda_1 e_1 + \cdots + \lambda_p e_p$.

1. Montrer que $p \geqslant n+1$. Donner un exemple de famille positivement génératrice de cardinal $n+1$.

2. On suppose $p \geqslant 2n+1$. Montrer qu'il existe une sous-famille stricte de $\mathcal{F}$ qui est encore positivement génératrice. Donner un exemple de famille positivement génératrice de cardinal $2n$ dont aucune sous-famille stricte ne l'est.

(École normale supérieure)

▷ **Solution.**

1. Comme la famille $\mathcal{F}$ est génératrice, on a $p \geqslant n$. Si $p = n$, $\mathcal{F}$ est une base de E, et alors le vecteur $-e_1$ (par exemple) ne saurait être

obtenu comme combinaison à coefficients strictement positifs des e_i (par unicité de la décomposition). On a donc $p \geqslant n+1$.

Donnons-nous une base quelconque $(f_1, \dots, f_n)$ de E et montrons que la famille $(f_1, \dots, f_n, -f_1 - \cdots - f_n)$ est positivement génératrice. En effet, soit $x \in \mathrm{E}$, $x = x_1 f_1 + \cdots + x_n f_n$. Pour tout réel t on a

$$x = (x_1 + t)f_1 + \cdots + (x_n + t)f_n + t\Big(-\sum_{i=1}^{n} f_i\Big).$$

Il suffit de choisir t assez grand pour que tous les coefficients soient strictement positifs.

Notons que l'on montre de la même façon que toute famille génératrice $(f_1, \dots, f_p)$ pour laquelle on a une relation de liaison $\lambda_1 f_1 + \cdots + \lambda_p f_p = 0$, avec des coefficients λ_i strictement positifs, est une famille positivement génératrice.

2. La famille $\mathcal{F}$ étant de rang n, il existe $\mathrm{I} \subset [\![1, p]\!]$ de cardinal n tel que $(e_i)_{i\in\mathrm{I}}$ soit une base de E. Posons $\mathrm{J} = [\![1, p]\!] \setminus \mathrm{I}$. Alors, la famille $(e_j)_{j\in\mathrm{J}}$ est liée car $|\mathrm{J}| \geqslant n+1$; on peut donc écrire une relation de liaison $\sum_{j\in\mathrm{J}} \lambda_j e_j = 0$, où les λ_j ne sont pas tous nuls. Par ailleurs, puisque la famille $(e_i)_{1\leqslant i\leqslant p}$ est positivement génératrice, il existe des coefficients strictement positifs $t_1, \dots, t_p$ tels que $t_1 e_1 + \cdots + t_p e_p = 0$. Pour tout réel x on a donc

$$\sum_{i\in\mathrm{I}} t_i e_i + \sum_{j\in\mathrm{J}} (t_j + x\lambda_j)e_j = 0.$$

On prend $x = -\dfrac{t_k}{\lambda_k}$, où $k \in \mathrm{J}$ est choisi de sorte que $\lambda_k \neq 0$ et que la valeur absolue du quotient $\dfrac{t_k}{\lambda_k}$ soit minimale. Pour cette valeur de x les coefficients $(t_j + x\lambda_j)$ restent tous positifs ou nuls. Notons K la partie formée des indices $j \in \mathrm{J}$ tels que $t_j + x\lambda_j > 0$. Il s'agit d'une partie stricte de J, car $k \notin \mathrm{K}$. La remarque faite à la fin de la question 1 montre que la famille $(e_i)_{i\in\mathrm{I}\cup\mathrm{K}}$ est encore positivement génératrice.

Si $(f_1, \dots, f_n)$ est une base quelconque de E, alors la famille $(f_1, \dots, f_n, -f_1, \dots, -f_n)$ est de cardinal $2n$, positivement génératrice, mais aucune de ses sous-familles ne l'est. ◁

L'exercice suivant va nous fournir une description des familles positivement génératrices du dual.

6.30. Familles positivement génératrices de E*

Soit E un $\mathbb{R}$-espace vectoriel de dimension n et p éléments $f_1, \dots, f_p$ de son dual.

1. Montrer l'équivalence des deux propositions suivantes :

(i) $\forall x \in \mathrm{E}$, $\min\limits_{1 \leqslant i \leqslant p} f_i(x) \leqslant 0$;

(ii) $\exists (a_1, \dots, a_p) \in (\mathbb{R}_+)^p - \{0\}$, $\sum\limits_{i=1}^{p} a_i f_i = 0$.

2. Montrer l'équivalence des deux propositions suivantes :

(i') $\forall x \in \mathrm{E} - \{0\}$, $\min\limits_{1 \leqslant i \leqslant p} f_i(x) < 0$;

(ii') $(f_1, \dots, f_p)$ est une famille positivement génératrice de E*.

(École normale supérieure)

▷ **Solution.**

1. • Supposons (ii) vérifiée et raisonnons par l'absurde. Soit $x \in \mathrm{E}$ tel que, pour $1 \leqslant i \leqslant p$, on ait $f_i(x) > 0$. On a $\sum\limits_{i=1}^{p} a_i f_i(x) = 0$. Les termes de cette somme étant tous positifs, on en déduit que, pour $1 \leqslant i \leqslant p$, on a $a_i f_i(x) = 0$ et donc $a_i = 0$, car $f_i(x) > 0$. On a donc $(a_1, \dots, a_p) = 0$, ce qui est contraire à l'hypothèse.

• Pour montrer que (i) implique (ii), raisonnons par récurrence sur p.

⋆ Si $p = 1$, alors, pour tout $x \in \mathrm{E}$, on a $f_1(x) \leqslant 0$ et $f_1(-x) \leqslant 0$ et donc $f_1(x) = 0$: f_1 est l'application nulle et $af_1 = 0$ pour tout $a > 0$.

⋆ Supposons que la famille $(f_1, \dots, f_p)$ vérifie la propriété (i). Si f_p est l'application nulle, on conclut comme précédemment. Si $(f_1, \dots, f_{p-1})$ vérifie (i), on applique directement l'hypothèse de récurrence et on prend $a_p = 0$. C'est fini. Ces cas étant écartés, soit H le noyau de f_p, et $e \in \mathrm{E}$ tel que $f_p(e) = 1$.

Considérons $x \in \mathrm{H}$ et $\alpha > 0$. Posons $y = x + \alpha e$. On obtient alors $f_p(y) = \alpha > 0$ et pour $1 \leqslant i \leqslant p-1$, $f_i(y) = f_i(x) + \alpha f_i(e)$. Supposons qu'il existe $x \in \mathrm{H}$ tel que $\min\limits_{1 \leqslant i \leqslant p-1} f_i(x) > 0$. Pour $\alpha > 0$, assez petit, on a alors $\min\limits_{1 \leqslant i \leqslant p-1} f_i(y) > 0$ et donc $\min\limits_{1 \leqslant i \leqslant p} f_i(y) > 0$, ce qui est contraire à l'hypothèse. On a donc, pour tout $x \in \mathrm{H}$, $\min\limits_{1 \leqslant i \leqslant p-1} f_i(x) \leqslant 0$. La famille $(g_1, \dots, g_{p-1})$, où $g_i = f_{i|\mathrm{H}}$, vérifie (i). D'après l'hypothèse de récurrence, il existe $(a_1, \dots, a_{p-1}) \in \mathbb{R}_+^{p-1} \setminus \{0\}$ tel que $\sum\limits_{i=1}^{p-1} a_i g_i = 0$. On a donc, pour tout $x \in \mathrm{H}$, $\sum\limits_{i=1}^{p-1} a_i f_i(x) = 0$. En posant $a_p = -\sum\limits_{i=1}^{p-1} a_i f_i(e)$, on obtient

finalement $\sum_{i=1}^{p} a_i f_i = 0$, puisque cette application linéaire s'annule en e et sur H et que $E = H \oplus \mathbb{R}e$.

Il reste à démontrer que $a_p \geqslant 0$. Raisonnons par l'absurde et supposons $a_p < 0$. La famille $(f_1, \ldots, f_{p-1}, -f_p)$ vérifie (ii) et donc aussi (i). Soit y un élément quelconque de E, $x \in H$ et $\alpha \in \mathbb{R}$ tel que $y = x + \alpha e$.

• Si $\alpha > 0$, on a $\min_{1 \leqslant i \leqslant p-1} f_i(y) = \min_{1 \leqslant i \leqslant p} f_i(y) \leqslant 0$, car $f_p(y) = \alpha > 0$.

• Si $\alpha < 0$, on obtient $\min_{1 \leqslant i \leqslant p-1} f_i(y) = \min(f_1(y), \ldots, f_{p-1}(y), -f_p(y))$ négatif ou nul, car $-f_p(y) = -\alpha > 0$.

Par continuité, la propriété reste vraie pour $\alpha = 0$ et on obtient pour tout $y \in E$, $\min_{1 \leqslant i \leqslant p-1} f_i(y) \leqslant 0$, cas qui a été écarté. On a donc $a_p \geqslant 0$, ce qui termine la démonstration.

2. • Supposons la propriété (i') vérifiée. Soit $f \in E^*$. Nous allons démontrer que pour $\lambda > 0$ assez petit les fonctions $(f_1 - \lambda f, \ldots, f_p - \lambda f)$ vérifient l'hypothèse (i) et appliquer le résultat de la première question.

Munissons E d'une norme quelconque (elles sont toutes équivalentes, car nous sommes en dimension finie) et notons S la sphère unité de E : S est compacte. Les fonctions f_i sont linéaires donc continues, puisque E est de dimension finie. On en déduit que la fonction $g = \min_{1 \leqslant i \leqslant p} f_i$ est elle aussi continue (cela se montre aisément par récurrence sur p, à partir de l'égalité $\min(f_1, f_2) = \frac{f_1 + f_2}{2} - \frac{|f_1 - f_2|}{2}$). Comme g est continue sur le compact S, elle y est majorée et sa borne supérieure M est atteinte. De plus, g étant strictement négative sur S, on a $M < 0$.

Pour les mêmes raisons, $|f|$ est majorée sur le compact S, par $a > 0$. Soit $\lambda > 0$ et $h = \min_{1 \leqslant i \leqslant p} f_i - \lambda f$. On a, pour tout $x \in S$ et $1 \leqslant i \leqslant p$,

$$f_i(x) - \lambda f(x) \leqslant f_i(x) + \lambda a \text{ (car } f(x) \geqslant -a).$$

On en déduit que, pour tout $x \in S$,

$$h(x) \leqslant g(x) + \lambda a \leqslant M + \lambda a.$$

Si donc on choisit $\lambda < -\frac{M}{a}$, alors, on a, pour tout $x \in S$, $h(x) < 0$. D'autre part, pour tout $x \in E - \{0\}$, il existe $u \in S$ et $\alpha \in \mathbb{R}_+^*$ tels que $x = \alpha u$. On a alors $h(x) = \alpha h(u) < 0$ de sorte que $h(x) \leqslant 0$ pour tout $x \in E$.

Les fonctions $g_1, \ldots, g_p$ définies par $g_i(x) = f_i(x) - \lambda f(x)$, pour $x \in E$ et $1 \leqslant i \leqslant p$ vérifient alors la condition (i) de la première question. On en déduit qu'il existe $(a_1, \ldots, a_p) \in \mathbb{R}_+^p - \{0\}$ tel que

$$\sum_{i=1}^{p} a_i g_i = \sum_{i=1}^{p} a_i (f_i - \lambda f) = 0.$$

ce qui peut s'écrire

$$f = \frac{1}{\lambda \sum_{i=1}^{p} a_i} \sum_{i=1}^{p} a_i f_i,$$

puisque $\lambda \sum_{i=1}^{p} a_i > 0$, ($\lambda > 0$ et les a_i sont positifs et non tous nuls).

Ainsi, tout élément $f \in \mathrm{E}^*$ s'écrit comme combinaison linéaire à coefficients positifs de $(f_1, \ldots, f_p)$. En considérant $f - \sum_{i=1}^{p} f_i$, au lieu de f, on peut même supposer que les coefficients sont strictement positifs.

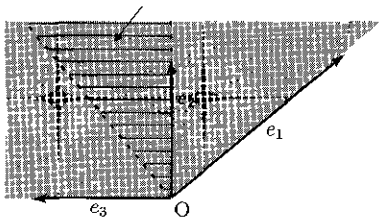
Cela prouve donc que $(f_1, \ldots, f_p)$ est une famille positivement génératrice de E^*.

• Supposons que tout élément de E^* soit combinaison linéaire à coefficients strictement positifs de $f_1, \ldots, f_p$. Soit $x \in \mathbf{E}$ et $f \in \mathbf{E}^*$ tel que $f(x) = -1$. Il existe $(a_1, \ldots, a_p) \in (\mathbb{R}_+^*)^p$ tel que $f = \sum_{i=1}^{p} a_i f_i$. On ne peut avoir $\min_{1 \leqslant i \leqslant p} f_i(x) \geqslant 0$, car sinon $f(x) \geqslant 0$. On a donc $\min_{1 \leqslant i \leqslant p} f_i(x) < 0$ et (i') est vérifiée. ◁

On peut en déduire une caractérisation géométrique des familles positivement génératrices de E *moyennant l'identification de* $\mathbf{E}$ *avec son bidual* E^{**} *qui est le dual de* E^*. *On rappelle que l'application qui à tout vecteur* $x \in \mathrm{E}$ *associe l'application* $\hat{x} \in \mathrm{E}^{**}$ *définie par* $\hat{x}(f) = f(x)$ *pour tout* $f \in \mathrm{E}^*$ *est un isomorphisme. L'équivalence de la question 2 se traduit ainsi : une famille* $(e_1, \ldots, e_p)$ *de* E *est positivement génératrice si, pour tout* $f \in \mathrm{E}^*$, *il existe* $i \in [\![1, p]\!]$ *tel que* $f(e_i) < 0$.

Si E *est muni d'une structure euclidienne, les formes linéaires sur* $\mathbf{E}$ *sont les* $f_x : y \longmapsto (x, y)$. *Il en résulte que* $(e_1, \ldots, e_p)$ *de* E *est positivement génératrice si tout vecteur non nul* x *de* $\mathbf{E}$, *il existe* $i \in [\![1, p]\!]$ *tel que* $(x, e_i) < 0$. *Autrement dit, si* $\mathbf{E}$ *la réunion des* H_{e_i}, *où* H_y *désigne le demi-espace ouvert défini par* $\mathrm{H}_y = \{x \in \mathrm{E}, \langle x, y \rangle < 0\}$ *(on pose* $\mathrm{H}_0 = \emptyset$*).*

Voici un exemple dans $\mathbb{R}^2$ *: en grisé figure le cône positif engendré par* e_1, e_2 *et* e_3, *c'est-à-dire l'ensemble des combinaisons linéaires à coefficients positifs de* e_1, e_2, e_3. *La famille n'est pas positivement génératrice car la zone grisée n'est pas recouverte par les* H_{e_i}.



Les derniers exercices de ce chapitre auront pour cadre des algèbres.

6.31. Caractérisation de $\mathbb{C}$

Soit K une algèbre sans diviseur de zéro sur $\mathbb{R}$ de dimension finie $n \geqslant 2$. On identifie $\mathbb{R}$ avec $\mathbb{R}.1$ où 1 est l'élément neutre de K pour le produit.

1. Montrer que tout élément non nul de K est inversible.

2. Soit $a \in \mathrm{K}$ n'appartenant pas à $\mathbb{R}$. Montrer que $(1, a)$ est libre et que $(1, a, a^2)$ est liée.

3. Montrer l'existence de $i \in \mathrm{K}$ tel que $i^2 = -1$.

4. Montrer que si K est commutative alors K est isomorphe à $\mathbb{C}$.

(École normale supérieure)

▷ **Solution.**

1. Soit a un élément non nul de K. Par intégrité de K, l'application $\mathbb{R}$-linéaire $x \mapsto ax$ est injective, donc bijective puisque K est de dimension finie. Il existe donc $b \in \mathrm{K}$ tel que $ab = 1$ ce qui montre que a est inversible à droite. De la même manière on prouve l'existence de c tel que $ca = 1$. En multipliant cette dernière égalité à droite par b on a $c = b$. Donc a est inversible.

2. Dire que a n'est pas réel revient à dire que la famille $(1, a)$ est $\mathbb{R}$-libre. Notons qu'un tel élément a existe car $n = \dim \mathrm{K} \geqslant 2$. Comme K est de dimension finie, la famille $(a^k)_{k \geqslant 0}$ est liée et l'idéal des polynômes $\mathrm{P} \in \mathbb{R}[\mathrm{X}]$ tels que $\mathrm{P}(a) = 0$ est non nul. Notons Q le générateur unitaire de cet idéal (c'est le polynôme minimal de a). Comme K est intègre, Q est irréductible sur $\mathbb{R}$. Comme $a \notin \mathbb{R}$, le degré de Q n'est pas égal à 1. Par suite Q est de degré 2 et on en déduit que la famille $(1, a, a^2)$ est liée.

3. Notons qu'un réel λ commute avec tout élément x de K : en effet, comme K est une $\mathbb{R}$-algèbre, on peut écrire

$$\lambda x = (\lambda 1)x = \lambda(1x) = \lambda(x1) = x(\lambda 1) = x\lambda.$$

Soit toujours a choisi comme précédemment et α, β les réels tels que $a^2 = \alpha + \beta a$. Bien entendu α est non nul car $a \notin \mathbb{R}$. On va chercher notre élément i dans le plan $\mathrm{Vect}(1, a)$. Pour $(x, y) \in \mathbb{R}^2$ on a

$$(x + ay)^2 = x^2 + 2xya + y^2a^2 = (x^2 + \alpha y^2) + (2xy + \beta y^2)a.$$

On cherche donc x, y tels que $\begin{cases} x^2 + \alpha y^2 = -1 \\ 2xy + \beta y^2 = 0 \end{cases}$. On doit prendre y non nul, ce qui impose $x = -\dfrac{\beta y}{2}$. En substituant dans la première équation

on obtient $y^2\left(\alpha+\dfrac{\beta^2}{4}\right)=-1$. Cette équation admet bien deux solutions car $\alpha+\dfrac{\beta^2}{4}<0$ puisque le polynôme $Q(X)=X^2-\beta X-\alpha$ est irréductible sur $\mathbb{R}$. Il existe donc deux éléments (opposés) i dans $\mathrm{Vect}(1,a)$ tels que $i^2=-1$.

4. Le plan vectoriel $\mathrm{Vect}(1,i)=\mathrm{Vect}(1,a)$ est clairement une sous-algèbre de K qui est isomorphe à $\mathbb{C}$. On va prouver que si K est supposée commutative alors $K=\mathrm{Vect}(1,i)$. Soit b un élément quelconque de $K\setminus\mathbb{R}$. La question précédente appliquée avec b à la place de a montre l'existence d'une racine carrée de -1 dans le plan $\mathrm{Vect}(1,b)$. Or, comme K est commutative et intègre, l'équation $z^2=-1$ n'a que deux solutions à savoir i et $-i$. On a donc forcément $\mathrm{Vect}(1,b)=\mathrm{Vect}(1,a)=\mathrm{Vect}(1,i)$. Par suite $K=\mathrm{Vect}(1,i)$ et K est isomorphe à $\mathbb{C}$. ◁

L'exercice suivant offre une version plus complète qui caractérise toutes les algèbres réelles sans diviseur de zéro et de dimension finie.

6.32. Théorème de Frobenius

Soit A une $\mathbb{R}$-algèbre sans diviseurs de 0 et de dimension finie $n\geqslant 2$. On identifie $\mathbb{R}$ avec $\mathbb{R}1$ où 1 est l'élément neutre de A pour le produit.

1. Soit $x\in A$. Montrer qu'il existe $(a,b)\in\mathbb{R}^2$ tel que

$$x^2+ax+b=0.$$

2. Soit $x\in A$ tel que $x^2\in\mathbb{R}^+$. Montrer que $x\in\mathbb{R}$.

3. Soit $V=\{x\in A,\ x^2\in\mathbb{R}^-\}$. Montrer que tout $x\in A$ s'écrit de manière unique $x=r+v$ avec $r\in\mathbb{R}$ et $v\in V$.

4. Soit (u,v) une famille libre d'éléments de V et $\lambda\in\mathbb{R}$. Montrer que $u-\lambda v\notin\mathbb{R}$.

5. Montrer que V est un sous-espace vectoriel de A.

6. On suppose A de dimension 2. Montrer que A est isomorphe à $\mathbb{C}$.

7. On pose $\langle u,v\rangle=-(uv+vu)$ pour tout $(u,v)\in A^2$. Montrer que cela définit un produit scalaire sur V.

8. On suppose que $\dim A\geqslant 3$. Montrer qu'il existe i,j dans V tels que $\langle i,j\rangle=0$, $i^2=-1$, $j^2=-1$ et que $(1,i,j,ij)$ est une $\mathbb{R}$-base de A.

(École normale supérieure)

▷ **Solution.**

1. Si x^2 est réel, on prend $a=0$ et $b=-x^2$. Sinon x n'est pas réel et la famille $(1,x)$ est libre. Comme A de dimension finie, la famille $(x^k)_{k\in\mathbb{N}}$ est liée et il existe $P\in\mathbb{R}[X]$, non nul tel que $P(x)=0$. On

décompose P en produit de polynômes irréductibles $P = P_1 P_2 \dots P_n$. Il existe $k \in [\![1, n]\!]$ tel que $P_k(x) = 0$, car A n'a pas de diviseur de 0. Le polynôme P_k est de degré 2 puisque la famille $(1, x)$ est libre. En divisant par le coefficient de x^2, on obtient $(a, b) \in \mathbb{R}^2$ tel que $x^2 + ax + b = 0$.

2. Notons qu'un élément de A commute avec tout réel. En effet, si $x \in \mathbb{R}$ et $\lambda \in \mathbb{R}$, $x\lambda = x(\lambda 1) = \lambda(x1) = \lambda x$, car A est une $\mathbb{R}$-algèbre.

Soit $x \in A$ tel que $x^2 \in \mathbb{R}_+$ et λ une racine carrée réelle de x^2. On a alors $x^2 - \lambda^2 = 0$. Mais $x^2 - \lambda^2 = (x - \lambda)(x + \lambda)$ car $\lambda x = x\lambda$. On a donc $x = \pm\lambda$, car A n'a pas de diviseur de 0, et x est dans $\mathbb{R}$.

3. Soit $x \in A$.

• Pour l'existence, on considère $(a, b) \in \mathbb{R}^2$ tel que $x^2 + ax + b = 0$. On peut écrire cette égalité $\left(x + \frac{a}{2}\right)^2 = \frac{a^2 - 4b}{4}$ (car x commute avec $\frac{a}{2}$).

Si $\frac{a^2 - 4b}{4} \in \mathbb{R}_+$, alors d'après la question 2, $x + \frac{a}{2}$ est réel donc x est réel. Il s'écrit $x = x + 0$, avec $x \in \mathbb{R}$ et $0 \in V$.

Si $\frac{a^2 - 4b}{4} \in \mathbb{R}_-$, alors $v = x + \frac{a}{2} \in V$ et $x = -\frac{a}{2} + v$.

• Montrons l'unicité de cette décomposition. Soit $(r, r') \in \mathbb{R}^2$, $(v, v') \in V^2$ tels que $r + v = r' + v'$. En élevant au carré, on en déduit que $v'^2 = (r - r' + v)^2 = (r - r')^2 + v^2 + 2(r - r')v$. Les carrés sont réels, donc si $r \neq r'$, v est réel et donc $v = 0$, puisque $v^2 \leqslant 0$. On obtient $v'^2 = (r - r')^2 \in \mathbb{R}_+$ et donc $v' = 0$, puis $r = r'$, ce qui est contradictoire. On conclut : $r = r'$ et $v = v'$.

4. Si $v \in V$ et $\lambda \in \mathbb{R}$, alors $(\lambda v)^2 = \lambda^2 v^2 \in \mathbb{R}_-$ et donc $\lambda v \in V$. S'il existe $r \in \mathbb{R}$ tel que $u = \lambda v + r$, par l'unicité de la décomposition étudiée dans la question 3, on obtient $r = 0$ et $u = \lambda v$, ce qui contredit la liberté de la famille (u, v).

5. L'ensemble V contient 0.

Si $v \in V$ et $\lambda \in V$, on a montré dans la question 4 que $\lambda v \in V$.

Soit u et v deux éléments de V. Si la famille (u, v) est liée, il existe par exemple $\lambda \in \mathbb{R}$ tel que $v = \lambda u$. On a alors $u + v = (1 + \lambda)u \in V$. Supposons donc la famille (u, v) libre.

En particulier, u n'est pas nul. Montrons que u est inversible. L'application $x \longmapsto ux$ est $\mathbb{R}$-linéaire. Son noyau est réduit à $\{0\}$ car A n'a pas de diviseur de 0. On en déduit qu'elle est bijective car A est de dimension finie. Ainsi il existe $u' \in A$ tel que $uu' = 1$. On montre de même qu'il existe $u'' \in A$ tel que $u''u = 1$. En considérant $u''uu'$, on montre que $u' = u''$, donc que u est inversible.

Considérons le produit uvu^{-1}. On a

$$(uvu^{-1})^2 = uv^2u^{-1} = v^2uu^{-1} = v^2,$$

car v^2 est réel, donc $uvu^{-1} = \pm v$, c'est-à-dire $uv = \pm vu$.

Si $uv = -vu$, $(u + v)^2 = u^2 + v^2 \in \mathbb{R}_-$ et $u + v \in V$. Si $uv = vu$, $(uv)^2 = u^2v^2 \in \mathbb{R}_+$ et $uv \in \mathbb{R}$. Mais alors $(u + v)^2 = u^2 + v^2 + 2uv \in \mathbb{R}$

et comme d'après la question précédente, $u+v \notin \mathbb{R}$, $(u+v)^2 \in \mathbb{R}_-$ et $u+v \in \mathrm{V}$. Dans tous les cas, $u+v \in \mathrm{V}$ et V est un sous-espace vectoriel de A.

6. Soit v un élément non nul de V (un tel élément existe sinon on a $\mathrm{A} = \mathbb{R}$). Comme $v^2 \in \mathbb{R}_-^*$, il existe $\lambda > 0$ tel que $v^2 = -\lambda^2$. Alors $\left(\dfrac{v}{\lambda}\right)^2 = -1$. On note $i = \dfrac{v}{\lambda}$. La famille $(1, i)$ est libre donc c'est une base de A et $\mathrm{A} = \mathbb{R} + \mathbb{R}i$ est isomorphe à $\mathbb{C}$.

7. Pour tout couple $(u, v) \in \mathrm{V}^2$, on a $\langle u, v\rangle = u^2 + v^2 - (u+v)^2 \in \mathbb{R}$, car $u+v \in \mathrm{V}$ donc chaque carré est réel.

L'application $(u, v) \longmapsto \langle u, v\rangle$ est clairement bilinéaire et symétrique. Cela résulte des propriétés d'une algèbre.

Enfin $\langle u, u\rangle = -2u^2 \in \mathbb{R}_+$ car $u^2 \in \mathbb{R}_-$ et $\langle u, u\rangle = 0$ implique $u = 0$, car A n'a pas de diviseur de 0. On a bien un produit scalaire.

8. On montre comme dans la question 6 qu'il existe $i \in \mathrm{V}$ tel que $i^2 = -1$. D'après la question 3, on a $\mathrm{A} = \mathbb{R} \oplus \mathrm{V}$. Comme $\dim \mathrm{A} \geqslant 3$, on a $\dim \mathrm{V} \geqslant 2$ et $\dim i^\perp \geqslant 1$. Soit $v \in i^\perp$, non nul. Il existe $\mu \in \mathbb{R}_+^*$ tel que $v^2 = -\mu^2$ et $j = \dfrac{v}{\mu}$ vérifie $j^2 = -1$ et $\langle i, j\rangle = 0$.

De $\langle i, j\rangle = 0$, on déduit $ji = -ij$ et $(ij)^2 = ijij = -i^2j^2 = -1$. Ainsi on a $ij \in \mathrm{V}$ et $ij \neq 0$. On a de plus $i(ij) = i^2j = -j = (ji)i = -(ij)i$ donc $\langle ij, i\rangle = 0$ et de même $\langle ij, j\rangle = 0$.

La famille (i, j, ij) est donc libre. Supposons que $\dim \mathrm{V} > 3$. On peut trouver $k \in (\mathrm{Vect}(i, j, ij))^\perp$, non nul. Il vérifie $ik = -ki$, $jk = -kj$ et $ijk = -kij$. On en déduit

$$ijk = -ikj = kij = -ijk$$

et donc $ijk = 0$, ce qui impossible car A ne possède pas de diviseur de 0. Ainsi V est de dimension 3, engendré par (i, j, ij). Comme $\mathrm{A} = \mathbb{R} \oplus \mathrm{V}$, on en déduit que $\mathrm{A} = \mathrm{Vect}(1, i, j, ij)$. ◁

Dans le cas où A *est de dimension supérieure ou égale à* 3, *on obtient une algèbre de dimension* 4, *de base* $(1, i, j, k)$ *où* i, j, k *vérifient* $i^2 = j^2 = k^2 = ijk = -1$. *Les éléments* i, j, k *anticommutent. Cette algèbre est un corps (puisque tout élément non nul est inversible), non commutatif, appelé corps des quaternions ; il est noté* $\mathbb{H}$.

Il résulte de l'exercice que les seules $\mathbb{R}$*-algèbres de dimension finie et sans diviseur de* 0 *sont* $\mathbb{R}$, $\mathbb{C}$ *et* $\mathbb{H}$. *Ce théorème a été démontré par le mathématicien allemand Georg Frobenius (1878) et indépendamment par le mathématicien et philosophe américain Charles Pierce (1880).*

6.33. Sous-algèbres de dimension finie de $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$

Déterminer les sous-algèbres de dimension finie de $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$.
(École normale supérieure)

▷ **Solution.**

Soit $f \in \mathcal{C}^0(\mathbb{R}, \mathbb{R})$, non constante. Alors la famille $(f^n)_{n\geqslant 0}$ est libre. En effet, une relation de liaison $\lambda_0 + \lambda_1 f + \cdots + \lambda_p f^p = 0$ implique que le polynôme $P = \lambda_0 + \lambda_1 X + \cdots + \lambda_p X^p$ s'annule sur l'image de f. Celle-ci étant un intervalle (d'après le théorème des valeurs intermédiaires), non réduit à un point par hypothèse, elle est de cardinal infini et $P = 0$. Il en résulte que la seule sous-algèbre de dimension finie est l'algèbre des fonctions constantes, qui est de dimension 1. ◁

Si A *est une* K*-algèbre, une dérivation de* A *est un endomorphisme* D *qui vérifie* $D(ab) = D(a)b + aD(b)$ *pour tout* $(a, b) \in A^2$. *Cette notion algébrique, directement inspirée de la dérivation analytique usuelle, a servi de thème à plusieurs exercices posés à l'ENS de Lyon.*

6.34. Racine carrée de la dérivation

Soit $E = \mathcal{C}^\infty(\mathbb{R}, \mathbb{C})$ et $D : E \to E$ l'opérateur de dérivation. Existe-t-il T dans $\mathcal{L}(E)$ tel que $T \circ T = D$?

(**École normale supérieure**)

▷ **Solution.**

Supposons que T existe. Considérons $\operatorname{Ker} D^2$, qui est un plan vectoriel isomorphe à $\mathbb{C}_1[X]$. Il est évidemment stable par D, mais aussi par T, car les endomorphismes T et D commutent, puisque $T \circ D = T^3 = D \circ T$. Notons d et t les restrictions de D et T à $\operatorname{Ker} D^2$. On a $t^2 = d$ et $d^2 = 0$. On en déduit que $t^4 = 0$. Ainsi t est nilpotent. Son indice de nilpotence est inférieur ou égal à la dimension de $\operatorname{Ker} D^2$, qui est 2 (voir l'exercice 6.8 et la remarque qui le suit), donc $d = t^2 = 0$. Mais c'est absurde, car $d \neq 0$.

Conclusion. *L'opérateur* D *n'a pas de racine carrée.* ◁

6.35. Φ-dérivation (1)

Soit A une $\mathbb{R}$-algèbre et $\Phi : A \to \mathbb{R}$ un morphisme d'algèbres. On appelle Φ-dérivation de A toute forme linéaire δ sur A qui vérifie, pour tout $(a, b) \in A^2$, $\delta(ab) = \Phi(a)\delta(b) + \delta(a)\Phi(b)$.

Déterminer tous les morphismes d'algèbre $\Phi : \mathbb{R}[X] \to \mathbb{R}$, puis toutes les Φ-dérivations de $\mathbb{R}[X]$.

(**École normale supérieure**)

▷ **Solution.**

• Soit Φ un morphisme de $\mathbb{R}[X]$ dans $\mathbb{R}$. Il est clair que Φ va être

déterminé par l'image de X. En effet, posons $\Phi(X) = \alpha$. On alors, pour tout $n \in \mathbb{N}$, $\Phi(X^n) = \alpha^n$, puis par linéarité de Φ, pour tout $P \in \mathbb{R}[X]$, $\Phi(P) = P(\alpha)$. Réciproquement, il est clair que, pour tout réel α, l'application $\Phi : P \longmapsto P(\alpha)$ est un morphisme d'algèbres de $\mathbb{R}[X]$ dans $\mathbb{R}$.

• Si Φ est le morphisme $P \longmapsto P(\alpha)$, on peut remarquer que l'application $P \longmapsto P'(\alpha)$ est alors une Φ-dérivation. Cherchons les autres. Soit D une forme linéaire sur $\mathbb{R}[X]$ vérifiant, pour tout $(A, B) \in \mathbb{R}[X]^2$, $D(AB) = \Phi(A)D(B) + \Phi(B)D(A)$. On a $D(1) = D(1 \times 1) = 2D(1)$ et donc $D(1) = 0$. Posons $D(X) = k \in \mathbb{R}$. On obtient alors successivement

$$D(X^2) = 2\Phi(X)D(X) = 2k\alpha,$$
$$D(X^3) = \Phi(X)D(X^2) + \Phi(X^2)D(X) = 2k\alpha^2 + k\alpha^2 = 3k\alpha^2.$$

On montre par récurrence sur n que $D(X^n) = kn\alpha^{n-1}$ pour tout $n \in \mathbb{N}^*$. C'est vrai pour $1 \leqslant n \leqslant 3$ et si $D(X^n) = kn\alpha^{n-1}$ alors

$$D(X^{n+1}) = \Phi(X)D(X^n) + \Phi(X^n)D(X) = \alpha(kn\alpha^{n-1}) + \alpha^n(k) = k(n+1)\alpha^n.$$

On note que $n\alpha^{n-1}$ est la dérivée en a de X^n. Par linéarité de D, on en déduit que, pour tout polynôme P, on a

$$D(P) = kP'(\alpha).$$

Réciproquement, si D est la forme linéaire définie par $D(P) = kP'(\alpha)$, on a, pour $(A, B) \in \mathbb{R}[X]^2$,

$$\begin{aligned} D(AB) &= k(AB)'(\alpha) = kA(\alpha)B'(\alpha) + kA'(\alpha)B(\alpha) \\ &= \Phi(A)D(B) + \Phi(B)D(A). \end{aligned}$$ ◁

6.36. Φ-dérivation (2)

Dans cet exercice, A désigne l'algèbre des suites réelles (a_n) telles que la série entière $\sum a_n x^n$ ait un rayon de convergence supérieur ou égal à $R > 0$ fixé.

1. Quels sont les morphismes d'algèbres de A dans $\mathbb{R}$?

2. Soit Φ un tel morphisme. Quelles sont les Φ-dérivations de l'algèbre A ?

(École normale supérieure)

▷ **Solution.**

1. Il résulte du cours d'analyse que A est bien une algèbre, pour l'addition usuelle des suites et le produit de convolution. Par ailleurs, l'application qui à une suite (a_n) de A associe l'application $f :]-R, R[\to \mathbb{R}$

définie par $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ est un morphisme injectif d'algèbres, permettant d'identifier A à une sous-algèbre de $\mathcal{F}(]-R, R[, \mathbb{R})$, l'algèbre des fonctions développables en série entière en 0 avec un rayon de convergence supérieur ou égal à R. Nous ferons cette identification dans la suite en regardant les éléments de A comme des fonctions définies sur $]-R, R[$.

Soit Φ un morphisme d'algèbres de A dans $\mathbb{R}$. L'ensemble $\mathcal{P}$ des fonctions polynômes de $]-R, R[$ dans $\mathbb{R}$ forme une sous-algèbre de A, isomorphe à $\mathbb{R}[X]$. La restriction de Φ à $\mathcal{P}$ est un morphisme d'algèbres de $\mathcal{P}$ dans $\mathbb{R}$. En reprenant la démonstration de l'exercice précédent, on montre qu'il existe $\alpha \in \mathbb{R}$ tel que $\Phi(P) = P(\alpha)$ pour toute fonction polynôme P.

Lorsque a est un réel de $]-R, R[$, il est clair que l'application $\Phi_a : A \to \mathbb{R}$ qui à $f \in A$ associe $f(a)$ est un morphisme d'algèbres de A dans $\mathbb{R}$, que nous appellerons morphisme d'évaluation en a. Il semble assez naturel de penser que le réel α défini ci-dessus appartient nécessairement à l'intervalle $]R, R[$ et que Φ n'est rien d'autre que le morphisme d'évaluation en α.

• Commençons par montrer que $\alpha = \Phi(\mathrm{Id}) \in]-R, R[$. Raisonnons par l'absurde et supposons que $|\alpha| \geqslant R$. Dans ce cas, la fonction f définie par $f(x) = \sum_{n=0}^{+\infty} \frac{x^n}{\alpha^n}$ appartient à A, puisque la série entière $\sum \frac{x^n}{\alpha^n}$ est de rayon $|\alpha| \geqslant R$. Or, nous savons que, pour $|x| < \alpha$, $f(x) = \dfrac{1}{1 - \dfrac{x}{\alpha}}$, ce qui s'écrit encore $\left(1 - \frac{1}{\alpha}\,\mathrm{Id}\right) f = 1$. Si on applique le morphisme Φ à cette égalité, il vient $\left(1 - \frac{1}{\alpha}\Phi(\mathrm{Id})\right) \Phi(f) = (1 - 1)\Phi(f) = 1$, soit $1 = 0$, ce qui est clairement absurde. Ainsi, le réel α appartient à $]-R, R[$.

• Soit $g \in A$. On va montrer que $\Phi(g) = g(\alpha)$. Cela revient à prouver que $\Phi(g - g(\alpha)) = 0$. Soit la fonction h définie par $h(x) = \dfrac{g(x) - g(\alpha)}{x - \alpha}$, prolongée par continuité en α par $h(\alpha) = g'(\alpha)$. Il suffit de montrer que $h \in A$. En effet, dans ce cas, on pourra écrire $g - g(\alpha) = (\mathrm{Id} - \alpha)h$ et en appliquant Φ, il viendra $\Phi(g) - g(\alpha) = (\Phi(\mathrm{Id}) - \alpha)\Phi(h) = 0$, c'est-à-dire $\Phi(g) = g(\alpha)$.

Si $\alpha = 0$, il est clair que $h \in A$: en posant $g(x) = \sum_{n=0}^{+\infty} a_n x^n$, on a simplement $h(x) = \sum_{n=1}^{+\infty} a_n x^{n-1}$.

Supposons α non nul. Puisque h est continue en α, la relation qui définit h équivaut à $h(x)(x - \alpha) = g(x) - g(\alpha)$, pour tout $x \in]-R, R[$. Considérons la suite (a_n) telle que $g(x) - g(\alpha) = \sum_{n=0}^{+\infty} a_n x^n$ et cherchons

une suite (b_n) telle que $h(x) = \sum_{n=0}^{+\infty} b_n x^n$. La relation qui lie g et h équivaut à : $-\alpha b_0 = a_0$ et pour tout $n \in \mathbb{N}$, $b_{n-1} - \alpha b_n = a_n$. On en déduit facilement par récurrence que, pour tout entier naturel n,

$$b_n = -\sum_{k=0}^{n} a_k \alpha^{k-n-1} = \sum_{k=n+1}^{+\infty} a_k \alpha^{k-n-1},$$

puisque $\sum_{k=0}^{+\infty} a_k \alpha^k = g(\alpha) - g(\alpha) = 0$. Si $r \in]|\alpha|, \mathrm{R}[$, la suite $(a_n r^n)$ est bornée. Considérons $\mathrm{M} > 0$ tel que $|a_n r^n| \leqslant \mathrm{M}$ pour tout n. On a alors, pour tout $n \in \mathbb{N}$,

$$|b_n r^n| = \left| \sum_{k=n+1}^{+\infty} a_k r^n \alpha^{k-n-1} \right| \leqslant \sum_{k=n+1}^{+\infty} \mathrm{M} r^{n-k} |\alpha|^{k-n-1} = \frac{\mathrm{M}}{r} \sum_{i=0}^{+\infty} \left(\frac{|\alpha|}{r} \right)^i$$
$$= \frac{\mathrm{M}}{r - |\alpha|}.$$

Comme la suite $(b_n r^n)$ est bornée, le rayon de convergence de la série entière $\sum b_n x^n$ est supérieur à r. Puisque cela vaut pour tout $r \in]|\alpha|, \mathrm{R}[$, le rayon de convergence est bien supérieur ou égal à R. On a donc, pour tout $x \in]-\mathrm{R}, \mathrm{R}[$, $h(x) = \sum_{n=0}^{+\infty} b_n x^n$: la fonction h appartient à A.

Conclusion. Les seuls morphismes d'algèbres de A dans $\mathbb{R}$ sont les morphismes d'évaluation en un point de l'intervalle $]-\mathrm{R}, \mathrm{R}[$.

2. Soit $\alpha \in]-\mathrm{R}, \mathrm{R}[$ et soit D une Φ_α-dérivation de A. On sait d'après l'exercice précédent qu'il existe $k \in \mathbb{R}$ tel que, pour toute fonction polynôme P, $\mathrm{D}(\mathrm{P}) = k\mathrm{P}'(\alpha)$. Soit $g \in \mathrm{A}$. Comme ci-dessus, on introduit la fonction $h \in \mathrm{A}$ telle que $g - g(\alpha) = (\mathrm{Id} - \alpha)h$ et on applique D à cette égalité. Il vient

$$\mathrm{D}(g) = \mathrm{D}(\mathrm{Id} - \alpha)\Phi_\alpha(h) + \Phi_\alpha(\mathrm{Id} - \alpha)\mathrm{D}(h) = k\Phi_\alpha(h) = kh(\alpha) = kg'(\alpha),$$

car, $\mathrm{Id} - \alpha$ étant une fonction polynôme, $\mathrm{D}(\mathrm{Id} - \alpha) = k$.

Conclusion. Comme dans le cas des polynômes, les seules Φ_α-dérivations de A sont les applications qui à $f \in \mathrm{A}$ associe $kf'(\alpha)$. ◁

6.37. Φ-dérivation (3)

Soit $\mathrm{E} = \mathcal{C}^0(\mathbb{R}, \mathbb{R})$. Déterminer les formes linéaires $\mathrm{D} : \mathrm{E} \to \mathbb{R}$ vérifiant, pour tout $(f, g) \in \mathrm{E}^2$,

$$\mathrm{D}(fg) = f(0)\mathrm{D}(g) + g(0)\mathrm{D}(f).$$

(École normale supérieure)

▷ **Solution.**

Si on note 1 la fonction constante $x \longmapsto 1$, on a $D(1) = D(1)+D(1)$ et donc $D(1) = 0$. On en déduit que D est nulle pour toutes les applications constantes. Si $f \in E$, on peut écrire $f = (f - f(0)) + f(0)$. Par linéarité, on obtient $D(f) = D(f - f(0))$ et on est donc ramené à étudier la restriction de D aux applications nulles en 0. Il est clair que si $f \in E$ est une application qui vérifie $f(0) = 0$, alors on a $D(f^2) = 0$.

Soit maintenant $f \in E$ telle que $f(0) = 0$. Si f est positive, on peut écrire $f = g^2$ où $g = \sqrt{f} \in E$ vérifie $g(0) = 0$. D'après ce qui précède, on a $D(f) = 0$. Dans le cas général, il est possible d'écrire f comme différence de deux fonctions continues, positives et nulles en 0. Il suffit de prendre

$$g(x) = \max(f(x), 0) \text{ et } h(x) = \max(-f(x), 0).$$

On a $f = g - h$, $g \geqslant 0$, $h \geqslant 0$ et $g(0) = h(0) = 0$. Par linéarité de D, on obtient $D(f) = D(g) - D(h) = 0$.

Conclusion. Le seule forme linéaire D qui convienne est donc la forme nulle. ◁

On termine cette série d'exercices sur les algèbres avec l'énoncé suivant.

6.38. Étude d'une algèbre

Soit $E = C^\infty(\mathbb{R}, \mathbb{R})$ et A l'ensemble des $\varphi \in \mathcal{L}(E)$ tels qu'il existe une suite $(A_i)_{i\in\mathbb{N}}$ à support fini d'éléments de $\mathbb{R}[X]$ vérifiant, pour tout $f \in E$,

$$\varphi(f) = \sum_i A_i f^{(i)},$$

où $f^{(i)}$ désigne la dérivée i-ième de f.

1. Montrer que $\varphi = 0$ équivaut à $A_i = 0$ pour tout $i \in \mathbb{N}$.

2. On admet que A est une algèbre. En déterminer les éléments inversibles.

3. Existe-t-il des morphismes d'algèbres de A dans $\mathbb{R}$?

(**École normale supérieure**)

▷ **Solution.**

1. Il est clair que si les A_i sont tous nuls, φ est nul. Réciproquement, si $\varphi = 0$, on montre par récurrence sur i que $A_i = 0$, pour tout $i \in \mathbb{N}$.

- En prenant $f = 1$, on obtient $\varphi(f) = A_0 = 0$.
- Si on suppose que $A_i = 0$, pour $0 \leqslant i \leqslant n - 1$, on obtient en

prenant $f : x \longmapsto x^n$, $\varphi(f) = \sum_{i \geqslant n} \mathrm{A}_i f^{(i)} = n!\,\mathrm{A}_n$, d'où l'on tire $\mathrm{A}_n = 0$. La propriété est établie.

On en déduit qu'un élément de A s'écrit de manière unique sous la forme $\varphi(f) = \sum_i \mathrm{A}_i f^{(i)}$.

2. Soit $\varphi \in \mathrm{A}$ défini par $\varphi(f) = \sum_i \mathrm{A}_i f^{(i)}$, pour tout $f \in \mathrm{E}$. Supposons que φ soit inversible et que son inverse ψ s'écrive $\psi(f) = \sum_j \mathrm{B}_j f^{(j)}$. On obtient alors $\psi \circ \varphi = \mathrm{Id}_\mathrm{E}$ c'est-à-dire, pour tout $f \in \mathrm{E}$,

$$\begin{aligned} f = (\psi \circ \varphi)(f) &= \sum_j \mathrm{B}_j \left(\sum_i \mathrm{A}_i f^{(i)} \right)^{(j)} \\ &= \sum_{i,j} \mathrm{B}_j \left(\sum_{0 \leqslant k \leqslant j} \mathrm{C}_j^k \mathrm{A}_i^{(j-k)} f^{(i+k)} \right) \\ &= \sum_p \left(\sum_{\substack{0 \leqslant i \leqslant p \\ i+j \geqslant p}} \mathrm{C}_j^{p-i} \mathrm{A}_i^{(i+j-p)} \mathrm{B}_j \right) f^{(p)}. \end{aligned}$$

en vertu de la formule de Leibniz.

Ni φ, ni ψ ne sont nuls. Soit i_0 (resp. j_0) le plus grand indice i (resp. j) pour lequel $\mathrm{A}_i \neq 0$ (resp. $\mathrm{B}_j \neq 0$). Considérons le coefficient de $f^{(i_0+j_0)}$ dans la somme précédente. Il est égal à $\mathrm{A}_{i_0}\mathrm{B}_{j_0}$. En effet, si $i+j \geqslant i_0+j_0$, on a, si $i > i_0$, $\mathrm{A}_i = 0$ et, si $i < i_0$, $j > j_0$ et donc $\mathrm{B}_j = 0$. Si $i_0 + j_0 \neq 0$, le coefficient de $f^{(i_0+j_0)}$ n'étant pas nul, on ne peut pas avoir, pour tout $f \in \mathrm{E}$,

$$f = \sum_p \left(\sum_{\substack{0 \leqslant i \leqslant p \\ i+j \geqslant p}} \mathrm{C}_j^{p-i} \mathrm{A}_i^{(i+j-p)} \mathrm{B}_j \right) f^{(p)},$$

d'après l'unicité de l'écriture démontrée à la première question. On a donc nécessairement $i_0 = j_0 = 0$. On obtient alors $f = \mathrm{A}_0\mathrm{B}_0 f$, pour tout $f \in \mathrm{E}$, ce qui n'est vérifié que si $\mathrm{A}_0\mathrm{B}_0 = 1$. Ceci nécessite que A_0 soit un polynôme constant non nul. Alors $\varphi = \mathrm{A}_0\,\mathrm{Id}_\mathrm{E}$, avec $\mathrm{A}_0 \in \mathbb{R}^*$.

Réciproquement, s'il existe $\lambda \in \mathbb{R}^*$ tel que $\varphi = \lambda\,\mathrm{Id}_\mathrm{E}$, alors φ est évidemment inversible, d'inverse $\psi = \dfrac{1}{\lambda}\,\mathrm{Id}_\mathrm{E}$.

Conclusion. Les éléments inversibles de A sont les endomorphismes de la forme $\lambda\,\mathrm{Id}_\mathrm{E}$, avec $\lambda \in \mathbb{R}^*$.

3. Supposons qu'il existe un morphisme d'algèbres F de A dans $\mathbb{R}$. Notons D et φ_1 les éléments de A définis, pour tout $f \in \mathrm{E}$, par $\mathrm{D}(f) = f'$

et $\varphi_1(f) = \mathrm{Id}_{\mathbb{R}} f : x \longmapsto xf(x)$. On a alors, pour tout $f \in \mathrm{E}$,

$$\mathrm{D} \circ \varphi_1(f) = \mathrm{D}(\mathrm{Id}_{\mathbb{R}} f) = f + \mathrm{Id}_{\mathbb{R}} f' = f + \varphi_1(f'),$$

et donc $\mathrm{D} \circ \varphi_1 = \mathrm{Id}_{\mathrm{E}} + \varphi_1 \circ \mathrm{D}$. On en déduit que

$$\mathrm{F}(\mathrm{Id}_{\mathrm{E}}) = \mathrm{F}(\mathrm{D} \circ \varphi_1) - \mathrm{F}(\varphi_1 \circ \mathrm{D}) = \mathrm{F}(\mathrm{D})\mathrm{F}(\varphi_1) - \mathrm{F}(\varphi_1)\mathrm{F}(\mathrm{D}) = 0.$$

Or par définition un morphisme d'algèbre doit envoyer l'élément unité de A sur 1.

Conclusion. Il n'y a pas de morphisme d'algèbres de A dans $\mathbb{R}$. $\lhd$

Chapitre 7

Matrices

Les systèmes de n équations linéaires à n inconnues sont étudiés dès le milieu du XVIII^{e} *siècle. Cramer décrit explicitement les formules donnant la solution sous forme du quotient de deux expressions qui porteront après Cauchy le nom de déterminants. Celui-ci donnera la formule générale définissant le produit de deux déterminants. La notion d'application linéaire, sous le nom de «substitution linéaire», apparaît au début du* XIX^{e} *siècle dans les travaux de Lagrange. puis de Gauss sur les formes quadratiques à coefficients entiers. Ce dernier, pour la première fois, représente la substitution linéaire*

$$(x, y, z) \longmapsto (ax + by + cz, a'x + b'y + c'z, a''x + b''y + c''z)$$

par un tableau $\begin{matrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{matrix}$ *. Il étudie explicitement ce qui se passe quand on compose de telles transformations linéaires, autrement dit, il donne dans ce cas la règle de multiplication des matrices. Il note que ce résultat s'étendrait à un nombre quelconque de variables. Dans ses travaux de théorie des nombres, Eisenstein élargit le symbolisme de Gauss et introduit la notation* $\frac{1}{\mathrm{S}}$ *quand* S *a un déterminant non nul. Le terme de matrice est inventé par Sylvester vers 1850. Cayley, qui définit un espace vectoriel de dimension n comme un système de n nombres réels ou complexes (vers 1845), écrit en 1858 le mémoire que l'on considère comme ayant créé la théorie des matrices : il y définit à la fois l'addition et la multiplication des matrices carrées et rectangulaires. Pour lui, une matrice est une façon abrégée de noter une substitution linéaire.*

L'étude plus approfondie des matrices va se faire à travers les formes quadratiques. Depuis le XVII^{e} *siècle, on recherche des transformations linéaires ramenant celles-ci à des formes simples. On est conduit à considérer des relations matricielles de la forme* $\mathrm{A}' = {}^{t}\mathrm{UAV}$. *À partir de 1850, une série de travaux va viser à obtenir des invariants dans le passage de* A *à* A$'$. *Ces résultats seront résumés et approfondis dans les mémoires de Frobenius, vers 1870. Il établit les formules de changement de bases* $\mathrm{A}' = \mathrm{U}^{-1}\mathrm{AV}$, *montre que toute matrice de rang r est équivalente à* $\begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$. *Il élucide de manière définitive le problème de la résolution des systèmes d'équations linéaires.*

Rappelons maintenant l'essentiel du cours. Soit E, F *deux* K*-espaces vectoriels de dimension finie non nulle, respectivement* p *et* n*. Le choix d'une base* $\mathcal{B}$ *de* E *et d'une base* $\mathcal{C}$ *de* F *permet de ramener l'étude d'une application linéaire* u *de* E *dans* F *à celle de sa matrice* $\mathrm{A} \in \mathcal{M}_{n,p}(\mathrm{K})$*, relativement aux bases* $\mathcal{B}$ *et* $\mathcal{C}$*. En effet,* u *s'identifie alors simplement à l'application linéaire de* K^p *dans* K^n *canoniquement associée à* A*. que l'on note souvent encore* A*. On dispose alors d'un outil analytique pour l'étude de* u *permettant le calcul du rang, et pour des endomorphismes, du déterminant, du polynôme caractéristique, du spectre, etc.*

Le fait que l'on puisse choisir librement les bases $\mathcal{B}$ *et* $\mathcal{C}$ *conduit à introduire la notion de matrices équivalentes : deux matrices* A *et* B *de* $\mathcal{M}_{n,p}(\mathrm{K})$ *sont équivalentes si elles représentent la même application linéaire dans des bases différentes. Pour cela, il faut et suffit qu'il existe* $(\mathrm{Q}, \mathrm{P}) \in \mathrm{GL}_n(\mathrm{K}) \times \mathrm{GL}_p(\mathrm{K})$ *tel que* $\mathrm{B} = \mathrm{Q}^{-1}\mathrm{AP}$*. C'est une relation d'équivalence, dont les classes sont paramétrées par le rang.*

Dans l'étude des endomorphismes de E*, prendre la même base à «l'arrivée» et au «départ», permet de faire correspondre le produit matriciel et la composition. Le changement de base se traduit alors par la notion de matrices semblables, relation d'équivalence bien plus fine que la relation précédente. On renvoie le lecteur au chapitre sur la réduction, dans le tome 2 d'algèbre, pour l'étude de cette notion. Il y trouvera de même un chapitre consacré spécifiquement au groupe linéaire.*

Commençons par une série d'exercices consacrés à la notion de rang. Le premier fournit un exemple de problème abstrait que l'on résout par une traduction matricielle. Nous en avons déjà donné une solution sans faire appel aux matrices dans l'exercice 6.5.

7.1. Condition pour que $\operatorname{rg} g \leqslant \operatorname{rg} f$

Soit E et F deux K-espaces vectoriels non nuls de dimension finie, f et g dans $\mathcal{L}(\mathrm{E}, \mathrm{F})$. Montrer que $\operatorname{rg} g \leqslant \operatorname{rg} f$ si et seulement s'il existe $h \in \mathrm{GL}(\mathrm{F})$ et $k \in \mathcal{L}(\mathrm{E})$ tels que $h \circ g = f \circ k$.

(École polytechnique)

▷ **Solution.**

L'exercice est traité matriciellement. On note n la dimension de F, p celle de E, r le rang de f. On fixe des bases quelconques de E et F et on considère la matrice $\mathrm{A} \in \mathcal{M}_{n,p}(\mathrm{K})$ de f. En écrivant l'égalité $h \circ g = f \circ k$ sous la forme $g = h^{-1} \circ f \circ k$ et en appelant P et Q les matrices de h^{-1} et k, on voit que le problème posé consiste à montrer que l'ensemble

$$\Omega_{\mathrm{A}} = \{\mathrm{PAM},\ \mathrm{P} \in \mathrm{GL}_n(\mathrm{K}), \mathrm{M} \in \mathcal{M}_p(\mathrm{K})\}$$

est l'ensemble des matrices de rang inférieur ou égal à $r = \operatorname{rg} \mathrm{A}$.

On observe que, pour toutes matrices $P \in GL_n(K)$ et $M \in \mathcal{M}_p(K)$, on a $\operatorname{rg}(PAM) = \operatorname{rg}(AM) \leqslant \operatorname{rg}(A)$, car $\operatorname{Im} AM \subset \operatorname{Im} A$, ce qui établit une inclusion.

On remarque que si $B \in \Omega_A$, toute matrice équivalente à B est aussi dans Ω_A. On note, pour $0 \leqslant k \leqslant \min(n,m)$, $J_k = (a_{ij})_{\substack{1 \leqslant i \leqslant n \\ 1 \leqslant j \leqslant p}} \in \mathcal{M}_{n,p}(K)$ la matrice définie par $a_{11} = a_{22} = \cdots = a_{kk} = 1$ et $a_{ij} = 0$ sinon. Elle est de rang k. Comme deux matrices de $\mathcal{M}_{n,p}(K)$ qui ont le même rang sont équivalentes, il nous suffit de prouver que pour tout $s \leqslant r$, $J_s \in \Omega_A$.

On sait que $J_r \in \Omega_A$, puisque $A \in \Omega_A$. Il suffit alors de constater que pour $s \leqslant r$, $J_s = J_r M$, où M est la matrice $\begin{pmatrix} I_s & 0 \\ 0 & 0 \end{pmatrix}$ de $\mathcal{M}_p(K)$, pour conclure. ◁

L'exercice suivant sera à rapprocher des importants lemmes de factorisation de l'exercice 6.4.

7.2. Factorisation bilatère

Soit A, B, C dans $\mathcal{M}_n(K)$. Donner une condition nécessaire et suffisante pour qu'il existe X dans $\mathcal{M}_n(K)$ telle que $C = AXB$.
(**École normale supérieure**)

▷ **Solution.**

Une condition nécessaire se voit facilement : il faut que $\operatorname{Ker} B \subset \operatorname{Ker} C$ et que $\operatorname{Im} C \subset \operatorname{Im} A$ (ce qui impose en particulier que $\operatorname{rg} C \leqslant \operatorname{rg} A$ et $\operatorname{rg} C \leqslant \operatorname{rg} B$). On va voir en fait qu'elle est suffisante.

Dans ce qui suit on note φ_M l'application linéaire canoniquement associée à une matrice M. On suppose donc $\operatorname{Ker} B \subset \operatorname{Ker} C$ et $\operatorname{Im} C \subset \operatorname{Im} A$. Choisissons H un supplémentaire de $\operatorname{Ker} B$ dans K^n. Les applications φ_C et φ_{AXB} coïncident sur $\operatorname{Ker} B$ quel que soit le choix de X. Il suffit de faire en sorte qu'elles coïncident aussi sur H. L'application φ_B établit un isomorphisme entre H et l'espace $\varphi_B(H) = \operatorname{Im} B$, isomorphisme que nous noterons u_B. Il suffit donc de trouver X telle que si f désigne la restriction de φ_X à $\operatorname{Im} B$ alors $\varphi_A \circ f = \varphi_C \circ u_B^{-1}$. Donnons-nous un supplémentaire K de $\operatorname{Ker} A$ de sorte que φ_A induise un isomorphisme u_A entre K et $\operatorname{Im} A$. Par hypothèse, $\operatorname{Im} C \subset \operatorname{Im} A$. Donc $f = u_A^{-1} \circ \varphi_C \circ u_B^{-1}$ est bien définie et convient. Il existe évidemment des matrices X telles que $\varphi_X|_{\operatorname{Im} B} = f$. ◁

L'exercice suivant montre qu'une application multiplicative de $\mathcal{M}_n(K)$ dans K, non constante, permet, comme le déterminant, de caractériser l'inversibilité.

7.3. Fonctions multiplicatives et inversibilité

Soit f une application non constante de $\mathcal{M}_n(\mathrm{K})$ dans K, telle que, pour toutes matrices A et B, on ait $f(\mathrm{AB}) = f(\mathrm{A})f(\mathrm{B})$.

Montrer que $f(\mathrm{A}) = 0$ si, et seulement si, A n'est pas inversible.

(École polytechnique)

▷ **Solution.**

• On observe d'abord que $f(0) = f(0)f(0)$ et donc que $f(0) \in \{0, 1\}$. Si $f(0) = 1$, on a $f(\mathrm{A}) = f(\mathrm{A})f(0) = f(0) = 1$, pour toute matrice A, et l'application f est constante, ce qui est contraire à l'hypothèse. On a donc $f(0) = 0$.

On obtient, de même, $f(\mathrm{I}_n) = f(\mathrm{I}_n)f(\mathrm{I}_n)$ et donc $f(\mathrm{I}_n) \in \{0, 1\}$. Si $f(\mathrm{I}_n) = 0$, on a $f(\mathrm{A}) = f(\mathrm{AI}_n) = f(\mathrm{A})f(\mathrm{I}_n) = 0$, pour toute matrice A, et f est constante. On a donc $f(\mathrm{I}_n) = 1$.

• Supposons A inversible. On peut alors écrire

$$f(\mathrm{A})f(\mathrm{A}^{-1}) = f(\mathrm{I}_n) = 1 \text{ et donc } f(\mathrm{A}) \neq 0.$$

• Soit A et B deux matrices équivalentes : $\mathrm{B} = \mathrm{QAP}$ avec (Q et P dans $\mathrm{GL}_n(\mathrm{K})$. Comme $f(\mathrm{B}) = f(\mathrm{Q})f(\mathrm{A})f(\mathrm{P})$, on a, d'après ce qui précède, $f(\mathrm{A}) = 0$ si et seulement si $f(\mathrm{B}) = 0$. Autrement dit, sur une classe d'équivalence, soit f est identiquement nulle, soit elle ne s'annule jamais.

Il suffit donc, pour la réciproque, de montrer que pour tout $r < n$ il existe une matrice A de rang r telle que $f(\mathrm{A}) = 0$. On peut prendre une matrice nilpotente de rang r, par exemple $\mathrm{K}_r = \left(\begin{array}{c|c} 0 & \mathrm{I}_r \\ \hline 0 & 0 \end{array}\right) \in \mathcal{M}_n(\mathrm{K})$. On a alors $(\mathrm{K}_r)^{r+1} = 0$ et donc $f(\mathrm{K}_r)^{r+1} = 0$ et $f(\mathrm{K}_r)$ est bien nul, tout comme $f(\mathrm{A})$. ◁

7.4. Degré et valuation du polynôme $\det(\mathrm{XA} + \mathrm{B})$

Soit A et B dans $\mathcal{M}_n(\mathbb{C})$. On pose $\mathrm{P}(t) = \det(t\mathrm{A} + \mathrm{B})$. Établir que

$$\deg \mathrm{P} \leqslant \operatorname{rg} \mathrm{A} \quad \text{et} \quad \operatorname{val} \mathrm{P} \geqslant n - \operatorname{rg} \mathrm{B}.$$

(École polytechnique)

▷ **Solution.**

• Si le rang de A est r, A est équivalente à la matrice $\mathrm{J}_r = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$. Elle s'écrit donc $\mathrm{A} = \mathrm{QJ}_r\mathrm{R}$ avec Q et R inversibles. On considère la matrice B′ telle que $\mathrm{B} = \mathrm{QB'R}$. On a

$$P(t) = \det(Q(tJ_r + B')R) = \det Q \det R \det(tJ_r + B').$$

En notant $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{C}^n$ et $C_1, \ldots, C_n$ les vecteurs colonnes de B', on obtient

$$P(t) = \det Q \det R \det(te_1 + C_1, \ldots, te_r + C_r, C_{r+1}, \ldots, C_n).$$

En utilisant la multilinéarité du déterminant, on voit que P est un polynôme en t de degré inférieur ou égal à r, le coefficient de t^r étant $\det Q \det R \det(e_1, \ldots, e_r, C_{r+1}, \ldots, C_n)$.

• Notons $Q(t) = \det(A + tB)$ le polynôme obtenu en échangeant A et B. On a d'après le point précédent, $\deg(Q) \leqslant \operatorname{rg} B$. Or, par multilinéarité du déterminant, $P(t) = t^n \det(A + \frac{1}{t}B) = t^n Q(\frac{1}{t})$. Cette expression montre que la valuation de P est égale à $n - \deg Q$. La seconde inégalité en résulte. ◁

On montre dans l'exercice suivant que, pour qu'un endomorphisme de $\mathcal{M}_n(\mathbb{C})$ conserve le rang, il suffit qu'il conserve le caractère inversible.

7.5. Endomorphismes de $\mathcal{M}_n(\mathbb{C})$ stabilisant le groupe linéaire

Soit φ un endomorphisme de $\mathcal{M}_n(\mathbb{C})$ tel que si M appartient à $GL_n(\mathbb{C})$, alors $\varphi(M)$ appartient à $GL_n(\mathbb{C})$.

1. Donner des exemples de tels endomorphismes.

2. Montrer que, pour tout $M \in \mathcal{M}_n(\mathbb{C})$, M appartient à $GL_n(\mathbb{C})$ si, et seulement si, $\varphi(M)$ appartient à $GL_n(\mathbb{C})$. Pour cela, on prouvera que si $\operatorname{rg} M < n$, il existe $P \in GL_n(\mathbb{C})$ tel que pour tout $\lambda \in \mathbb{C}$, $P - \lambda M$ est inversible.

3. Montrer que $\operatorname{rg}\varphi(M) \geqslant \operatorname{rg} M$. Pour cela, on montrera que si $\operatorname{rg} M = r$, il existe $Q \in GL_n(\mathbb{C})$ tel que $Q - \lambda M$ soit non inversible pour r valeurs de λ exactement.

4. Montrer que φ conserve le rang.

(**École normale supérieure**)

▷ **Solution.**

1. Si P et Q sont dans $GL_n(\mathbb{C})$, l'endomorphismes $M \longmapsto QMP$ convient. La transposition convient également. Par suite tous les endomorphismes du type $M \longmapsto Q\,{}^tMP$ vérifient la propriété de l'énoncé.

2. Soit $M \in \mathcal{M}_n(\mathbb{C})$.

• Montrons tout d'abord l'indication. Supposons $\operatorname{rg} M < n$ et M non nulle (si $M = 0$ toute matrice inversible P convient). Géométriquement, on souhaite prouver qu'il existe une droite affine de $\mathcal{M}_n(\mathbb{C})$ dirigée par M

et incluse dans le groupe linéaire. En multipliant par P^{-1}, le problème se ramène à montrer l'existence de $P \in GL_n(\mathbb{C})$ tel que pour tout $\lambda \in \mathbb{C}$, on ait $I_n - \lambda P^{-1}M \in GL_n(\mathbb{C})$. Cette condition équivaut à dire que la seule valeur propre de $P^{-1}M$ est 0, c'est-à-dire que $P^{-1}M$ est nilpotente.

On utilise encore les matrices équivalentes. Comme $\operatorname{rg} M = r < n$, elle peut s'écrire $M = AK_rB$ où $K_r = \begin{pmatrix} 0 & I_r \\ 0 & 0 \end{pmatrix}$ et $(A, B) \in GL_n(\mathbb{C})^2$. Comme la matrice K_r est nilpotente, la matrice $B^{-1}A^{-1}M = B^{-1}K_rB$, qui lui est semblable, est aussi nilpotente. Ainsi $P = AB$ convient.

• On suppose toujours $\operatorname{rg}(M) < n$ et on va montrer maintenant que $\varphi(M)$ est non inversible. La matrice P étant choisie de manière à remplir la condition ci-dessus, on a $\varphi(P - \lambda M) = \varphi(P) - \lambda\varphi(M) \in GL_n(\mathbb{C})$ pour tout λ dans $\mathbb{C}$. Comme P est inversible, $\varphi(P)$ l'est aussi et pour tout $\lambda \in \mathbb{C}$,

$$I_n - \lambda\varphi(M)\left(\varphi(P)\right)^{-1} \in GL_n(\mathbb{C}).$$

Donc 0 est la seule valeur propre de $\varphi(M)\left(\varphi(P)\right)^{-1}$: celle-ci est nilpotente, donc non inversible. Nécessairement, $\varphi(M)$ est non inversible. On a donc bien l'équivalence demandée.

3. L'inégalité $\operatorname{rg}(\varphi(M)) \geqslant \operatorname{rg} M$ est vérifiée si M est inversible. Supposons dans la suite $r = \operatorname{rg} M < n$.

• On commence par démontrer l'indication en suivant une démarche analogue à celle de la question 2. On écrit cette fois $M = AJ_rB$, avec $J_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$ et $(A, B) \in GL_n(\mathbb{C})^2$. Soit $D = \operatorname{Diag}(1, 2, \ldots, n) \in GL_n(\mathbb{C})$. Posons $Q = ADB$. On a $Q - \lambda M = A(D - \lambda J_r)B$. Elle est non inversible si et seulement si $\lambda \in \{1, 2, \ldots, r\}$.

• D'après la question précédente, $\varphi(Q - \lambda M) = \varphi(Q) - \lambda\varphi(M)$ est non inversible pour exactement r valeurs (non nulles) de λ. Il en va de même de $I_n - \lambda\varphi(Q)^{-1}\varphi(M)$. On en déduit que la matrice $\varphi(Q)^{-1}\varphi(M)$ a exactement r valeurs propres non nulles distinctes : elle est donc au moins de rang r. On en déduit que $\varphi(M)$ est au moins de rang r puisqu'elle est de même rang que $\varphi(Q)^{-1}\varphi(M)$.

4. Il résulte de la question précédente que φ est injectif ; c'est donc un automorphisme de l'espace vectoriel $\mathcal{M}_n(\mathbb{C})$. Si $M \in GL_n(\mathbb{C})$, $\varphi^{-1}(M)$ est inversible en vertu de la question 2 puisque $M = \varphi(\varphi^{-1}(M))$. L'endomorphisme φ^{-1} vérifie donc la même hypothèse que φ. D'après la question précédente appliquée à φ^{-1}, il vient

$$\operatorname{rg} M = \operatorname{rg}\varphi^{-1}(\varphi(M)) \geqslant \operatorname{rg}\varphi(M) \text{ et finalement } \operatorname{rg} M = \operatorname{rg}\varphi(M).$$

Conclusion. L'endomorphisme φ de $\mathcal{M}_n(\mathbb{C})$ conserve le rang. ◁

L'exercice suivant aborde la détermination de tous les endomorphismes de l'espace vectoriel $\mathcal{M}_n(\mathbb{C})$ qui conservent le rang. On remarque que ce sont nécessairement des automorphismes. Notons que ce résultat a fait l'objet d'un problème d'écrit posé au concours de l'École polytechnique en 1988.

7.6. Endomorphismes de $\mathcal{M}_n(\mathbb{C})$ conservant le rang

1. Décrire les sous-espaces vectoriels de $\mathcal{M}_n(\mathbb{C})$ formés de matrices de rang au plus 1.

2. Déterminer les automorphismes de l'espace vectoriel $\mathcal{M}_n(\mathbb{C})$ qui conservent le rang.

(École normale supérieure)

▷ **Solution.**

On supposera dans la suite que $n \geqslant 2$, sinon tous les sous-espaces vectoriels de $\mathcal{M}_n(\mathbb{C})$, puis tous les automorphismes de $\mathcal{M}_n(\mathbb{C})$ conviennent.

1. • On commence par décrire les matrices de rang 1. Soit A une matrice de rang 1 de $\mathcal{M}_n(\mathbb{C})$. Les vecteurs colonnes de A sont proportionnels. Si on fait la convention d'identifier $\mathcal{M}_{n,1}(\mathbb{C})$ avec $\mathbb{C}^n$, il existe donc des vecteurs $X \in \mathbb{C}^n \setminus \{0\}$ et $Y = {}^t(y_1, \ldots, y_n) \in \mathbb{C}^n \setminus \{0\}$ tels que les vecteurs colonnes de A soient $(y_1 X, \ldots, y_n X)$. On a donc $A = X\,{}^tY$. Réciproquement, toute matrice qui s'écrit $X\,{}^tY$, où X et Y sont des vecteurs colonnes non nuls est de rang 1, car elle est non nulle et ses colonnes sont toutes colinéaires à X.

L'écriture de A sous la forme $X\,{}^tY$ n'est pas unique. On vérifie aisément que deux matrices $A = X\,{}^tY$ et $A' = X'\,{}^tY'$ de rang 1 écrites sous cette forme, sont colinéaires si et seulement si les deux familles (X, X') et (Y, Y') sont liées et en particulier, on a $A = A'$ si et seulement s'il existe $\lambda \in \mathbb{C}^*$ tel que $X' = \lambda X$ et $Y' = \dfrac{1}{\lambda} Y$.

• On dispose dès maintenant d'un moyen simple pour fabriquer des sous-espaces vectoriels de $\mathcal{M}_n(\mathbb{C})$ formés de matrices de rang au plus 1. Soit $X \in \mathbb{C}^n \setminus \{0\}$ une colonne non nulle. L'application $f_X : \mathbb{C}^n \to \mathcal{M}_n(\mathbb{C})$ qui à Y associe $X\,{}^tY$ est linéaire. L'image d'un sous-espace vectoriel L de $\mathbb{C}^n$ par f_X est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{C})$ formé de matrices de rang au plus 1. Posons $E_{X,L} = f_X(L)$. Comme le noyau de f_X est réduit à 0 d'après ce qui précède, f_X est injective et $E_{X,L}$ a même dimension que L.

De la même manière, on peut fixer une colonne $Y \in \mathbb{C}^n$ non nulle et prendre les images des sous-espaces vectoriels de $\mathbb{C}^n$ par l'application linéaire $g_Y : \mathbb{C}^n \to \mathcal{M}_n(\mathbb{C})$ qui à X associe $X\,{}^tY$. Si L est un sous-

espace de $\mathbb{C}^n$, on notera $\mathrm{E}_{\mathrm{L},\mathrm{Y}}$ le sous-espace $g_{\mathrm{Y}}(\mathrm{L})$. On a également $\dim \mathrm{E}_{\mathrm{L},\mathrm{Y}} = \dim \mathrm{L}$.

• On va montrer que les sous-espaces obtenus dans le point précédent sont les seuls possibles. Soit E un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{C})$ formé de matrices de rang au plus 1. Si $\dim \mathrm{E} = 1$, le résultat est clair. On suppose donc $\dim \mathrm{E} \geqslant 2$. Soit A et A′ deux matrices non colinéaires de E (donc non nulles), que nous écrivons $\mathrm{A} = \mathrm{X}\,{}^t\mathrm{Y}$ et $\mathrm{A}' = \mathrm{X}'\,{}^t\mathrm{Y}'$. On note que

$$\mathrm{rg}(\mathrm{A} + \mathrm{A}') \leqslant 1 \Longrightarrow (\mathrm{X}, \mathrm{X}') \text{ liée ou } (\mathrm{Y}, \mathrm{Y}') \text{ liée.}$$

En effet, supposons que les familles $(\mathrm{X}, \mathrm{X}')$ et $(\mathrm{Y}, \mathrm{Y}')$ sont libres. Si on note $\mathrm{Y} = {}^t(y_1, \ldots, y_n)$ et $\mathrm{Y}' = {}^t(y'_1, \ldots, y'_n)$, les vecteurs-colonnes de $\mathrm{A} + \mathrm{A}'$ ont pour coordonnées $\begin{pmatrix} y_1 \\ y'_1 \end{pmatrix}$, …, $\begin{pmatrix} y_n \\ y'_n \end{pmatrix}$ dans la base $(\mathrm{X}, \mathrm{X}')$. Ils ne sont pas tous colinéaires car la famille $(\mathrm{Y}, \mathrm{Y}')$ est libre et on a $\mathrm{rg}(\mathrm{A} + \mathrm{A}') \geqslant 2$.

⋆ Supposons alors que $(\mathrm{X}, \mathrm{X}')$ est libre. D'après ce qui précède, comme $\mathrm{A} + \mathrm{A}' \in \mathrm{E}$, la famille $(\mathrm{Y}, \mathrm{Y}')$ est liée.

Soit maintenant une autre matrice $\mathrm{A}'' = \mathrm{X}''\,{}^t\mathrm{Y}''$, non nulle, de E. Supposons que la famille $(\mathrm{Y}, \mathrm{Y}'')$ est libre. Alors $(\mathrm{Y}', \mathrm{Y}'')$ est également libre et nécessairement les familles $(\mathrm{X}, \mathrm{X}'')$ et $(\mathrm{X}', \mathrm{X}'')$ sont liées (toujours grâce à l'implication ci-dessus). Mais comme $\mathrm{X}'' \neq 0$, la famille $(\mathrm{X}, \mathrm{X}')$ est liée, ce qui est contraire à notre hypothèse. On a donc Y″ colinéaire à Y. Comme Y″ n'est défini qu'à une constante multiplicative près, on peut supposer $\mathrm{Y}'' = \mathrm{Y}$. Tout élément de E peut donc s'écrire $\mathrm{X}''\,{}^t\mathrm{Y}$, avec $\mathrm{Y} \in \mathbb{C}^n \setminus \{0\}$ fixé. Comme Y est fixé, une telle écriture d'un élément de E est unique. La matrice $\mathrm{X}''\,{}^t\mathrm{Y}$ décrivant un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{C})$, X″ décrit un sous-espace vectoriel L de $\mathbb{C}^n$. Finalement, on a $\mathrm{E} = \mathrm{E}_{\mathrm{L},\mathrm{Y}}$.

⋆ On montre de la même manière que lorsque $(\mathrm{Y}, \mathrm{Y}')$ est libre, il existe un sous-espace vectoriel L de $\mathbb{C}^n$ tel que $\mathrm{E} = \mathrm{E}_{\mathrm{X},\mathrm{L}}$.

Conclusion. Il y a donc deux types de sous-espaces vectoriels de $\mathcal{M}_n(\mathbb{C})$ formés de matrices de rang au plus 1, qui sont les sous-espaces $\mathrm{E}_{\mathrm{X},\mathrm{L}} = \{\mathrm{X}\,{}^t\mathrm{Y}, \mathrm{Y} \in \mathrm{L}\}$ et $\mathrm{E}_{\mathrm{L},\mathrm{Y}} = \{\mathrm{X}\,{}^t\mathrm{Y}, \mathrm{X} \in \mathrm{L}\}$, X et Y étant des vecteurs non nuls de $\mathbb{C}^n$ et L un sous-espace vectoriel de $\mathbb{C}^n$. La dimension d'un tel sous-espace est celle de L ; elle est au plus égale à n.

2. Soit Φ un automorphisme de l'espace vectoriel $\mathcal{M}_n(\mathbb{C})$ conservant le rang. Il transforme tout sous-espace vectoriel de $\mathcal{M}_n(\mathbb{C})$ formé de matrices de rang $\leqslant 1$ en un sous-espace formé de matrices de rang $\leqslant 1$, et de même dimension. Nous nous intéresserons particulièrement aux sous-espaces de dimension n. Un tel sous-espace est de la forme $\mathrm{E}_{\mathrm{X},\mathbb{C}^n}$ ou $\mathrm{E}_{\mathbb{C}^n,\mathrm{Y}}$. Pour alléger les notations, nous poserons désormais $\mathrm{E}_{\mathrm{X},\mathbb{C}^n} = \mathrm{E}^1_{\mathrm{X}}$ et $\mathrm{E}_{\mathbb{C}^n,\mathrm{Y}} = \mathrm{E}^2_{\mathrm{Y}}$.

• Soit X_0 un élément non nul, fixé de $\mathbb{C}^n \setminus \{0\}$. Il existe $X_0' \in \mathbb{C}^n \setminus \{0\}$ tel que $\Phi(E^1_{X_0}) = E^1_{X_0'}$ ou $E^2_{X_0'}$. Nous nous contenterons d'examiner le premier cas, le deuxième pouvant s'en déduire. En effet, considérons l'endomorphisme T de $\mathcal{M}_n(\mathbb{C})$: $A \longmapsto {}^tA$. C'est un automorphisme de $\mathcal{M}_n(\mathbb{C})$ qui conserve le rang et $T \circ \Phi$ également. Mais si $\Phi(E^1_{X_0}) = E^2_{X_0'}$, alors $T \circ \Phi(E^1_{X_0}) = E^1_{X_0'}$ et en remplaçant Φ par $T \circ \Phi$ on est ramené au premier cas. On suppose donc dans la suite qu'il existe $X_0' \in \mathbb{C}^n \setminus \{0\}$ tel que $\Phi(E^1_{X_0}) = E^1_{X_0'}$.

⋆ Si $X \in \mathbb{C}^n \setminus \{0\}$ est colinéaire à X_0, $\Phi(E^1_X) = \Phi(E^1_{X_0}) = E^1_{X_0'}$. Si la famille (X_0, X) est libre, alors il résulte des conditions d'égalité de deux matrices de rang 1 (cf. question 1), que $E^1_{X_0} \cap E^1_X = \{0\}$. On en déduit, par injectivité de Φ que $E^1_{X_0'} \cap \Phi(E^1_X) = \{0\}$. S'il existe $X' \in \mathbb{C}^n \setminus \{0\}$ tel que $\Phi(E^1_X) = E^2_{X'}$ alors $E^1_{X_0'} \cap \Phi(E^1_X)$ est égal à $\mathrm{Vect}(X_0' {}^tX')$; c'est impossible. Nous avons donc démontré que, pour tout $X \in \mathbb{C}^n \setminus \{0\}$, il existe $X' \in \mathbb{C}^n \setminus \{0\}$ tel que $\Phi(E^1_X) = E^1_{X'}$.

⋆ Si on prend $Y \in \mathbb{C}^n \setminus \{0\}$, on a $E^1_{X_0} \cap E^2_Y = \mathrm{Vect}(X_0 {}^tY)$ et $\Phi(E^1_{X_0}) \cap \Phi(E^2_Y) = E^1_{X_0'} \cap \Phi(E^2_Y)$. Si $\Phi(E^2_Y) = E^1_{Y'}$, alors $\Phi(E^1_{X_0}) \cap \Phi(E^2_Y)$ est $\{0\}$ ou $E^1_{X_0'}$, selon que X_0' et Y' sont ou non colinéaires. Ce n'est pas une droite vectorielle ; c'est impossible. Nous avons donc démontré que, pour tout $Y \in \mathbb{C}^n \setminus \{0\}$, il existe $Y' \in \mathbb{C}^n \setminus \{0\}$ tel que $\Phi(E^2_Y) = E^2_{Y'}$.

Le problème est que X' et Y' ne sont pas définis de manière unique.

• Soit X un élément non nul, fixé de $\mathbb{C}^n \setminus \{0\}$. Supposons choisi un vecteur X' tel que $\Phi(E^1_X) = E^1_{X'}$. Pour tout $Y \in \mathbb{R}^n$, $\Phi(X {}^tY)$ appartient à $\Phi(E^1_X) = E^1_{X'}$. Il existe donc $Z \in \mathbb{R}^n$ tel que $\Phi(X {}^tY) = X' {}^tZ$. Le vecteur X' étant fixé, ceci définit un Z unique. Considérons l'application $G_X : Y \in \mathbb{C}^n \longmapsto Z \in \mathbb{C}^n$. Puisque Φ est linéaire, l'application G_X est aussi linéaire. D'autre part, on a

$$G_X(Y) = 0 \Longleftrightarrow \Phi(X {}^tY) = 0 \Longrightarrow X {}^tY = 0 \Longrightarrow Y = 0, \text{ car } X \neq 0.$$

On en déduit que $G_X \in \mathrm{GL}_n(\mathbb{C})$. On définit en particulier G_{X_0} par le choix de X_0' effectué tout au début.

Soit alors $Y \in \mathbb{C}^n \setminus \{0\}$ et Y' tel que $\Phi(E^2_Y) = E^2_{Y'}$. On a, pour tout $X \in \mathbb{C}^n \setminus \{0\}$,

$$X' {}^tG_X(Y) = \Phi(X {}^tY) \in \Phi(E^2_Y) = E^2_{Y'}.$$

On en déduit que, pour tout $X \in \mathbb{C}^n \setminus \{0\}$, $G_X(Y)$ est colinéaire à Y' et donc que $G_X(Y)$ est colinéaire à $G_{X_0}(Y)$. Ce qui peut encore s'écrire : $G_{X_0}^{-1} \circ G_X(Y)$ est colinéaire à Y, ceci pour tout $Y \in \mathbb{C}^n$. C'est un résultat classique (rappelé avant l'exercice 6.6) qui permet alors d'affirmer que $G_{X_0}^{-1} \circ G_X$ est un homothétie : il existe $\lambda_X \in \mathbb{C}^*$ tel que $G_X = \lambda_X G_{X_0}$. Mais l'application G_X est définie à une constante multiplicative près, qui

dépend du choix de X'. En remplaçant X' par un vecteur colinéaire, on peut supposer que $G_X = G_{X_0}$, pour tout $X \in \mathbb{C}^n \setminus \{0\}$.

Pour simplifier on note $G_{X_0} = G$. On obtient donc que, pour tout $X \in \mathbb{C}^n$, il existe $X' \in \mathbb{C}^n$ tel que $\Phi(X\,{}^tY) = X'\,{}^tG(Y)$. Le vecteur X' est maintenant déterminé de manière unique. Comme précédemment pour G, on montre aisément que l'application $F : X \in \mathbb{C}^n \longmapsto X' \in \mathbb{C}^n$ est linéaire et injective.

Soit P et Q les matrices respectives de F et G dans la base canonique. Ce sont des matrices inversibles. On a, pour tout $(X, Y) \in (\mathbb{C}^n)^2$,

$$\Phi(X\,{}^tY) = F(X)\,{}^tG(Y) = PX\,{}^t(QY) = PX\,{}^tY\,{}^tQ.$$

Autrement dit, on obtient $\Phi(A) = PA\,{}^tQ$, pour toute matrice de rang 1. Toute matrice de $\mathcal{M}_n(\mathbb{C})$ étant somme de matrices de rang 1 (par exemple somme de matrices colinéaires aux matrices de base E_{ij}), cette égalité est vraie pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$.

Pour terminer, regardons ce qu'on obtient pour Φ dans le cas où $\Phi(E^1_{X_0})$ est de la forme $E^2_{X'_0}$. En composant avec T on obtient une application du type précédent : il existe donc deux matrices inversibles P et Q telles que, pour tout $A \in \mathcal{M}_n(\mathbb{C})$,

$${}^t(\Phi(A)) = T \circ \Phi(A) = PA\,{}^tQ, \text{ et donc } \Phi(A) = Q\,{}^tA\,{}^tP.$$

Conclusion. Si Φ est un automorphisme de $\mathcal{M}_n(\mathbb{C})$ conservant le rang, alors il existe deux matrices inversibles de $\mathcal{M}_n(\mathbb{C})$ telles que l'on ait soit $\Phi(A) = PAQ$, pour tout $A \in \mathcal{M}_n(\mathbb{C})$, soit $\Phi(A) = P\,{}^tAQ$, pour tout $A \in \mathcal{M}_n(\mathbb{C})$. Il est clair que, réciproquement, les applications ainsi définies sont des automorphismes de $\mathcal{M}_n(\mathbb{C})$ qui conservent le rang. ◁

Notons que la preuve ci-dessus n'utilise en aucune manière le fait que le corps de base est $\mathbb{C}$. Aussi le résultat est-il valide pour un corps quelconque.

La série d'exercices qui suit concerne la trace. On commence par une petite équation matricielle qui fait intervenir la trace.

7.7. Équation matricielle $X + {}^tX = (\mathrm{Tr}\,X)A$

Soit $A \in \mathcal{M}_n(\mathbb{R})$. Résoudre l'équation $X + {}^tX = (\mathrm{Tr}\,X)A$ où l'inconnue X est dans $\mathcal{M}_n(\mathbb{R})$.

(École polytechnique)

▷ **Solution.**

Si X est solution, on obtient en prenant la trace des deux membres de l'équation $2\,\mathrm{Tr}\,X = \mathrm{Tr}\,X\,\mathrm{Tr}\,A$.

• Si $\mathrm{Tr}\,X = 0$, alors ${}^tX = -X$: X est antisymétrique. Réciproquement, toute matrice antisymétrique est solution de l'équation.

• Si $\mathrm{Tr}\,X \neq 0$, nécessairement, $\mathrm{Tr}\,A = 2$ et $A = \frac{1}{\mathrm{Tr}\,X}(X + {}^tX)$ est symétrique. On a alors

$$\left(X - \frac{\mathrm{Tr}\,X}{2}A\right) + {}^t\left(X - \frac{\mathrm{Tr}\,X}{2}A\right) = 0$$

et donc $X - \frac{\mathrm{Tr}\,X}{2}A$ est antisymétrique. Autrement dit, il existe B antisymétrique telle que

$$X = \frac{\mathrm{Tr}\,X}{2}A + B.$$

Réciproquement, si $\lambda \in \mathbb{R}$ et B est antisymétrique, $X = \lambda A + B$ est solution. En effet, on a $\mathrm{Tr}\,X = \lambda\,\mathrm{Tr}\,A = 2\lambda$ et $X + {}^tX = 2\lambda A = (\mathrm{Tr}\,X)A$.

Conclusion. Si $\mathrm{Tr}\,A \neq 2$ ou si A n'est pas symétrique, les seules solutions sont les matrices antisymétriques. Si $\mathrm{Tr}\,A = 2$ et A symétrique, les solutions s'écrivent $\lambda A + B$ avec $\lambda \in \mathbb{R}$ et B antisymétrique quelconques. ◁

En combinant la trace et le produit, on obtient sur $\mathcal{M}_n(\mathrm{K})$ la forme bilinéaire $(X, Y) \longmapsto \mathrm{Tr}(XY)$ qui permet d'établir un isomorphisme canonique entre $\mathcal{M}_n(\mathrm{K})$ et son dual, comme le montre l'exercice suivant.

7.8. Dual de $\mathcal{M}_n(\mathrm{K})$

1. Si $A \in \mathcal{M}_n(\mathrm{K})$, on note f_A la forme linéaire définie, pour tout $X \in \mathcal{M}_n(\mathrm{K})$, par $f_A(X) = \mathrm{Tr}(AX)$. Montrer que l'application f qui à $A \in \mathcal{M}_n(\mathrm{K})$ associe f_A établit un isomorphisme entre $\mathcal{M}_n(\mathrm{K})$ et son dual.

2. Soit $f : \mathcal{M}_n(\mathrm{K}) \longrightarrow \mathrm{K}$ une forme linéaire vérifiant, pour tout (X, Y) dans $\mathcal{M}_n(\mathrm{K})^2$, $f(XY) = f(YX)$. Montrer l'existence de $\lambda \in \mathrm{K}$ tel que, pour tout $X \in \mathcal{M}_n(\mathrm{K})$, $f(X) = \lambda\,\mathrm{Tr}(X)$.

(**École normale supérieure**)

▷ **Solution.**

On notera $(E_{ij})_{1 \leqslant i,j \leqslant n}$ la base canonique de $\mathcal{M}_n(\mathrm{K})$. On rappelle que, pour tout $1 \leqslant i, j, k, l \leqslant n$, on a $E_{ij}E_{kl} = \delta_{jk}E_{il}$.

1. On vérifie aisément que f est linéaire. Pour des raisons de dimension, il suffit donc de prouver qu'elle est injective. Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ telle que $f_A = 0$. On a donc, pour $1 \leqslant i_0, j_0 \leqslant n$,

$$0 = \mathrm{Tr}(\mathrm{A}\mathrm{E}_{i_0 j_0}) = \mathrm{Tr}\Big(\sum_{1\leqslant i,j\leqslant n} a_{ij}\mathrm{E}_{ij}\mathrm{E}_{i_0 j_0}\Big) = \mathrm{Tr}\Big(\sum_{i=1}^{n} a_{ii_0}\mathrm{E}_{ii_0}\mathrm{E}_{i_0 j_0}\Big)$$
$$= \sum_{i=1}^{n} a_{ii_0}\,\mathrm{Tr}\,(\mathrm{E}_{ij_0}) = a_{j_0 i_0}.$$

Donc A est nulle.

2. • On peut donner une solution directe de cette question. On obtient, si $1 \leqslant i, j \leqslant n$ et $i \neq j$,

$$f(\mathrm{E}_{ij}) = f(\mathrm{E}_{ii}\mathrm{E}_{ij}) = f(\mathrm{E}_{ij}\mathrm{E}_{ii}) = f(0) = 0.$$

On a également

$$f(\mathrm{E}_{ii}) = f(\mathrm{E}_{ij}\mathrm{E}_{ji}) = f(\mathrm{E}_{ji}\mathrm{E}_{ij}) = f(\mathrm{E}_{jj}).$$

Si on note λ la valeur commune des $f(\mathrm{E}_{ii})$, on remarque que les formes linéaires f et $\lambda\,\mathrm{Tr}$ coïncident sur la base canonique de $\mathcal{M}_n(\mathrm{K})$. Elles sont donc égales.

• On peut aussi utiliser la première question. Soit $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$ telle que $f = f_{\mathrm{A}}$. On a, pour tout $(\mathrm{X}, \mathrm{Y}) \in \mathcal{M}_n(\mathrm{K})^2$, $\mathrm{Tr}(\mathrm{AXY}) = \mathrm{Tr}(\mathrm{AYX})$. Comme $\mathrm{Tr}(\mathrm{AYX}) = \mathrm{Tr}(\mathrm{XAY})$, on en déduit que $\mathrm{Tr}((\mathrm{AX} - \mathrm{XA})\mathrm{Y}) = 0$. Cela étant valable pour toute matrice Y, on a, d'après la question 1, $\mathrm{AX} = \mathrm{XA}$. Ainsi A commute avec toute matrice X ; il s'agit donc d'une matrice scalaire. Cela résulte du fait (démontré page 269) que le centre de $\mathcal{L}(\mathrm{E})$ est l'ensemble des homothéties. Donnons-en une démonstration directe. Si $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n}$, on a pour $1 \leqslant i, j \leqslant n$,

$$\mathrm{A}\mathrm{E}_{ij} = \sum_{1\leqslant k,l\leqslant n} a_{kl}\mathrm{E}_{kl}\mathrm{E}_{ij} = \sum_{k=1}^{n} a_{ki}\mathrm{E}_{kj}$$
$$= \mathrm{E}_{ij}\mathrm{A} = \sum_{1\leqslant k,l\leqslant n} a_{kl}\mathrm{E}_{ij}\mathrm{E}_{kl} = \sum_{l=1}^{n} a_{jl}\mathrm{E}_{il}.$$

Par unicité de l'écriture, on obtient $a_{ki} = 0$ pour $k \neq i$ et $a_{ii} = a_{jj}$: A est bien une matrice scalaire. On retrouve ainsi que $f = f_{\mathrm{A}}$ est colinéaire à la trace. ◁

Notons que la seconde question fournit une caractérisation de la trace (à un scalaire près) parmi les formes linéaires. Les formes linéaires permettent de caractériser analytiquement l'appartenance à un hyperplan. Le résultat que l'on vient de voir est donc naturellement utilisé dans l'exercice suivant.

7.9. Tout hyperplan de $\mathcal{M}_n(\mathrm{K})$ coupe $\mathrm{GL}_n(\mathrm{K})$

Montrer que pour $n \geqslant 2$, tout hyperplan de $\mathcal{M}_n(\mathrm{K})$ rencontre $\mathrm{GL}_n(\mathrm{K})$.

(École polytechnique)

▷ **Solution.**

Soit H un hyperplan de $\mathcal{M}_n(\mathrm{K})$. C'est donc le noyau d'une forme linéaire non nulle f. D'après l'exercice précédent, il existe $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$ non nulle, telle que pour tout $\mathrm{X} \in \mathcal{M}_n(\mathrm{K})$, $f(\mathrm{X}) = \mathrm{Tr}(\mathrm{AX})$.

Le problème est donc de montrer qu'il existe X inversible telle que AX soit de trace nulle. On va une fois de plus utiliser la notion de matrice équivalente. Notons $r \geqslant 1$ le rang de A. Nous savons qu'il existe P et Q dans $\mathrm{GL}_n(\mathrm{K})$ tels que $\mathrm{PAQ} = \mathrm{J}_r = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$. Alors, si $\mathrm{X} \in \mathcal{M}_n(\mathrm{K})$, $\mathrm{Tr}(\mathrm{AX}) = \mathrm{Tr}(\mathrm{PJ}_r\mathrm{QX}) = \mathrm{Tr}(\mathrm{J}_r\mathrm{QXP})$. Il nous suffit donc de trouver une matrice inversible Y telle que $\mathrm{Tr}(\mathrm{J}_r\mathrm{Y}) = 0$ (on posera $\mathrm{X} = \mathrm{Q}^{-1}\mathrm{YP}^{-1}$ qui sera donc dans $\mathrm{GL}_n(\mathrm{K})$ et dans H). La matrice de permutation

$$\mathrm{Y} = \begin{pmatrix} 0 & 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \ddots & & & 0 \\ 0 & 1 & 0 & & & \vdots \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 1 & 0 \end{pmatrix}$$

convient car $\mathrm{J}_r\mathrm{Y}$ a sa diagonale nulle. ◁

Une question se pose alors naturellement. Quelle est la dimension maximale d'un sous-espace vectoriel de $\mathcal{M}_n(\mathrm{K})$ qui ne rencontre pas le groupe linéaire ? C'est au moins $n(n-1)$ puisque le sous-espace formé des matrices dont la dernière colonne est nulle convient. L'exercice suivant détermine plus généralement la dimension maximale d'un sous-espace vectoriel de $\mathcal{M}_n(\mathrm{K})$ dont tous les éléments sont de rang inférieur à p, le cas $p = n - 1$ répondant à la question posée au début.

7.10. Dimension maximale de sous-espaces de matrices de rang $\leqslant p$

Soit V un sous-espace de $\mathcal{M}_n(\mathbb{R})$ tel que, pour tout $\mathrm{M} \in \mathrm{V}$, on ait $\mathrm{rg}\,\mathrm{M} \leqslant p$. On souhaite démontrer que $\dim \mathrm{V} \leqslant pn$.

1. Montrer que que l'on peut se ramener au cas où V contient la matrice $\left(\begin{array}{c|c} \mathrm{I}_p & 0 \\ \hline 0 & 0 \end{array}\right)$. Cette condition sera supposée réalisée par la suite.

Montrer que toute matrice de V peut s'écrire $\left(\begin{array}{c|c} A & C \\ \hline B & 0 \end{array}\right)$, avec $A \in \mathcal{M}_p(\mathbb{R})$, $B \in \mathcal{M}_{n-p,p}(\mathbb{R})$, $C \in \mathcal{M}_{p,n-p}(\mathbb{R})$ et $BC = 0$.

2. Conclure, en considérant l'application qui à tout élément $\left(\begin{array}{c|c} A & C \\ \hline B & 0 \end{array}\right)$ de V associe $(A, {}^tB + C)$, élément de l'espace vectoriel $E = \mathcal{M}_p(\mathbb{R}) \times \mathcal{M}_{p,n-p}(\mathbb{R})$.

3. Montrer que le résultat reste vrai lorsqu'on remplace le corps de base $\mathbb{R}$ par un corps quelconque infini K. Pour cela, on considérera l'application q qui à $\left(\begin{array}{c|c} A & C \\ \hline B & 0 \end{array}\right)$ associe $\left(\begin{array}{c|c} 0 & C \\ \hline B & 0 \end{array}\right)$, $W = q(V)$ et l'application $\Phi : \left(\begin{array}{c|c} 0 & C \\ \hline B & 0 \end{array}\right) \in W \longmapsto C \in \mathcal{M}_{p,n-p}(K)$ et on montrera que si $\left(\begin{array}{c|c} 0 & 0 \\ \hline B & 0 \end{array}\right) \in \operatorname{Ker}\Phi$, alors pour tout $C' \in \operatorname{Im}\Phi$, on a $BC' = 0$. On en déduira que $\dim W \leqslant p(n-p)$, avant de conclure.

(**École normale supérieure**)

▷ **Solution.**

1. • On peut supposer que V contient une matrice de rang p (sinon on remplace p par le rang maximal d'un élément de V). Une telle matrice peut s'écrire $R\left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array}\right)S$, avec R et S inversibles. On considère alors $V' = \{R^{-1}MS^{-1}, M \in V\}$, qui est un sous-espace de $\mathcal{M}_n(\mathbb{R})$ de même dimension que V (puisque l'application $M \longmapsto R^{-1}MS^{-1}$ est un automorphisme de $\mathcal{M}_n(\mathbb{R})$) et dont les éléments sont encore de rang inférieur à p. Il suffit donc de démontrer le résultat pour V', qui contient la matrice $P = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array}\right)$. On suppose désormais que V contient P. Soit $M = \left(\begin{array}{c|c} A & C \\ \hline B & D \end{array}\right)$ dans V, avec $A \in \mathcal{M}_p(\mathbb{R})$, $B \in \mathcal{M}_{n-p,p}(\mathbb{R})$, $C \in \mathcal{M}_{p,n-p}(\mathbb{R})$ et $D \in \mathcal{M}_{n-p}(\mathbb{R})$.

• Considérons, pour $\lambda \in \mathbb{R}$, la matrice $M - \lambda P = \left(\begin{array}{c|c} A - \lambda I_p & C \\ \hline B & D \end{array}\right)$. Celle-ci est dans V ; son rang est donc inférieur ou égal à p. Considérons une matrice d'ordre $p+1$ extraite de $M-\lambda P$ et bordant la matrice $A-\lambda I_p$. Elle sera donc de la forme $\left(\begin{array}{c|c} A - \lambda I_p & Y_j \\ \hline X_i & d_{ij} \end{array}\right)$, avec $1 \leqslant i, j \leqslant n-p$, d_{ij} étant un coefficient quelconque de la matrice D, X_i la i-ème ligne de B et Y_j la j-ème colonne de C. Les indices i et j étant fixés, considérons l'application f qui, à tout réel λ associe le déterminant de la matrice

$\left(\begin{array}{c|c} \mathrm{A}-\lambda \mathrm{I}_p & \mathrm{Y}_j \\ \hline \mathrm{X}_i & d_{ij} \end{array}\right)$. C'est une fonction polynomiale. La matrice $\mathrm{M}-\lambda \mathrm{P}$ étant de rang inférieur ou égal à p, puisque dans V, on a $f(\lambda)=0$ pour tout réel λ : f est la fonction nulle.

$\star$ Le coefficient de λ^p dans f est $(-1)^p d_{ij}$. On a donc $d_{ij}=0$. Comme cela est vrai pour tout (i,j) tel que $1 \leqslant i,j \leqslant n-p$, on en déduit que $\mathrm{D}=0$.

$\star$ Examinons maintenant le coefficient de λ^{p-1}. Nommons e_1, $e_2,\ldots$, e_n les vecteurs de la base canonique de $\mathbb{R}^n$, $\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_p$ les vecteurs colonnes de la matrice $\begin{pmatrix} \mathrm{A} \\ \mathrm{X}_i \end{pmatrix}$ et $\mathrm{Z}=\begin{pmatrix} \mathrm{Y}_j \\ 0 \end{pmatrix}$. Alors, pour tout $\lambda \in \mathbb{R}$,

$$f(\lambda)=\det(\mathrm{C}_1-\lambda e_1, \mathrm{C}_2-\lambda e_2, \ldots, \mathrm{C}_p-\lambda e_p, \mathrm{Z}).$$

On calcule ce déterminant en utilisant la multilinéarité et on trouve que le coefficient de λ^{p-1} est

$$\sum_{k=1}^{p} \det(-\lambda e_1, \ldots, -\lambda e_{k-1}, \mathrm{C}_k, -\lambda e_{k+1}, \ldots, -\lambda e_p, \mathrm{Z}).$$

En posant $\mathrm{X}_i=(x_1, x_2, \ldots, x_p)$ et $\mathrm{Y}_j = {}^t(y_1, y_2, \ldots, y_p)$, on obtient, pour $1 \leqslant k \leqslant p$,

$$\det(-\lambda e_1, \ldots, -\lambda e_{k-1}, \mathrm{C}_k, -\lambda e_{k+1}, \ldots, -\lambda e_p, \mathrm{Z}) =$$

$$\begin{vmatrix} -\lambda & & a_{1k} & & & y_1 \\ & \ddots & \vdots & & & \vdots \\ & & a_{kk} & & & y_k \\ & & \vdots & \ddots & & \vdots \\ & & a_{pk} & & -\lambda & y_p \\ 0 & \ldots & x_k & \ldots & 0 & 0 \end{vmatrix} = (-\lambda)^{p-1} \begin{vmatrix} a_{kk} & y_k \\ x_k & 0 \end{vmatrix} = (-\lambda)^p x_k y_k.$$

Le coefficient de λ^{p-1} est donc $(-1)^p \sum_{k=1}^{p} x_k y_k = (-1)^{p-1} \mathrm{X}_i \mathrm{Y}_j$. On a donc, pour $1 \leqslant i,j \leqslant n-p$, $\mathrm{X}_i \mathrm{Y}_j = 0$. Puisque X_i est le i-ième vecteur ligne de B et Y_j le j-ième vecteur colonne de C, $\mathrm{X}_i \mathrm{Y}_j$ est le coefficient d'indice ij de BC. On obtient finalement $\mathrm{BC}=0$.

Nous avons démontré que tout élément de V s'écrit $\mathrm{M}=\left(\begin{array}{c|c} \mathrm{A} & \mathrm{C} \\ \hline \mathrm{B} & 0 \end{array}\right)$, avec $\mathrm{A} \in \mathcal{M}_p(\mathbb{R})$, $\mathrm{B} \in \mathcal{M}_{n-p,p}(\mathbb{R})$, $\mathrm{C} \in \mathcal{M}_{p,n-p}(\mathbb{R})$ et $\mathrm{BC}=0$.

2. Considérons l'application Φ qui à toute matrice $\mathrm{M} \in \mathrm{V}$, écrite comme ci-dessus, associe le couple $(\mathrm{A}, {}^t\mathrm{B}+\mathrm{C})$, élément de l'espace vectoriel $\mathrm{E}=\mathcal{M}_p(\mathbb{R}) \times \mathcal{M}_{p,n-p}(\mathbb{R})$. L'application Φ est clairement

linéaire. Montrons qu'elle est injective. L'égalité $\Phi(M) = 0$ équivaut à $A = {}^tB + C = 0$. Sachant que $BC = 0$, on a alors $B\,{}^tB = -BC = 0$. Pour tout $k \in [\![1, p]\!]$, le coefficient d'indice (k,k) de $B\,{}^tB$ est $\sum_{j=1}^{n-p} b_{jk}^2$. On a donc $b_{jk} = 0$ pour $1 \leqslant k \leqslant p$ et $1 \leqslant j \leqslant n-p$. Finalement $B = 0$ et donc $C = 0$. La matrice M est nulle.

De l'injectivité de Φ, on déduit que $\dim V = \dim(\operatorname{Im}\Phi) \leqslant \dim E$. Sachant que $\dim E = p^2 + p(n-p) = np$, on peut donc affirmer que $\dim V \leqslant np$.

3. Par des arguments analogues à ceux développés dans la première question, on se ramène au cas où toute matrice de V s'écrit $\left(\begin{array}{c|c} A & C \\ \hline B & 0 \end{array}\right)$, avec $A \in \mathcal{M}_p(K)$, $B \in \mathcal{M}_{n-p,p}(K)$, $C \in \mathcal{M}_{p,n-p}(K)$ et $BC = 0$ (la nullité du polynôme f est assurée par le fait que tout élément de K, qui est supposé infini, est racine de f).

L'application q étant linéaire, W est un sous-espace vectoriel de $\mathcal{M}_n(K)$. D'après ce qui précède, tout élément $\left(\begin{array}{c|c} 0 & C \\ \hline B & 0 \end{array}\right)$ de W vérifie $BC = 0$. Soit $\left(\begin{array}{c|c} 0 & 0 \\ \hline B & 0 \end{array}\right) \in \operatorname{Ker}\Phi$. Si $C' = \Phi\left(\left(\begin{array}{c|c} 0 & C' \\ \hline B' & 0 \end{array}\right)\right) \in \operatorname{Im}\Phi$, alors la matrice $\left(\begin{array}{c|c} 0 & C' \\ \hline B+B' & 0 \end{array}\right)$ est dans W, comme somme de deux éléments de W, et on a donc

$$0 = (B+B')C' = BC' + B'C' = BC',$$

car $B'C' = 0$. En identifiant les matrices aux endomorphismes canoniquement associés, on obtient donc, pour tout $C' \in \operatorname{Im}\Phi$, $\operatorname{Im} C' \subset \operatorname{Ker} B$.

Considérons $W' = \left\{B \in \mathcal{M}_{n-p,p}(K),\ \left(\begin{array}{c|c} 0 & 0 \\ \hline B & 0 \end{array}\right) \in W\right\}$. Cet espace vectoriel est isomorphe à $\operatorname{Ker}\Phi$. Posons $E = \bigcap_{B \in W'} \operatorname{Ker} B$. On a, pour tout $C' \in \operatorname{Im}\Phi$, $\operatorname{Im} C' \subset E$. Les éléments de $\operatorname{Im}\Phi$ s'identifient donc à des éléments de $\mathcal{L}(K^{n-p}, E)$. On en déduit que $\dim \operatorname{Im}\Phi \leqslant (n-p)\dim E$. Si on considère un supplémentaire E' de E dans K^p, on a, pour tout $B \in W'$, $E \subset \operatorname{Ker} B$ et les éléments de W' s'identifient à des éléments de $\mathcal{L}(E', K^{n-p})$. On a donc $\dim \operatorname{Ker}\Phi = \dim W' \leqslant (n-p)\dim E'$. Du théorème du rang, on déduit alors que

$$\dim W = \dim \operatorname{Im}\Phi + \dim \operatorname{Ker}\Phi \leqslant (n-p)(\dim E + \dim E') = (n-p)p.$$

Considérons maintenant l'ensemble W'' des matrices de $\mathcal{M}_n(K)$ de la forme $\left(\begin{array}{c|c} A & 0 \\ \hline 0 & 0 \end{array}\right)$, avec $A \in \mathcal{M}_p(K)$. Il est clair que $V \subset W \oplus W''$. On en déduit que

$$\dim V \leqslant \dim W + \dim W'' \leqslant (n-p)p + p^2 \leqslant np. \quad \lhd$$

Cette dimension maximale est atteinte : on peut exhiber sans mal un sous-espace de dimension np formé de matrices de rang $\leqslant p$, par exemple l'ensemble des matrices ayant leurs $n-p$ dernières colonnes nulles.

L'exercice qui suit exploite les propriétés de l'orthogonalité duale et la connaissance des formes linéaires sur $\mathcal{M}_n(\mathrm{K})$. Pour des rappels sur la dualité, on se reportera à la page 298 du chapitre 6 (Espaces vectoriels).

7.11. Orthogonalité duale

Soit $(\mathrm{A},\mathrm{B}) \in \mathcal{M}_n(\mathrm{K})^2$. Montrer l'équivalence des deux assertions suivantes :

(i) $\exists \mathrm{X} \in \mathcal{M}_n(\mathrm{K})$, $\mathrm{AX}+\mathrm{XA}=\mathrm{B}$;

(ii) $\forall \mathrm{C} \in \mathcal{M}_n(\mathrm{K})$, $\mathrm{AC}+\mathrm{CA}=0 \Longrightarrow \mathrm{Tr}(\mathrm{BC})=0$.

(**École normale supérieure**)

▷ **Solution.**

• L'implication $(i) \Longrightarrow (ii)$ résulte d'un simple calcul. Soit C dans $\mathcal{M}_n(\mathrm{K})$ telle que $\mathrm{AC}+\mathrm{CA}=0$. On a alors

$$\begin{aligned}\mathrm{Tr}(\mathrm{BC}) &= \mathrm{Tr}(\mathrm{AXC}+\mathrm{XAC}) = \mathrm{Tr}(\mathrm{AXC})+\mathrm{Tr}(\mathrm{XAC})\\ &= \mathrm{Tr}(\mathrm{CAX})+\mathrm{Tr}(\mathrm{ACX}) = \mathrm{Tr}((\mathrm{AC}+\mathrm{CA})\mathrm{X}) = 0,\end{aligned}$$

car on sait que, pour tout couple $(\mathrm{M},\mathrm{N}) \in \mathcal{M}_n(\mathrm{K})^2$, $\mathrm{Tr}(\mathrm{MN})=\mathrm{Tr}(\mathrm{NM})$.

• Pour montrer que $(ii) \Longrightarrow (i)$, interprétons les deux conditions. L'application $f : \mathcal{M}_n(\mathrm{K}) \to \mathcal{M}_n(\mathrm{K})$ définie par $f(\mathrm{X})=\mathrm{AX}+\mathrm{XA}$ est un endomorphisme de $\mathcal{M}_n(\mathrm{K})$. L'assertion (i) signifie que B appartient au sous-espace $\mathrm{Im}\, f$. Essayons de dégager la signification de (ii). Pour tout $\mathrm{C} \in \mathcal{M}_n(\mathrm{K})$, l'application $\mathrm{T_C} : \mathrm{M} \longmapsto \mathrm{Tr}(\mathrm{CM})$ est une forme linéaire sur $\mathcal{M}_n(\mathrm{K})$ et on a vu dans l'exercice 7.8 que l'application T qui à C associe la forme $\mathrm{T_C}$ est un isomorphisme de $\mathcal{M}_n(\mathrm{K})$ sur son dual. Notons F le sous-espace $\mathrm{T}(\mathrm{Ker}\, f)$ de $\mathcal{M}_n(\mathrm{K})^*$. La condition (ii) s'écrit : pour tout $\mathrm{C} \in \mathrm{Ker}\, f$, $\mathrm{T_C}(\mathrm{B})=0$. Elle signifie simplement que B appartient à F°, orthogonal dual de F. La première implication a montré que $\mathrm{Im}\, f \subset \mathrm{F}^\circ$. L'égalité provient simplement du fait que les deux espaces ont la même dimension, puisque $\dim \mathrm{F}^\circ = n^2 - \dim \mathrm{F} = n^2 - \dim \mathrm{Ker}\, f = \dim \mathrm{Im}\, f$. ◁

Si A, B *sont deux matrices de* $\mathcal{M}_n(\mathrm{K})$, *on appelle crochet de Lie de* A *et* B, *la matrice* $[\mathrm{A},\mathrm{B}] = \mathrm{AB}-\mathrm{BA}$. *Il s'agit d'une nouvelle loi de composition interne sur* $\mathcal{M}_n(\mathrm{K})$, *qui en fait une algèbre de Lie*[1].

1. Le lecteur intéressé pourra consulter : MNEIMNÉ (R.) et TESTARD (F.), *Introduction à la théorie des groupes de Lie classiques*, Hermann, 1986.

L'exercice suivant se propose de déterminer, en caractéristique nulle, les matrices qui sont des crochets de Lie.

7.12. Crochets de Lie de $\mathcal{M}_n(\mathrm{K})$

Soit K un corps de caractéristique nulle.

1. Montrer qu'une matrice de $\mathcal{M}_n(\mathrm{K})$ de trace nulle est semblable à une matrice de diagonale nulle.

2. Montrer que si une matrice $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$ est de trace nulle, il existe B et C dans $\mathcal{M}_n(\mathrm{K})$ telles que $\mathrm{A} = \mathrm{BC} - \mathrm{CB} = [\mathrm{B}, \mathrm{C}]$ (crochet de Lie).

(**École normale supérieure**)

▷ **Solution.**

1. On va procéder par récurrence sur $n \geqslant 1$, le résultat étant trivial pour $n = 1$. Supposons $n \geqslant 2$ et A non nulle. Alors, A n'est pas la matrice d'une homothétie car si $\lambda \in \mathrm{K}^*$, la trace de $\lambda \mathrm{I}$ n'est pas nulle, puisque K est un corps de caractéristique nulle. Il existe donc $e \in \mathrm{K}^n$ tel que $(e, \mathrm{A}e)$ soit libre (d'après le résultat rappelé avant l'exercice 6.6). On pose $e'_1 = e$ et $e'_2 = \mathrm{A}(e)$ et on complète le système libre (e'_1, e'_2) en une base $\mathcal{B}' = (e'_1, e'_2, \ldots, e'_n)$ de K^n. On note $\mathcal{B} = (e_1, e_2, \ldots, e_n)$ la base canonique de K^n. Si P est la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$, $\mathrm{P}^{-1}\mathrm{AP}$ s'écrit

$$\mathrm{P}^{-1}\mathrm{AP} = \left(\begin{array}{c|c} 0 & \times \ldots\ldots \times \\ \hline \mathbf{1} & \\ 0 & \\ \vdots & \mathrm{A}' \\ 0 & \end{array}\right) \quad \text{avec } \mathrm{A}' \in \mathcal{M}_{n-1}(\mathrm{K}).$$

On a alors $\operatorname{Tr} \mathrm{A} = 0 + \operatorname{Tr} \mathrm{A}' = 0$. D'après l'hypothèse de récurrence, il existe $\mathrm{Q} \in \mathrm{GL}_{n-1}(\mathrm{K})$ tel que $\mathrm{Q}^{-1}\mathrm{A}'\mathrm{Q}$ soit à diagonale nulle. Posons

$$\mathrm{P}' = \left(\begin{array}{c|c} \mathbf{1} & 0\,0\,\ldots\,0 \\ \hline 0 & \\ \vdots & \mathrm{Q} \\ 0 & \end{array}\right) \in \mathrm{GL}_n(\mathrm{K}).$$

On obtient, d'après les règles de multiplication par blocs,

$$\mathrm{P}'^{-1}\mathrm{P}^{-1}\mathrm{APP}' = \left(\begin{array}{c|c} 0 & \times \ldots\ldots \times \\ \hline \times & \\ \vdots & \mathrm{Q}^{-1}\mathrm{A}'\mathrm{Q} \\ \times & \end{array}\right).$$

Comme $Q^{-1}A'Q$ est de diagonale nulle, $(PP')^{-1}APP'$ l'est aussi. La récurrence est achevée.

Ce résultat ne demeure pas en caractéristique non nulle. Par exemple, pour tout p premier, Id_p a une trace nulle dans $\mathcal{M}_p(\mathbb{Z}/p\mathbb{Z})$ et n'est évidemment semblable qu'à elle-même.

2. Nous savons que $\mathrm{Tr}(BC) = \mathrm{Tr}(CB)$, pour tous B et C dans $\mathcal{M}_n(K)$. On en déduit que tout crochet de Lie est de trace nulle. Il faut démontrer la réciproque. Observons, pour commencer, que si $A = [B, C]$ est un crochet de Lie, alors toute matrice semblable à A en est un aussi, puisque, pour tout $P \in \mathrm{GL}_n(K)$, $P^{-1}AP = [P^{-1}BP, P^{-1}CP]$. La question précédente permet donc de supposer que la diagonale de A est nulle.

L'idée est de fixer une matrice B «simple», par exemple diagonale, et de regarder l'ensemble des crochets de Lie $[B, C]$, où C décrit $\mathcal{M}_n(K)$. Considérons $B = \mathrm{Diag}(b_1, \ldots, b_n)$ et considérons l'endomorphisme ad_B de $\mathcal{M}_n(K)$ définie par $ad_B(C) = [B, C]$. En notant $C = (c_{ij})$, on obtient $ad_B(C) = [B, C] = ((b_i - b_j)c_{ij})$. On voit déjà que $\mathrm{Im}\, ad_B$ est inclus dans l'ensemble des matrices de diagonale nulle. Si on prend soin de choisir les b_i deux à deux distincts (ce qui est possible puisque K est infini), on a alors $[B, C] = 0$ si et seulement si C est diagonale. Le noyau de ad_B (qui n'est autre que le commutant de B) est dans ce cas de dimension n. Par le théorème du rang, on obtient $\dim(\mathrm{Im}\, ad_B) = n^2 - n$. C'est exactement la dimension du sous-espace vectoriel de $\mathcal{M}_n(K)$ formé des matrices de diagonale nulle ; celui-ci est donc égal à $\mathrm{Im}\, ad_B$.

Conclusion. Toute matrice de trace nulle est un crochet de Lie. ◁

Le lecteur intéressé pourra consulter le problème de l'ENSET de 1983 où l'on démontre que, pour $n \geqslant 3$, le résultat de la seconde question reste vrai pour tout corps K différent de $\mathbb{Z}/2\mathbb{Z}$.

7.13. Diagonales possibles dans une classe de similitude

Soit $A \in \mathcal{M}_n(K)$. Quels sont les n-uplets $(a_1, \ldots, a_n) \in K^n$ tels qu'il existe une matrice de diagonale $(a_1, \ldots, a_n)$ semblable à A ?

(**École normale supérieure**)

▷ **Solution.**

On identifiera matrice et endomorphisme canoniquement associé. Il est connu que si A est une matrice scalaire λI_n, elle n'est semblable qu'à elle même. Dans ce cas particulier, seul $(\lambda, \ldots, \lambda)$ convient.

Supposons désormais que A n'est pas une matrice scalaire. Nous savons que la trace est un invariant de similitude. Si $(a_1, \ldots, a_n)$ convient, nécessairement $a_1 + \cdots + a_n = \mathrm{Tr}\, A$. Nous allons démontrer par récurrence sur n que $a_1 + \cdots + a_n = \mathrm{Tr}\, A$ est une condition suffisante pour A soit semblable à une matrice de diagonale $(a_1, \ldots, a_n)$.

Prenons $n \geqslant 2$, $A \in \mathcal{M}_n(K)$ et $(a_1, \dots, a_n) \in K^n$ de somme $\operatorname{Tr} A$. La matrice $A' = A - a_1 I_n$ n'est pas scalaire. Classiquement, il existe donc e_1 un vecteur de K^n tel que $(e_1, A'e_1)$ soit libre (voir page 269). On complète ce système libre en une base $(e_1, e_2, \dots, e_n)$ de K^n (avec $e_2 = A'e_1$). Via ce changement de base, la matrice A' est donc semblable à une matrice qui admet comme première colonne $\begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ et A est semblable à

$$A'' = \begin{pmatrix} a_1 & \times & \times & \times \\ 1 & & & \\ 0 & & B & \\ \vdots & & & \\ 0 & & & \end{pmatrix}.$$

Si $n = 2$, B est un scalaire b qui vérifie $\operatorname{Tr} A = \operatorname{Tr} A'' = a_1 + b$; c'est forcément a_2. Le résultat est vérifié.

Supposons $n \geqslant 3$. Si B n'est pas une matrice scalaire, comme sa trace est

$$\operatorname{Tr} B = \operatorname{Tr} A - a_1 = a_2 + \cdots + a_n,$$

il existe, par hypothèse de récurrence, P inversible de taille $n-1$ telle que $P^{-1}BP$ soit de diagonale $(a_2, \dots, a_n)$. Si on pose $Q = \begin{pmatrix} 1 & 0 \\ 0 & P \end{pmatrix}$, un produit par blocs donne

$$Q^{-1}A''Q = \begin{pmatrix} a_1 & \times \\ \times & P^{-1}BP \end{pmatrix},$$

qui est une matrice de diagonale $(a_1, \dots, a_n)$.

Reste à traiter le cas où B est scalaire de la forme λI_{n-1}. Changeons dans la base e_3 en $e_3' = e_3 + e_1$. La famille $(e_1, e_2, e_3', e_4, \dots, e_n)$ est encore une base de K^n. Par ce changement de base, on voit que la matrice A est semblable à

$$\begin{pmatrix} a_1 & \times & \times & \times & \times & \times \\ 1 & \lambda & 1 & 0 & \vdots & 0 \\ 0 & 0 & \lambda & & & \cdots \\ \vdots & \vdots & & \ddots & & \vdots \\ \vdots & \vdots & & & \ddots & 0 \\ 0 & 0 & \dots & \dots & 0 & \lambda \end{pmatrix}$$

ce qui nous ramène au cas où B n'est pas scalaire. Le résultat est ainsi démontré. ◁

Voici, pour terminer cette série sur les traces, un joli petit exercice de congruence qu'on peut regarder comme une généralisation du petit théorème de Fermat.

7.14. Traces modulo p

Soit p un nombre premier, A et B deux matrices de $\mathcal{M}_n(\mathbb{Z})$.

1. Montrer que $\operatorname{Tr}(A+B)^p \equiv \operatorname{Tr} A^p + \operatorname{Tr} B^p \ [p]$.

2. En déduire que $\operatorname{Tr} A^p \equiv \operatorname{Tr} A \ [p]$.

3. Soit la suite récurrente (u_n) définie par $u_0 = 3$, $u_1 = 0$, $u_2 = 2$ et, pour tout $n \in \mathbb{N}$, $u_{n+3} = u_{n+1} + u_n$. Montrer que p divise u_p pour tout p premier.

(École normale supérieure)

▷ **Solution.**

1. Si les matrices A et B commutent, on peut appliquer la formule du binôme : $(A+B)^p = \sum_{k=0}^{n} C_p^k A^k B^{p-k}$. On prend la trace et on passe modulo p. Comme p divise C_p^k pour $1 \leqslant k \leqslant p-1$, il vient

$$\operatorname{Tr}(A+B)^p \equiv \sum_{k=0}^{p} C_p^k \operatorname{Tr}(A^k B^{p-k}) \equiv \operatorname{Tr} A^p + \operatorname{Tr} B^p \ [p].$$

Dans le cas général, en absence de commutativité, le développement de $(A+B)^p$ donne une somme de 2^p termes :

$$(A+B)^p = \sum_{(A_1,A_2,\ldots,A_p)\in\{A,B\}^p} A_1 A_2 \ldots A_p.$$

Dans la somme de droite, certains p-uplets vont conduire à la même trace grâce à la propriété $\operatorname{Tr}(XY) = \operatorname{Tr}(YX)$.

Précisons cela en utilisant une opération de groupe. Soit G le sous-groupe de $\mathcal{S}_p$ engendré par le p-cycle $c = (1, 2, \ldots, p)$. C'est un groupe cyclique de cardinal p. On définit une opération de groupe de G sur $\{A,B\}^p$ par $c \cdot (A_1, \ldots, A_p) = (A_2, \ldots, A_p, A_1)$. La propriété de la trace rappelée ci-dessus montre que si $(A_1, \ldots, A_p)$ et $(A'_1, \ldots, A'_p)$ sont dans la même orbite, alors $\operatorname{Tr}(A_1 \ldots A_p) = \operatorname{Tr}(A'_1 \ldots A'_p)$. Or les orbites ont pour cardinal un diviseur de $|G| = p$. Elles sont donc toutes de cardinal p, excepté celles qui sont de cardinal 1, c'est-à-dire les orbites de $(A, A, \ldots, A)$ et $(B, B, \ldots, B)$. Il en résulte que $\operatorname{Tr}(A+B)^p \equiv \operatorname{Tr} A^p + \operatorname{Tr} B^p \ [p]$.

2. Soit E le sous-ensemble de $\mathcal{M}_n(\mathbb{Z})$ formé des matrices A vérifiant $\mathrm{Tr}(A^p) \equiv \mathrm{Tr}(A)\ [p]$. La question précédente montre que E est stable pour l'addition. Si $A \in E$, une récurrence immédiate montre que $kA \in E$ pour tout $k \in \mathbb{N}$. On a ensuite, en distinguant les cas $p = 2$ et p impair,

$$\mathrm{Tr}(-A)^p \equiv (-1)^p \,\mathrm{Tr}\, A^p \equiv (-1)^p A \equiv -A\ [p],$$

ce qui montre que $-A \in E$. Finalement kA est dans E, pour tout $k \in \mathbb{Z}$. Le lecteur vérifiera facilement que les matrices E_{ij} de la base canonique sont dans E ($E_{ij}^p = 0$ si $i \neq j$ et $E_{ij}^p = E_{ij}$ si $i = j$). Il en résulte que $E = \mathcal{M}_n(\mathbb{Z})$, ce qui est le résultat demandé.

3. Considérons la matrice $A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ associée à la relation de récurrence linéaire : si $X_n = \begin{pmatrix} u_n \\ u_{n+1} \\ u_{n+2} \end{pmatrix}$ on a $X_{n+1} = AX_n$ pour tout n. On constate que $A^0 = I_3$ est de trace 3, que A est de trace nulle et enfin que $A^2 = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$ est de trace 2. Autrement dit, on a $u_n = \mathrm{Tr}\, A^n$ pour $n \in \{0, 1, 2\}$.

Or, comme $A^3 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = A + I_3$ on a $A^{n+3} = A^{n+1} + A^n$ pour tout n et la suite $(\mathrm{Tr}\, A^n)_{n\geqslant 0}$ vérifie la même relation de récurrence que la suite $(u_n)_{n\geqslant 0}$. On en déduit que $u_n = \mathrm{Tr}\, A^n$ pour tout n. En particulier, pour tout entier p premier, on a $u_p = \mathrm{Tr}\, A^p \equiv \mathrm{Tr}\, A \equiv 0$ modulo p. D'où le résultat. $\lhd$

Le résultat de la seconde question s'obtient directement si on admet que $\mathbb{Z}/p\mathbb{Z}$ possède un sur-corps K, algébriquement clos. Notons $\bar{A}$ la matrice obtenue à partir de A en prenant les classes modulo p de ses coefficients. La matrice $\bar{A}$ est trigonalisable dans K. Si on note $\lambda_1, \ldots, \lambda_p$ les valeurs propres de $\bar{A}$ comptées avec multiplicité, on a

$$\mathrm{Tr}(\bar{A}^p) = \sum_{i=1}^{p} \lambda_i^p = \left(\sum_{i=1}^{p} \lambda_i\right)^p = (\mathrm{Tr}\, \bar{A})^p,$$

l'avant-dernière égalité provenant du fait que K est de caractéristique p (pour x et y dans K, on $(x+y)^p = \sum_{k=0}^{p} C_p^k x^k y^{p-k} = x^p + y^p$, puisque p divise C_p^k si $1 \leqslant k \leqslant p-1$). Comme la trace de $\bar{A}$ est le résidu modulo p de la trace de A, on obtient $\mathrm{Tr}\, A^p \equiv (\mathrm{Tr}\, A)^p$ et on conclut avec le petit théorème de Fermat.

Les exercices qui suivent concernent les matrices réelles positives, c'est-à-dire dont tous les coefficients sont positifs. Dans le tome 2 d'algèbre, on trouvera plusieurs exercices sur les propriétés spectrales des matrices positives, notamment le théorème de Perron-Frobenius (exercice 2.10).

7.15. Matrices monotones

On dit qu'une matrice à coefficients réels A est positive, ce qu'on note $A \geqslant 0$, si tous ses coefficients sont positifs ou nuls. On dit que $A \in \mathcal{M}_n(\mathbb{R})$ est monotone si elle est inversible et si $A^{-1} \geqslant 0$.

1. Soit $A \in \mathcal{M}_n(\mathbb{R})$. Montrer que $A \geqslant 0$ si et seulement si pour tout vecteur colonne X on a : $X \geqslant 0 \Longrightarrow AX \geqslant 0$.

2. Soit $A \in \mathcal{M}_n(\mathbb{R})$. Montrer que A est monotone si et seulement si pour tout vecteur colonne X on a : $AX \geqslant 0 \Longrightarrow X \geqslant 0$.

3. Soit $(c_1, \ldots, c_n) \in (\mathbb{R}_+)^n$. Montrer que la matrice suivante est monotone :

$$\begin{pmatrix} 2+c_1 & -1 & 0 & \ldots & 0 \\ -1 & 2+c_2 & -1 & \ldots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & -1 \\ 0 & \ldots & \ldots & -1 & 2+c_n \end{pmatrix}.$$

4. Déterminer les matrices qui sont à la fois positives et monotones.

(École polytechnique)

▷ **Solution.**

1. Il est clair que le produit de deux matrices positives est encore une matrice positive. Ainsi, si $A \geqslant 0$, on a pour tout $X \geqslant 0$, $AX \geqslant 0$. Réciproquement, les vecteurs $e_1, \ldots, e_n$ de la base canonique de $\mathbb{R}^n$ sont positifs. Il en résulte que les colonnes $Ae_1, \ldots, Ae_n$ de A sont toutes positives. D'où l'équivalence voulue.

2. Supposons A monotone. Comme A^{-1} est positive, si X est tel que $AX \geqslant 0$ alors $A^{-1}AX = X \geqslant 0$ d'après la question précédente.

Réciproquement, supposons que pour tout vecteur colonne X, $AX \geqslant 0$ implique $X \geqslant 0$. Montrons d'abord que A est inversible. Soit $X \in \operatorname{Ker} A$. Comme $AX = 0 \geqslant 0$, on a $X \geqslant 0$. Mais comme $-X$ est aussi dans le noyau de A, on a également $-X \geqslant 0$, de sorte que $X = 0$. A appartient donc à $\mathrm{GL}_n(\mathbb{R})$. L'hypothèse peut s'écrire : $AX \geqslant 0$ implique $A^{-1}AX \geqslant 0$ et la question précédente permet de conclure que A^{-1} est positive.

3. On va utiliser la caractérisation obtenue dans la question 2.

Soit $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ tel que $AX \geqslant 0$. En posant $x_0 = x_{n+1} = 0$, cette condition s'écrit : pour tout $k \in [\![1, n]\!]$, $(2 + c_k).x_k \geqslant x_{k-1} + x_{k+1}$. Notons p le plus petit des indices ℓ de $[\![1, n]\!]$ tels que $x_\ell = \min_{1 \leqslant k \leqslant n} x_k$.

• Si $p = 1$ ou n, on tire des inégalités précédentes $(2 + c_p)x_p \geqslant x_p$, soit $(1 + c_p)x_p \geqslant 0$ et donc $x_p \geqslant 0$.

• Si $p \in [\![2, n-1]\!]$, on obtient $(2 + c_p)x_p \geqslant 2x_p$, c'est-à-dire $c_p x_p \geqslant 0$. Si $c_p > 0$, on conclut comme précédemment $x_p \geqslant 0$. Mais il est impossible d'avoir $c_p = 0$, car l'inégalité $2x_p \geqslant x_{p-1} + x_{p+1}$ impose que l'on ait $x_p = x_{p-1} = x_{p+1}$, ce qui est contraire à la définition de p.

On a donc $x_p \geqslant 0$, ce qui implique $X \geqslant 0$, par définition de p.

4. Montrons que les matrices positives et monotones sont les matrices positives ayant exactement un terme non nul par ligne et par colonne, autrement dit les matrices $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ pour lesquelles il existe une permutation $\sigma \in \mathcal{S}_n$ telle que, pour tout $i \in [\![1, n]\!]$, $a_{i,\sigma(i)} > 0$ et $a_{ij} = 0$ si $j \neq \sigma(i)$.

Un telle matrice est évidemment positive et inversible. D'autre part, si on considère $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, on obtient $AX = \begin{pmatrix} a_{1,\sigma(1)}x_{\sigma(1)} \\ \vdots \\ a_{n,\sigma(n)}x_{\sigma(n)} \end{pmatrix}$. La condition $AX \geqslant 0$ impose $x_{\sigma(i)} \geqslant 0$, pour tout i et donc $X \geqslant 0$. La matrice A est donc monotone d'après la question 2.

Réciproquement, soit A une matrice positive et monotone. Les termes de sa première colonne sont positifs et l'un au moins est strictement positif, puisque A est inversible. Notons $J = \{i \in [\![1, n]\!], a_{i,1} > 0\}$. Soit $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ et $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = AX$. Fixons $x_2, \ldots, x_n$ strictement positifs. Si $i \notin J$, alors $y_i \geqslant 0$, quel que soit le choix de x_1. Si $i \in J$, alors la condition $y_i \geqslant 0$ équivaut à $x_1 \geqslant -\dfrac{1}{a_{i1}} \sum_{j=2}^{n} a_{ij}x_j$. Notons z_i le terme de droite de cette inégalité. Si, pour tout $i \in J$, on a $z_i < 0$, alors on pourra trouver $x_1 < 0$ vérifiant cette série d'inégalités. On aura alors $AX \geqslant 0$ sans avoir $X \geqslant 0$, ce qui est contradictoire avec A monotone. Il existe donc i tel que $z_i = 0$, c'est-à-dire tel que $a_{ij} = 0$ pour tout $j \neq 1$. Un tel i est unique, sinon A aurait deux lignes proportionnelles. On peut effectuer le même raisonnement pour toutes les colonnes. Pour

tout $j \in [\![1,n]\!]$, il existe un unique $i \in [\![1,n]\!]$ tel que tous les termes de la i-ième ligne soient nuls sauf le j-ième. L'application $j \longmapsto i$ est bijective et en la notant σ^{-1}, A a la forme voulue. $\lhd$

L'exercice suivant étudie les puissances d'une matrice strictement stochastique. Les matrices stochastiques interviennent en probabilités (d'où leur nom). Si X *et* Y *sont deux variables aléatoires à valeurs dans le même ensemble fini* $\mathrm{E} = [\![1,k]\!]$, *la matrice* $\mathrm{A} = (a_{ij}) \in \mathcal{M}_k(\mathbb{R})$ *définie par* $a_{ij} = \mathrm{P}(\mathrm{Y} = j \mid \mathrm{X} = i)$ *est stochastique, ce qui par définition signifie qu'on a* $a_{ij} \geqslant 0$ $(1 \leqslant i,j \leqslant k)$ *et* $\sum_{j=1}^{k} a_{ij} = 1$ $(1 \leqslant i \leqslant k)$.

Indiquons dans quel contexte on rencontre les matrices stochastiques. L'évolution d'un système susceptible de prendre un nombre fini d'états 1, ..., k *est représentée mathématiquement par une suite* $(\mathrm{X}_n)_{n\geq 0}$ *de variables aléatoires à valeurs dans* E. *C'est ce qu'on appelle un processus aléatoire (ou stochastique). Si* X_{n+1} *s'obtient à partir de la valeur de* X_n *et d'un tirage au sort effectué selon une loi ne dépendant que de cette valeur, on dit que le processus est une chaîne de Markov. Les exemples abondent : marches aléatoires, fortune d'un joueur, modélisation de l'alternance des voyelles et des consonnes dans un poème de Pouchkine (par Markov lui-même), ou prévision (en probabilité) des états successifs d'un signal pour améliorer la compression en traitement du signal (Shannon).*

Techniquement, on dit qu'une suite de variables aléatoires (X_n) *est une chaîne de Markov si «la loi de l'état* $n+1$ *conditionnelle au passé ne dépend que de l'état antérieur* n*», ce qui se traduit par*

$$\mathrm{P}(\mathrm{X}_{n+1} = j \mid \mathrm{X}_0 = i_0, \mathrm{X}_1 = i_1, \ldots, \mathrm{X}_n = i_n) = \mathrm{P}(\mathrm{X}_{n+1} = j \mid \mathrm{X}_n = i_n).$$

Si la matrice $\mathrm{A} = (a_{ij}) \in \mathcal{M}_k(\mathbb{R})$ *définie par* $a_{ij} = \mathrm{P}(\mathrm{X}_{n+1} = j \mid \mathrm{X}_n = i)$ *est indépendante de* n, *on dit que la chaîne de Markov est stationnaire.*

Si, dans ce dernier cas, on introduit pour $n \geqslant 0$ *le vecteur-colonne* $\mathrm{Y}_n = \begin{pmatrix} \mathrm{P}(\mathrm{X}_n = 1) \\ \vdots \\ \mathrm{P}(\mathrm{X}_n = k) \end{pmatrix}$, *on obtient* $\mathrm{Y}_{n+1} = \mathrm{A}\mathrm{Y}_n$ *et donc* $\mathrm{Y}_n = \mathrm{A}^n \mathrm{Y}_0$.

Le comportement (probabiliste) d'une chaîne de Markov stationnaire, et notamment son comportement asymptotique, est donc entièrement décrit par la donnée de la loi initiale Y_0 *et des puissances de* A.

7.16. Puissances d'une matrice strictement stochastique

Soit $\mathrm{A} = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ une matrice strictement stochastique, c'est-à-dire vérifiant

$$\forall (i,j) \in [\![1,n]\!]^2, \quad a_{ij} > 0 \quad \text{et} \quad \forall i \in [\![1,n]\!], \ \sum_{j=1}^{n} a_{ij} = 1.$$

Soit $\varepsilon = \min_{i,j} a_{ij} > 0$. Pour $k \in \mathbb{N}$, on pose $\mathrm{A}^k = (a_{ij}^{(k)})$ et, pour tout $j \in [\![1, n]\!]$, $\alpha_j^{(k)} = \min_i a_{ij}^{(k)}$, $\beta_j^{(k)} = \max_i a_{ij}^{(k)}$ et $\delta_j^{(k)} = \beta_j^{(k)} - \alpha_j^{(k)}$.

1. Montrer que pour tout $k \in \mathbb{N}^*$, A^k est strictement stochastique.

2. Montrer que

$$\alpha_j^{(k)} \leqslant \alpha_j^{(k+1)} \leqslant \beta_j^{(k+1)} \leqslant \beta_j^{(k)} \quad \text{et} \quad \delta_j^{(k+1)} \leqslant (1 - 2\varepsilon)\delta_j^{(k)}.$$

3. En déduire que la suite (A^k) converge vers une matrice que l'on précisera.

(École polytechnique)

▷ **Solution.**

1. Une matrice est strictement stochastique si et seulement si ses coefficients sont tous strictement positifs et si le vecteur $\mathrm{U} = {}^t(1, 1, \ldots, 1)$ est vecteur propre pour la valeur propre 1. Il en résulte que le produit de deux matrices strictement stochastiques l'est encore. Donc, pour tout $k \in \mathbb{N}^*$, A^k est strictement stochastique.

2. $\alpha_j^{(k)}$ et $\beta_j^{(k)}$ sont respectivement les plus petit et plus grand coefficients de la j-ième colonne de la matrice A^k. Comme $\mathrm{A}^{k+1} = \mathrm{A} \cdot \mathrm{A}^k$, on a

$$a_{ij}^{(k+1)} = \sum_{\ell=1}^{n} a_{i\ell} a_{\ell j}^{(k)}.$$

En majorant les $a_{\ell j}^{(k)}$ par $\beta_j^{(k)}$, on obtient

$$a_{ij}^{(k+1)} \leqslant \sum_{\ell=1}^{n} a_{i\ell} \beta_j^{(k)} = \beta_j^{(k)} \underbrace{\sum_{\ell=1}^{n} a_{i\ell}}_{=1} = \beta_j^{(k)}.$$

Ceci vaut pour tout $i \in [\![1, n]\!]$, donc $\beta_j^{(k+1)} = \max_i a_{ij}^{(k+1)} \leqslant \beta_j^{(k)}$. On procède de même pour la minoration $\alpha_j^{(k)} \leqslant \alpha_j^{(k+1)}$.

Soit j fixé, i_1 un indice tel que $\beta_j^{(k+1)} = a_{i_1 j}^{(k+1)}$. On a alors

$$\begin{aligned} \beta_j^{(k)} - \beta_j^{(k+1)} &= \beta_j^{(k)} - a_{i_1 j}^{(k+1)} = \beta_j^{(k)} - \sum_{\ell=1}^{n} a_{i_1 \ell} a_{\ell j}^{(k)} \\ &= \sum_{\ell=1}^{n} a_{i_1 \ell} (\beta_j^{(k)} - a_{\ell j}^{(k)}) \geqslant \varepsilon \delta_j^{(k)}, \end{aligned}$$

car tous les termes sont positifs et il existe au moins un indice ℓ tel que $a_{\ell j}^{(k)} = \alpha_j^{(k)}$ (et $a_{i_1 l} \geqslant \varepsilon$).

De même, si i_2 un indice tel que $\alpha_j^{(k+1)} = a_{i_2 j}^{(k+1)}$, on obtient

$$\alpha_j^{(k+1)} - \alpha_j^k = \sum_{\ell=1}^{n} a_{i_2 \ell}(a_{\ell j}^{(k)} - \alpha_j^{(k)}) \geqslant \varepsilon \delta_j^{(k)}.$$

En additionnant les deux inégalités, il vient $\delta_j^{(k)} - \delta_j^{(k+1)} \geqslant 2\varepsilon\delta_j^{(k)}$, c'est-à-dire $\delta_j^{(k+1)} \leqslant (1-2\varepsilon)\delta_j^{(k)}$.

3. On suppose bien entendu $n \geqslant 2$, sinon le problème est trivial. Dans ce cas, on a $\varepsilon \leqslant \dfrac{1}{2}$ et $0 \leqslant 1-2\varepsilon < 1$. Pour tout j, la suite $\delta_j^{(k)}$ tend donc vers 0 lorsque k tend vers l'infini. Les suites $(\alpha_j^{(k)})_{k\geqslant 1}$ et $(\beta_j^{(k)})_{k\geqslant 1}$ sont donc adjacentes et convergent vers une même limite $\ell_j > 0$. Il résulte de la définition de $\alpha_j^{(k)}$ et $\beta_j^{(k)}$ que pour tout $(i,j) \in [\![1,n]\!]^2$, la suite $(a_{ij}^{(k)})_{k\geqslant 1}$ converge vers ℓ_j. De l'égalité $\sum\limits_{j=1}^{n} a_{ij}^{(k)} = 1$, valable pour tous $k \geqslant 1$ et $i \in [\![1,n]\!]$, on déduit que $\sum\limits_{j=1}^{n} \ell_j = 1$.

Ceci montre que la suite $(\mathrm{A}^k)_{k\geqslant 0}$ converge vers une matrice M strictement stochastique de rang 1, dont les colonnes sont toutes proportionnelles au vecteur U. En passant à la limite dans l'égalité $\mathrm{A}^{2k} = \mathrm{A}^k\mathrm{A}^k$, on obtient $\mathrm{M}^2 = \mathrm{M}$. La matrice M est la matrice d'une projection sur Vect(U). ◁

L'exercice suivant donne une condition nécessaire et suffisante pour que tous les termes qui apparaissent dans le développement du déterminant soient nuls.

7.17. Théorème de Frobenius-König (1912-1916)

Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$. Pour $\sigma \in \mathcal{S}_n$ on appelle serpent de A associé à σ l'ensemble $\mathrm{S}_\sigma = \{a_{1,\sigma(1)}, \ldots, a_{n,\sigma(n)}\}$.

1. Montrer qu'il y a équivalence entre :

(i) tous les serpents de A contiennent 0 ;

(ii) il existe une sous-matrice nulle de A de taille (r,s), avec $r+s = n+1$.

2. On suppose que A est à coefficients positifs et que la somme des coefficients de chaque ligne et de chaque colonne vaut 1 (on dit que A est une matrice bistochastique). Montrer que A possède un serpent ne contenant pas 0.

(École normale supérieure)

▷ **Solution.**

1. • On commence par l'implication facile $(ii) \Longrightarrow (i)$. Pour I, J parties non vides de $[\![1,n]\!]$ on note $\mathrm{A}_{\mathrm{I},\mathrm{J}}$ la matrice extraite de A obtenue en gardant les coefficients $a_{i,j}$ d'indice $(i,j) \in \mathrm{I} \times \mathrm{J}$. Par hypothèse on peut trouver I (resp. J) de cardinal r (resp. s) telles que $\mathrm{A}_{\mathrm{I},\mathrm{J}} = 0$ et $r+s = n+1$. Soit $\sigma \in \mathcal{S}_n$. Il suffit de montrer que $\sigma(\mathrm{I}) \cap \mathrm{J} \neq \emptyset$ pour affirmer que le serpent S_σ contient 0. C'est évident, car sinon on aurait $r = |\sigma(\mathrm{I})| \leqslant |[\![1,n]\!] \backslash \mathrm{J}| = n - s$, ce qui contredit l'hypothèse $r + s = n + 1 > n$.

• Pour l'implication contraire, on raisonne par récurrence sur n. On note que si A vérifie (i), toute matrice obtenue à partir de A en permutant les lignes ou les colonnes vérifie encore (i).

Pour $n = 1$, le résultat est trivial, car A est nulle et $r = s = 1$ conviennent. Supposons le résultat vrai jusqu'au rang $n-1$ et considérons $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ vérifiant (i). Si A est nulle, c'est terminé. On suppose donc A non nulle et quitte à effectuer une permutation des lignes et des colonnes, on peut supposer $a_{nn} \neq 0$. Notons A' la matrice extraite $\mathrm{A}_{[\![1,n-1]\!],[\![1,n-1]\!]}$. Comme $a_{nn} \neq 0$, tous les serpents de A' contiennent 0. Par hypothèse de récurrence, on peut trouver une sous-matrice nulle de A' de taille (r,s) avec $r+s = n$. Ainsi, quitte à permuter de nouveau lignes et colonnes, on peut supposer que $\mathrm{A} = \left(\begin{array}{c|c} \mathrm{B} & \mathrm{C} \\ \hline 0 & \mathrm{D} \end{array}\right)$, où B est carrée de taille s et D carrée de taille r. Supposons que B admette un serpent qui ne contienne pas 0. Alors tous les serpents de D contiennent 0. Par hypothèse de récurrence, D admet une sous-matrice nulle D' de taille (r',s') avec $r'+s' = r+1$. La matrice extraite de A, obtenue en gardant les r' lignes correspondant aux lignes de D' ainsi que les s premières colonnes et les s' colonnes correspondant aux colonnes de D' est nulle et elle est de taille $(r', s+s')$, avec $r'+s+s' = r+s+1 = n+1$. Si tous les serpents de B contiennent 0, on applique l'hypothèse de récurrence à B et on conclut de même.

2. Supposons par l'absurde que tous les serpents de A contiennent 0. Il existe alors deux matrices de permutation P et Q telles que

$$\mathrm{PAQ} = \left(\begin{array}{c|c} \mathrm{B} & \mathrm{C} \\ \hline 0 & \mathrm{D} \end{array}\right) = \mathrm{A}',$$

où la matrice nulle est de taille (r,s) avec $r+s = n+1$. Mais la matrice A', obtenue en permutant les lignes et les colonnes de A est encore bistochastique. La somme des éléments de chaque colonne de B vaut donc 1, de même que la somme des éléments de chaque ligne de D. Mais alors la somme de tous les éléments de B et D est égale à $r+s = n+1$ ce qui est impossible, puisque la somme des éléments de A (et donc de A') est égale à n. ◁

Le premier résultat a été obtenu par Frobenius en 1912. En 1915, König en donne une preuve plus élémentaire à l'aide de la théorie des

graphes. Les deux mathématiciens se sont alors disputé la paternité du résultat. Le corollaire de la seconde question est dû à König. Il signifie que le permanent d'une matrice bistochastique A, *qui est défini par* $\mathrm{per}(\mathrm{A}) = \sum_{\sigma \in \mathcal{S}_n} \prod_{i=1}^{n} a_{i\sigma(i)}$ *est strictement positif. En 1916, Van der Waerden a conjecturé que le minimum du permanent sur le compact formé des matrices bistochastiques est* $\dfrac{n!}{n^n}$, *ce minimum n'étant atteint que pour la matrice dont tous les termes valent* $1/n$. *Ce problème est resté ouvert jusqu'en 1981 où il a été résolu indépendamment par Egorichev et Falikman. Le beau problème posé aux ENS Lyon-Cachan en 1996, en donne une preuve.*

L'énoncé suivant constitue une autre présentation de l'exercice 6.3 sur les drapeaux, qui met en évidence ce qui était sous-jacent dans celui-ci : la décomposition de Bruhat. Celle-ci s'établit à partir des opérations élémentaires sur les lignes et les colonnes utilisées dans la méthode du pivot de Gauss.

7.18. Décomposition de Bruhat et drapeaux

On pose $\mathrm{G} = \mathrm{GL}_n(\mathrm{K})$ et on note T_s le sous-groupe de G formé des matrices triangulaires supérieures inversibles. On désigne enfin par $\mathcal{D}$ l'ensemble des drapeaux de K^n.

1. Montrer, en adaptant l'algorithme du pivot de Gauss, que toute matrice $\mathrm{A} \in \mathrm{G}$ s'écrit sous la forme $\mathrm{T_1 P_\sigma T_2}$, où $\mathrm{T_1}$ et $\mathrm{T_2}$ sont dans T_s et où P_σ est une matrice de permutation, c'est-à-dire $\mathrm{P}_\sigma = (\delta_{i,\sigma(j)})_{1 \leqslant i,j \leqslant n}$ avec $\sigma \in \mathcal{S}_n$. Montrer que la permutation σ est définie de manière unique. La partition $\mathrm{G} = \bigcup_{\sigma \in \mathcal{S}_n} \mathrm{T}_s \mathrm{P}_\sigma \mathrm{T}_s$ de $\mathrm{GL}_n(\mathrm{K})$ ainsi obtenue, est appelé *décomposition de Bruhat.*

2. Montrer que G opère naturellement sur $\mathcal{D}$, de manière transitive (*i.e.* il n'y a qu'une orbite), et en déduire que $\mathcal{D}$ s'identifie à l'ensemble G/T_s des classes à gauche modulo T_s.

3. On considère l'action de groupe de G sur $\mathrm{G}/\mathrm{T}_s \times \mathrm{G}/\mathrm{T}_s$ définie par $\mathrm{A} \cdot (\overline{\mathrm{X}}, \overline{\mathrm{Y}}) = (\overline{\mathrm{AX}}, \overline{\mathrm{AY}})$. Comment cette action de groupe se traduit-elle sur les drapeaux ? Déduire de la question 1 que l'ensemble des orbites, pour l'une de ces actions, s'identifie à $\mathcal{S}_n$.

▷ **Solution.**

1. On va utiliser les matrices correspondant aux opérations élémentaires sur les lignes et les colonnes, mais en se limitant à celles qui sont dans T_s. Rappelons ce dont il s'agit. On notera $(\mathrm{E}_{ij})_{1 \leqslant i,j \leqslant n}$ la base canonique de $\mathcal{M}_n(\mathrm{K})$. On appelle matrice de transvection, toute matrice

de la forme $T_{ij}(\lambda) = Id + \lambda E_{ij}$ avec $i \neq j$ et $\lambda \in K$. Multiplier une matrice A à gauche par $T_{ij}(\lambda)$ revient à rajouter à la i-ième ligne de A, la j-ième ligne multipliée par λ et multiplier A à droite par $T_{ij}(\lambda)$ revient à rajouter à la j-ième colonne, la i-ième multipliée par λ. On appelle matrice de dilatation, toute matrice de la forme $D_i(\alpha) = Id + (\alpha - 1)E_{ii}$ où α est un scalaire non nul. Multiplier A à gauche (resp. à droite) par $D_i(\alpha)$ revient à multiplier la i-ième ligne (resp. colonne) par α. Tout matrice de dilatation est triangulaire supérieure. Une matrice $T_{ij}(\lambda)$ l'est, pour $i < j$.

Soit $A \in G$. On construit un algorithme permettant de transformer A en une matrice de permutation, en la multipliant par des matrices de transvections ou de dilatations triangulaires supérieures. D'après ce qui précède, on peut ajouter à toute ligne une combinaison linéaire de lignes d'indices supérieurs et ajouter à toute colonne une combinaison linéaire de colonnes d'indices inférieurs. La première colonne de A n'est pas nulle car $A \in G$. Notons i_1 le plus grand indice k tel que $a_{k1} \neq 0$. En multipliant A par des matrices $T_{ki_1}(\lambda)$, on peut annuler tous les coefficients a_{k1} pour $k < i_1$. À l'aide d'une matrice de dilatation, on peut remplacer le coefficient en position $(i_1, 1)$ par 1 et à l'aide d'opérations sur les colonnes, on peut annuler tous les autres coefficients de la i_1-ième ligne. Notons A_1 la matrice obtenue. Comme $A_1 \in G$, sa seconde colonne n'est pas nulle. Notons i_2 le plus grand indice k tel que le coefficient en position $(k, 2)$ de A_1 soit non nul. Comme précédemment, on peut égaler ce coefficient à 1 et mettre des 0 sur la i_2-ième ligne et la deuxième colonne partout ailleurs. On vérifie que ces opérations ne modifient pas les termes de la i_1-ième ligne et de la i_1-ième colonne. On continue l'algorithme de même avec la matrice A_2 obtenue. On construit ainsi une suite injective $i_1, i_2, \ldots, i_n$ d'entiers entre 1 et n, et en notant σ la bijection qui envoie k sur i_k, la matrice A_n obtenue à la fin de l'algorithme est la matrice de permutation P_σ. La matrice A a été multiplié à gauche et à droite par un produit d'éléments de T_s, qui est encore dans T_s. En considérant l'inverse de ces matrices, qui sont dans T_s, on écrit $T_1^{-1} A T_2^{-1} = A_n = P_\sigma$, soit encore $A = T_1 P_\sigma T_2$.

Montrons que la permutation σ est définie de manière unique. Il suffit de vérifier que, si σ et σ' sont deux permutations et T et T' deux matrices de T_s, l'égalité $TP_\sigma = P_{\sigma'} T'$ impose $\sigma = \sigma'$. On remarque, pour commencer, que la matrice TP_σ s'obtient en permutant les colonnes de T (la j-ième colonne de TP_σ est la $\sigma(j)$-ième colonne de T) et que $P_{\sigma'} T'$ s'obtient en permutant les lignes de T' (la i-ième ligne de $P_{\sigma'} T'$ est la ${\sigma'}^{-1}(i)$-ième ligne de T'). Pour tout indice j, les coefficients d'indice (i, j) de TP_σ sont donc nuls pour $i > \sigma(j)$. Le coefficient d'indice $(\sigma'(j), j)$ de $P_{\sigma'} T'$, c'est-à-dire le coefficient d'indice (j, j) de T' n'est pas nul,

car la matrice T' est triangulaire supérieure et inversible. On en déduit que $\sigma'(j) \leqslant \sigma(j)$. Par symétrie, on a, de même, $\sigma'(j) \leqslant \sigma(j)$. On obtient finalement $\sigma = \sigma'$.

Conclusion. On a obtenu $\mathrm{GL}_n(\mathrm{K}) = \bigcup_{\sigma \in \mathcal{S}_n} \mathrm{T}_s \mathrm{P}_\sigma \mathrm{T}_s$ où la réunion est disjointe, ce qui constitue la décomposition de Bruhat.

2. Si $\mathrm{A} \in \mathrm{G}$ et si $d = (\mathrm{F}_0, \mathrm{F}_1, \ldots, \mathrm{F}_n)$ est un drapeau, alors $\mathrm{A} \cdot d = (\mathrm{A}(\mathrm{F}_0), \mathrm{A}(\mathrm{F}_1), \ldots, \mathrm{A}(\mathrm{F}_n))$ est également un drapeau, puisque l'image par A d'un sous-espace vectoriel est un sous-espace de même dimension. Il est clair que cela définit une opération du groupe G sur l'ensemble $\mathcal{D}$ des drapeaux. Montrons que cette opération est transitive. Soit $(e_1, \ldots, e_n)$ la base canonique de K^n et δ le drapeau formé des sous-espaces $\mathrm{Vect}(e_1, \ldots, e_k)$, $0 \leqslant k \leqslant n$. Soit $d = (\mathrm{F}_0, \mathrm{F}_1, \ldots, \mathrm{F}_n)$ un drapeau quelconque. On peut clairement trouver une base $(f_1, \ldots, f_n)$ de K^n telle que pour tout $k \in [\![1, n]\!]$, $(f_1, f_2, \ldots, f_k)$ soit une base de F_k. L'unique matrice A vérifiant $\mathrm{A}(e_i) = f_i$ pour tout i, est inversible et vérifie $\mathrm{A} \cdot \delta = d$. L'opération de G sur $\mathcal{D}$ est donc transitive, et $\mathcal{D}$ est alors en bijection avec $\mathrm{G}/\mathrm{G}_\delta$, où G_δ est le stabilisateur du drapeau δ. Or il est clair que pour $\mathrm{A} \in \mathrm{G}$, $\mathrm{A} \cdot \delta = \delta$ si et seulement si A est triangulaire supérieure. On a donc $\mathrm{G}_\delta = \mathrm{T}_s$ et $\mathcal{D}$ est en bijection avec G/T_s.

3. Si d est un drapeau, on considère, avec les notations de la question 2, B dans G tel que $d = \mathrm{B} \cdot \delta$. On identifie d avec l'élément $\overline{\mathrm{B}}$ de G/T_s. Si A est un élément de G, on a $\mathrm{A} \cdot \overline{\mathrm{B}} = \overline{\mathrm{AB}}$ qui s'identifie à $\mathrm{AB} \cdot \delta = \mathrm{A} \cdot d$. L'action de G sur $\mathrm{G}/\mathrm{T}_s \times \mathrm{G}/\mathrm{T}_s$ se traduit donc, sur les couples de drapeaux, par l'action naturelle $\mathrm{A} \cdot (d, d') = (\mathrm{A} \cdot d, \mathrm{A} \cdot d')$, où l'action de G sur les drapeaux est celle qui est décrite dans la question précédente.

Si X et Y sont dans G, on peut écrire $(\overline{\mathrm{X}}, \overline{\mathrm{Y}}) = \mathrm{X}(\overline{\mathrm{I}_n}, \overline{\mathrm{X}^{-1}\mathrm{Y}})$. En décomposant $\mathrm{X}^{-1}\mathrm{Y}$ en $\mathrm{T}_1 \mathrm{P}_\sigma \mathrm{T}_2$, comme dans la question 1, on obtient

$$(\overline{\mathrm{I}_n}, \overline{\mathrm{X}^{-1}\mathrm{Y}}) = \mathrm{T}_1(\overline{\mathrm{T}_1^{-1}}, \overline{\mathrm{P}_\sigma \mathrm{T}_2}) = \mathrm{T}_1(\overline{\mathrm{I}_n}, \overline{\mathrm{P}_\sigma}),$$

car T_1 et T_2 sont éléments de T_s, et finalement $(\overline{\mathrm{X}}, \overline{\mathrm{Y}}) = \mathrm{X}\mathrm{T}_1(\overline{\mathrm{I}_n}, \overline{\mathrm{P}_\sigma})$. Toute orbite contient donc un élément de la forme $(\overline{\mathrm{I}_n}, \overline{\mathrm{P}_\sigma})$. La permutation σ est définie de manière unique, car s'il existe $\mathrm{A} \in \mathrm{G}$ tel que $(\overline{\mathrm{I}_n}, \overline{\mathrm{P}_\sigma}) = \mathrm{A}(\overline{\mathrm{I}_n}, \overline{\mathrm{P}_{\sigma'}}) = (\overline{\mathrm{A}}, \overline{\mathrm{AP}_{\sigma'}})$, alors A est dans T_s et il existe $\mathrm{T} \in \mathrm{T}_s$ tel que $\mathrm{AP}_{\sigma'} = \mathrm{P}_\sigma \mathrm{T}$. D'après la question 1, ceci implique $\sigma = \sigma'$. On obtient ainsi une bijection de l'ensemble des orbites sur $\mathcal{S}_n$. ◁

On retrouve ainsi le résultat de l'exercice 6.3 : le nombre de classe de $\mathrm{D} \times \mathrm{D}$ *sous l'action de* $\mathrm{GL}_n(\mathrm{K})$ *est* $n!$.

Lorsque que le corps K *est* $\mathbb{R}$, *le cas le plus probable est celui où le plus grand coefficient non nul de la* k*-ième colonne de* A_{k-1} *est toujours sur la* $(n+1-k)$*-ième ligne. On obtient alors la permutation* $\tau : k \longmapsto n+1-k$. *On peut effectivement démontrer que l'orbite de* P_τ *est la plus grande :*

elle est ouverte et dense dans G *de sorte que les autres orbites sont d'intérieur vide*[2].

On termine le chapitre avec une série d'exercices sur les matrices à coefficients dans $\mathbb{Z}$. *Pour les lois usuelles,* $\mathcal{M}_n(\mathbb{Z})$ *est muni d'une structure d'anneau. On note* $\mathrm{GL}_n(\mathbb{Z})$ *le groupe des inversibles de cet anneau.*

• *Si* $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$, *il existe* $\mathrm{N} \in \mathcal{M}_n(\mathbb{Z})$ *telle que* $\mathrm{MN} = \mathrm{NM} = \mathrm{I}_n$. *En passant au déterminant, on a* $\det \mathrm{M} \det \mathrm{N} = 1$. *Donc* $\det \mathrm{M}$ *est inversible dans* $\mathbb{Z}$ *et, par conséquent,* $\det \mathrm{M} = \pm 1$.

• *Réciproquement, supposons que* $\det \mathrm{M} = \varepsilon = \pm 1$. *Notons* N *la transposée de la comatrice de* M. *C'est une matrice à coefficients entiers. L'égalité* $\mathrm{MN} = \mathrm{NM} = \det \mathrm{M}\, \mathrm{I}_n = \varepsilon \mathrm{I}_n$ *montre que la matrice* $\varepsilon \mathrm{N}$, *qui appartient à* $\mathcal{M}_n(\mathbb{Z})$, *est l'inverse de* M. *On conclut que*

$$\boxed{\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z}) \iff \det \mathrm{M} = \pm 1}.$$

7.19. Bases d'un groupe abélien

Étant donné un groupe abélien G, on dit que $(e_1, \ldots, e_n) \in \mathrm{G}^n$ est une base de G si tout élément x de G s'écrit de manière unique

$$x = k_1 e_1 + k_2 e_2 + \cdots + k_n e_n, \text{ où } (k_1, k_2, \ldots, k_n) \in \mathbb{Z}^n.$$

Soit G un groupe abélien, possédant une base $(e_1, \ldots, e_n)$.

1. Si $(\varepsilon_1, \ldots, \varepsilon_m)$ est une famille de m éléments de G, on écrit pour $1 \leqslant j \leqslant p$, $\varepsilon_j = \sum_{i=1}^{n} k_{ij} e_i$, avec les k_{ij} dans $\mathbb{Z}$.

Montrer que $(\varepsilon_1, \ldots, \varepsilon_m)$ est une base de G si, et seulement si, $m = n$ et la matrice $\mathrm{P} = (k_{ij})_{1 \leqslant i,j \leqslant n}$ est dans $\mathrm{GL}_n(\mathbb{Z})$.

2. On considère un sous-groupe G′ de G. Montrer que G′ admet une base de cardinal $r \leqslant n$.

(École polytechnique)

▷ **Solution.**

1. • Supposons que $(\varepsilon_1, \ldots, \varepsilon_m)$ est une base de G. Pour $1 \leqslant p \leqslant m$, on peut écrire $e_p = \sum_{i=1}^{m} \ell_{ip} \varepsilon_i$, où les ℓ_{ip} sont dans $\mathbb{Z}$. On a alors

2. Le lecteur intéressé pourra se reporter à MNEIMNÉ (R.), TESTARD (F.), *Groupes de Lie classiques*, Hermann, 1986, p. 48-49.

$$(*) \qquad e_p = \sum_{i=1}^{m} \ell_{ip}\Big(\sum_{j=1}^{n} k_{ji}e_j\Big) = \sum_{j=1}^{n}\Big(\sum_{i=1}^{m} k_{ji}\ell_{ip}\Big)e_j$$

et, par unicité de l'écriture, $\sum_{i=1}^{n} k_{ji}\ell_{ip} = \delta_{jp}$. En considérant les matrices $\mathrm{P} = (k_{ij})_{\substack{1\leqslant i\leqslant n\\ 1\leqslant j\leqslant m}} \in \mathcal{M}_{n,m}(\mathbb{Z})$ et $\mathrm{Q} = (\ell_{ij})_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}} \in \mathcal{M}_{m,n}(\mathbb{Z})$, on obtient $\mathrm{PQ} = \mathrm{I}_n$. On en déduit, en considérant les rangs des matrices en tant que matrices à coefficients dans $\mathbb{Q}$, que $n = \operatorname{rg}\mathrm{I}_n \leqslant \operatorname{rg}\mathrm{P} \leqslant m$. Les deux bases jouant des rôles symétriques, on a, de même, $n \geqslant m$ et donc $m = n$. Les matrices P et Q sont donc carrées de taille n et l'égalité $\mathrm{PQ} = \mathrm{I}_n$ implique que P appartient à $\mathrm{GL}_n(\mathbb{Z})$.

• Réciproquement, supposons que $m = n$ et que P est dans $\mathrm{GL}_n(\mathbb{Z})$; notons son inverse $\mathrm{Q} = (\ell_{ij})_{1\leqslant i,j\leqslant n}$. Le calcul $(*)$ du sens direct montre que $e_p = \sum_{i=1}^{n} \ell_{ip}\varepsilon_i$ pour tout $1 \leqslant p \leqslant n$. Il s'ensuit que les e_p sont dans le sous-groupe engendré par les ε_i. Puisque les e_p engendrent G, $(\varepsilon_1, \ldots, \varepsilon_n)$ est un système générateur de G. Il reste à montrer l'unicité de l'écriture de tout $x \in \mathrm{G}$ comme combinaison linéaire à coefficients entiers des ε_i. Clairement, il suffit de le vérifier pour 0. Soit $(k_1, \ldots, k_n) \in \mathbb{Z}^n$ tel que $k_1\varepsilon_1 + \cdots + k_n\varepsilon_n = 0$. On a alors

$$0 = \sum_{i=1}^{n} k_i\Big(\sum_{j=1}^{n} k_{ji}e_j\Big) = \sum_{j=1}^{n}\Big(\sum_{i=1}^{n} k_ik_{ji}\Big)e_j,$$

d'où pour tout $1 \leqslant i \leqslant n$, $\sum_{i=1}^{n} k_{ji}k_i = 0$ et comme ce système dont les inconnues sont les k_i, est de Cramer (la matrice des k_{ij} est inversible dans $\mathcal{M}_n(\mathbb{Q})$), les k_i sont tous nuls.

2. On suppose que $\mathrm{G}' \neq \{0\}$, sinon la famille vide convient, et on raisonne par récurrence sur n.

• Traitons le cas $n = 1$. L'application $\varphi : x \in \mathbb{Z} \longmapsto xe_1 \in \mathrm{G}$ est un isomorphisme de groupes, donc $\varphi^{-1}(\mathrm{G}')$ est un sous-groupe de $\mathbb{Z}$, qui s'écrit $a\mathbb{Z}$ avec $a > 0$. Tout élément de G' s'écrit de manière unique kae_1, avec $k \in \mathbb{Z}$ et (ae_1) est une base de G'.

• Supposons que $n \geqslant 2$ et que la propriété est vérifiée au rang $n-1$. Soit G' un sous-groupe de G, groupe qui possède une base de n vecteurs $(e_1, \ldots, e_n)$. Considérons $\mathrm{G}_1 = \sum_{i=1}^{n-1} \mathbb{Z}e_i$ et $\mathrm{G}'_1 = \mathrm{G}_1 \cap \mathrm{G}'$. Alors G'_1 est un sous-groupe de G_1, groupe qui possède une base de $n-1$ vecteurs. Par hypothèse de récurrence, G'_1 possède une base $(e'_1, \ldots, e'_{r-1})$ $(1 \leqslant r \leqslant n)$.

Si $\mathrm{G}'_1 = \mathrm{G}'$, c'est-à-dire si $\mathrm{G}' \subset \mathrm{G}_1$, la démonstration est achevée, sinon on considère le morphisme de groupes $\nu : \mathrm{G} \longmapsto \mathbb{Z}$ défini par $\nu\Big(\sum_{i=1}^{n} k_ie_i\Big) = k_n$ et on étudie $\nu(\mathrm{G}')$. C'est un sous-groupe de $\mathbb{Z}$,

non réduit à $\{0\}$, qui s'écrit donc $a\mathbb{Z}$, avec $a > 0$. Il existe $e'_r \in G'$ tel que $\nu(e'_r) = a$.

Montrons que $(e'_1, \dots, e'_r)$ est une base de G'.

$\star$ Supposons que $k_1 e'_1 + \cdots + k_r e'_r = 0$, avec les k_i dans $\mathbb{Z}$. On a alors

$$0 = \nu(0) = k_1 \nu(e'_1) + k_2 \nu(e'_2) + \cdots + k_r \nu(e'_r) = k_r a,$$

d'où l'on tire $k_r = 0$ et l'égalité $k_1 e'_1 + \cdots + k_{r-1} e'_{r-1} = 0$. Comme $(e'_1, \dots, e'_{r-1})$ est une base de G'_1, on a $k_1 = \cdots = k_{r-1} = 0$.

$\star$ Si $x \in G'$, alors $\nu(x)$ appartient à $a\mathbb{Z}$ et il existe $k_r \in \mathbb{Z}$ tel que $\nu(x) = k_r a = \nu(k_r e'_r)$. On en déduit que $x - k_r e'_r \in G'_1$: il existe $k_1, \dots, k_{r-1}$ dans $\mathbb{Z}$ tels que $x - k_r e'_r = k_1 e'_1 + \cdots k_{r-1} e'_{r-1}$. On obtient donc $x = k_1 e'_1 + \cdots + k_r e'_r$.

On conclut que $(e'_1, \dots, e'_r)$ est une base du sous-groupe G' de G, avec $r \leqslant n$. $\lhd$

Un groupe abélien possédant une base de cardinal fini est dit libre de type fini. *Le cardinal n d'une base quelconque est appelé* rang *du groupe. Un tel groupe est isomorphe à $\mathbb{Z}^n$. La question 1 de l'exercice montre, en particulier, que si $\mathbb{Z}^m$ est isomorphe à $\mathbb{Z}^n$, alors $m = n$. La question 2 démontre qu'un sous-groupe d'un groupe libre de type fini est libre de type fini. En particulier, tout sous-groupe de $\mathbb{Z}^n$ possède une base comportant moins de n vecteurs.*

On cherche dans l'exercice suivant à décrire les vecteurs-colonnes qui peuvent figurer dans une matrice de $\mathrm{GL}_n(\mathbb{Z})$. Quitte à échanger deux colonnes, on peut se limiter à la première. Au vu de l'exercice précédent, le problème s'interprète comme suit : à quelle condition un vecteur de $\mathbb{Z}^n$ peut-il être complété en une base de $\mathbb{Z}^n$ (au sens défini dans l'exercice précédent) ? Une autre solution de ce problème sera donnée dans le chapitre sur le déterminant du tome 2 d'algèbre.

7.20. Première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$

1. Soit $u = (u_1, \dots, u_n) \in \mathbb{Z}^n$ et

$$\begin{array}{ccc} \mathbb{Z}^n & \dashrightarrow & \mathbb{Z} \\ \tilde{u} : (x_1, \dots, x_n) & \longmapsto & \displaystyle\sum_{i=1}^{n} u_i x_i. \end{array}$$

On suppose qu'il existe $x \in \mathbb{Z}^n$ tel que $\tilde{u}(x) = 1$. Démontrer que $\mathbb{Z}^n = \mathbb{Z}x \oplus \mathrm{Ker}\, \tilde{u}$.

2. Prouver que $x \in \mathbb{Z}^n$ est la première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$ si, et seulement si, ses composantes sont des nombres premiers entre eux.

(École polytechnique)

▷ **Solution.**

1. Il faut préciser la notation $\oplus$ puisqu'il ne s'agit plus de la somme directe de sous-espaces vectoriels. Si G est un groupe abélien et si $\mathrm{H}_1, \ldots, \mathrm{H}_k$ des sous-groupes de G, on écrira $\mathrm{G} = \mathrm{H}_1 \oplus \cdots \oplus \mathrm{H}_k$ si tout élément g de G s'écrit de manière unique sous la forme $g = h_1 + \cdots + h_k$ avec $(h_1, \ldots, h_k) \in \mathrm{H}_1 \times \cdots \times \mathrm{H}_k$. On remarque qu'il suffit de vérifier l'unicité pour 0.

Soit donc $x \in \mathbb{Z}^n$ tel que $\tilde{u}(x) = 1$. Pour $y \in \mathbb{Z}^n$, considérons $k = \tilde{u}(y)$ et $z = y - kx$. Comme $\tilde{u}$ est visiblement un morphisme du groupe $\mathbb{Z}^n$ dans $\mathbb{Z}$, on obtient

$$\tilde{u}(z) = \tilde{u}(y) - k\tilde{u}(x) = k - k = 0 \quad \text{et} \quad z \in \operatorname{Ker} \tilde{u}.$$

Par conséquent, $y = kx + z \in \mathbb{Z}x + \operatorname{Ker} \tilde{u}$.

Si $kx + z = 0$ avec $z \in \operatorname{Ker} \tilde{u}$ et $k \in \mathbb{Z}$, on obtient, en prenant l'image par $\tilde{u}$,

$$0 = k\tilde{u}(x) + \tilde{u}(z) = k + 0 = k \quad \text{et} \quad z = 0.$$

On conclut que

$$\boxed{\mathbb{Z}^n = \mathbb{Z}x \oplus \operatorname{Ker} \tilde{u}}.$$

2. • La condition proposée est nécessaire. En effet, soit $\mathrm{M} \in \mathrm{GL}(\mathbb{Z})$ et $(\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_n)$ ses colonnes. Soit d le pgcd des coefficients de la première colonne. On a

$$\det \mathrm{M} = \pm 1 = \det(\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_n) = d \det \Big(\underbrace{\frac{1}{d}\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_n}_{\in \mathcal{M}_n(\mathbb{Z})} \Big) \in d\mathbb{Z},$$

donc d est nécessairement égal à 1. Les composantes de C_1 sont donc des nombres premiers entre eux.

• Inversement, soit $x = (x_1, \ldots, x_n) \in \mathbb{Z}^n$ dont les composantes sont des nombres premiers eux. D'après l'identité de Bezout, il existe $(u_1, \ldots, u_n) \in \mathbb{Z}^n$ tel que

$$1 = u_1x_1 + \cdots + u_nx_n.$$

Notons $\tilde{u}$ l'application

$$(y_1, \ldots, y_n) \in \mathbb{Z}^n \longmapsto \sum_{i=1}^{n} u_i y_i \in \mathbb{Z}.$$

Par construction, on a $\tilde{u}(x) = 1$. De la question précédente, on déduit donc $\mathbb{Z}^n = \mathbb{Z}x \oplus \operatorname{Ker} \tilde{u}$. Puisque $\operatorname{Ker} \tilde{u}$ est un sous-groupe de $\mathbb{Z}^n$, il existe, d'après l'exercice précédent, $\varepsilon_2, \ldots, \varepsilon_r$ dans $\mathbb{Z}^n$, tels que

$$\operatorname{Ker} \tilde{u} = \mathbb{Z}\varepsilon_2 \oplus \cdots \oplus \mathbb{Z}\varepsilon_r \quad \text{et donc} \quad \mathbb{Z}^n = \mathbb{Z}x \oplus \mathbb{Z}\varepsilon_2 \oplus \cdots \oplus \mathbb{Z}\varepsilon_r$$

On vérifie alors facilement que $(x, \varepsilon_2, \ldots, \varepsilon_r)$ est une base de $\mathbb{Z}^n$. On a donc forcément $r = n$ et la matrice de cette famille dans la base canonique de $\mathbb{Z}^n$ est dans $\mathrm{GL}_n(\mathbb{Z})$. Sa première colonne contenant le vecteur x, on a le résultat souhaité. $\lhd$

Intéressons-nous à la notion de matrices équivalentes dans $\mathcal{M}_n(\mathbb{Z})$. Comme dans $\mathcal{M}_n(\mathrm{K})$, deux matrices A et B de $\mathcal{M}_n(\mathbb{Z})$ sont dites équivalentes s'il existe deux matrices P et Q de $\mathrm{GL}_n(\mathbb{Z})$ telles que $\mathrm{B} = \mathrm{PAQ}$. C'est une relation d'équivalence. Les classes d'équivalence dans $\mathcal{M}_n(\mathbb{Z})$ ne sont plus paramétrées par un seul entier (le rang), comme dans $\mathcal{M}_n(\mathrm{K})$, mais par une suite finie d'entiers. L'outil essentiel de l'exercice suivant est l'algorithme du pivot de Gauss.

7.21. Matrices équivalentes dans $\mathcal{M}_n(\mathbb{Z})$

On note G le sous-groupe de $\mathrm{GL}_n(\mathbb{Z})$ engendré par les matrices de la forme $\mathrm{I} - 2\mathrm{E}_{ii}$ et $\mathrm{I} + a\mathrm{E}_{ij}$ pour tout $1 \leqslant i, j \leqslant n$, $i \neq j$ et $a \in \mathbb{Z}$.

On envisage la relation $\mathcal{R}$ dans $\mathcal{M}_n(\mathbb{Z})$, définie pour A et B dans $\mathcal{M}_n(\mathbb{Z})$ par

$$\mathrm{A}\mathcal{R}\mathrm{B} \iff \exists (\mathrm{P}, \mathrm{Q}) \in \mathrm{G}^2,\ \mathrm{A} = \mathrm{PBQ}.$$

1. Montrer que $\mathcal{R}$ est une relation d'équivalence.

Pour $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{Z})$ non nulle, on note

$$d(\mathrm{A}) = \min\{|a_{ij}|, 1 \leqslant i, j \leqslant n,\ a_{ij} \neq 0\}.$$

2. Montrer qu'il existe dans la classe d'équivalence de A une matrice $\mathrm{N} = (n_{ij})_{1 \leqslant i,j \leqslant n}$ telle que $d(\mathrm{N})$ soit minimal dans cette classe et $d(\mathrm{N}) = n_{11}$.

Montrer que n_{11} divise les n_{1j}, puis qu'il divise les n_{ij}.

3. Montrer que l'on peut trouver N de la forme

$$\mathrm{N} = \mathrm{Diag}(n_{11}, k_1 n_{11}, k_1 k_2 n_{11}, \ldots, k_1 \cdots k_{n-1} n_{11}),$$

où les k_i sont dans $\mathbb{N}$.

4. Déduire de ce qui précède que $\mathrm{G} = \mathrm{GL}_n(\mathbb{Z})$.

(École polytechnique)

▷ **Solution.**

1. Soit A, B et C dans $\mathcal{M}_n(\mathbb{Z})$.

Comme $\mathrm{I} \in \mathrm{G}$, car G est un sous-groupe, et que $\mathrm{A} = \mathrm{IAI}$, on a $\mathrm{A}\mathcal{R}\mathrm{A}$: la relation $\mathcal{R}$ est réflexive.

Supposons $\mathrm{A}\mathcal{R}\mathrm{B}$. Il existe alors P et Q dans G tels que $\mathrm{A} = \mathrm{PBQ}$.

Mais alors, on a $B = P^{-1}AQ^{-1}$, où P^{-1} et Q^{-1} sont dans G : $\mathcal{R}$ est symétrique.

Supposons $A\mathcal{R}B$ et $B\mathcal{R}C$. Il existe donc P, Q, R et S dans G tels que $A = PBQ$ et $B = RCS$. On en déduit que $A = (PR)C(SQ)$ et $A\mathcal{R}C$, puisque PR et SQ sont dans G. La relation est transitive. C'est donc une relation d'équivalence.

2. • Notons $T_{ij}(a) = I + aE_{ij}$ pour $1 \leqslant i \neq j \leqslant n$ et $a \in \mathbb{Z}$. Il s'agit là d'une matrice de transvection. Rappelons rapidement l'effet de la multiplication par $T_{ij}(a)$: si $M \in \mathcal{M}_n(\mathbb{Z})$ est de colonnes $(C_1, \ldots, C_n)$ et de lignes $(L_1, \ldots, L_n)$, la matrice $T_{ij}(a)M$ est la transformée de M par l'opération élémentaire $L_i \leftarrow L_i + aL_j$. De même, la matrice $MT_{ij}(a)$ est la transformée de M par l'opération élémentaire $C_j \leftarrow C_j + aC_i$. La multiplication à gauche par $I - 2E_{ii}$ correspond à $L_i \leftarrow -L_i$, et la multiplication à droite à $C_i \leftarrow -C_i$.

Il est possible de trouver dans G une matrice réalisant la permutation de deux lignes ou deux colonnes. En effet, le lecteur vérifiera aisément que la matrice $(I - 2E_{ii})T_{ij}(-1)T_{ji}(1)T_{ij}(-1) \in G$ $(1 \leqslant i, j \leqslant n, i \neq j)$ effectue, par multiplication à gauche, l'opération élémentaire $L_i \leftrightarrow L_j$. De même, $T_{ji}(-1)T_{ij}(1)T_{ij}(-1)(I - 2E_{ii}) \in G$ effectue par multiplication à droite l'opération $C_i \leftrightarrow C_j$. Comme les transpositions engendrent le groupe symétrique, G contient donc toutes les matrices de permutation.

Il en résulte que dans la classe d'équivalence de $A \in GL_n(\mathbb{Z})$, on trouve toutes les matrices obtenues à partir de A par une succession d'opérations élémentaires sur les lignes et les colonnes telles que nous venons de les décrire.

• On remarque que si A est en relation avec B, alors $B \neq 0$ et $d(B)$ est donc défini. L'ensemble des $d(B)$, pour B décrivant la classe d'équivalence de A, est une partie non vide de $\mathbb{N}^*$. Elle admet donc un plus petit élément. Prenons $N = (n_{ij})_{1 \leqslant i,j \leqslant n}$ dans la classe de A telle que $d(N)$ soit ce plus petit élément.

Soit $(i_0, j_0) \in [\![1, n]\!]$ tel que $|n_{i_0 j_0}| = d(N)$. Quitte à effectuer l'opération élémentaire $L_{i_0} \leftrightarrow L_1$, on peut supposer $i_0 = 1$: on reste bien dans la classe de A et $d(N)$ est inchangé. Quitte à effectuer l'opération $C_{j_0} \leftarrow C_1$, on peut également supposer que $j_0 = 1$. Enfin, quitte à effectuer l'opération $C_1 \leftarrow -C_1$, on peut supposer $n_{11} > 0$. On a donc trouvé N dans la classe de A telle que $d(\mathbb{N})$ soit minimal avec $d(N) = n_{11}$.

• Soit $2 \leqslant j \leqslant n$. Opérons la division euclidienne de n_{1j} par n_{11}. Il existe $q \in \mathbb{Z}$ et $0 \leqslant r \leqslant n_{11} - 1$ tels que $n_{1j} = n_{11}q + r$. Effectuons l'opération élémentaire, $C_j \leftarrow C_j - qC_1$. Le coefficient d'indice $(1, j)$ de la matrice obtenue, toujours équivalente à A, est r. Comme $r < d(\mathbb{N})$ la minimalité de $d(N)$ impose $r = 0$. Donc $n_{11} | n_{1j}$.

On montre de même en agissant sur les lignes que n_{11} divise les n_{i1}.

• En effectuant pour chaque $2 \leqslant j \leqslant n$, l'opération $C_j - q_j C_1$ où q_j est le quotient de n_{1j} par n_{11}, on trouve dans la classe de A une nouvelle matrice $N' = (n'_{ij})_{1 \leqslant i,j \leqslant n}$, avec $n'_{11} = n_{11}$, les $n'_{1j} = 0$, $n'_{i1} = n_{i1}$ et $n'_{ij} = n_{ij} - q_j n_{i1} \equiv n_{ij} \pmod{n_{11}}$ pour $2 \leqslant i, j \leqslant n$.

Soit $2 \leqslant i, j \leqslant n$. On effectue sur la matrice N′ l'opération élémentaire $C_1 \leftarrow C_1 + C_j$. On obtient alors une nouvelle matrice $N'' = (n''_{ij})_{1 \leqslant i,j \leqslant n}$ dans la classe de A, avec toujours $n''_{11} = n_{11} = d(N)$.

Ce qui était valable pour N le reste pour N″ : n_{11} divise encore les $n''_{i1} = n_{i1} + n'_{ij} \equiv n'_{ij} \equiv n_{ij} \pmod{n_{11}}$. Autrement dit, n_{11} divise n_{ij}.

3. À partir de la matrice N construite à la question 2, on construit une matrice équivalente telle que $n_{i1} = 0$ si $i \geqslant 2$ et $n_{1j} = 0$ si $j \geqslant 2$. Pour cela, on effectue, pour $i \geqslant 2$, les opérations élémentaires $L_i \leftarrow L_i - \dfrac{n_{i1}}{n_{11}} L_1$, ce qui amène des 0 sur les termes de la première colonne, sauf le premier, puis, pour $j \geqslant 2$, les opérations élémentaires $C_j \leftarrow C_j - \dfrac{n_{j1}}{n_{11}} C_1$, ce qui amène des 0 sur les termes de la première ligne, sauf le premier (sans toucher à la première colonne). Ceci est possible, car tous les termes de la matrice sont divisibles par n_{11} et le restent après chacune de ces transformations. D'après ce qui précède, on peut donc trouver dans la classe de A une matrice que nous noterons encore N du type

$$N = \begin{pmatrix} n_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & n_{11}A' & \\ 0 & & & \end{pmatrix}.$$

Il est naturel de procéder par récurrence sur n, pour démontrer le résultat demandé. La propriété est évidente si $n = 1$. On suppose qu'elle est vérifiée au rang $n-1$ et on la démontre pour une matrice de taille n.

• Si $A' = 0$, on prend simplement $k_i = 0$ pour tout $1 \leqslant i \leqslant n-1$.

• Si $A' \neq 0$, d'après l'hypothèse de récurrence, il existe des matrices P et Q de taille $n-1$ qui sont produits de matrices de taille $n-1$ du type $I - 2E_{ii}$, $T_{ij}(a)$ ou de leur inverse et $k_1, \dots, k_n$ dans N tels que $PA'Q = N'$ où $N' = \mathrm{Diag}(k_1, k_1k_2, \dots, k_1 \dots k_{n-1})$. Comme on a $(I - 2E_{ii})^{-1} = I - 2E_{ii}$ et $T_{ij}(a)^{-1} = T_{ij}(-a)$, la matrice P est produit de matrices de taille $n-1$ du type $I - 2E_{ii}$ ou $T_{ij}(a)$. Si $M = I - 2E_{ii}$ ou $M = T_{ij}(a)$, la matrice

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & M & \\ 0 & & & \end{pmatrix}$$

est une matrice de taille n faisant partie des générateurs de G décrit par l'énoncé. Il s'ensuit que, d'après la règle de multiplication des matrices

par blocs, la matrice $P' = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \boxed{P} & \\ 0 & & & \end{pmatrix}$ est dans G. De même, $Q' = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \boxed{Q} & \\ 0 & & & \end{pmatrix}$ est dans G. Il en résulte que $P'NQ'$ est encore dans la classe de A. Ainsi, on a

$$P'NQ' = \begin{pmatrix} n_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \boxed{n_{11}PA'Q} & \\ 0 & & & \end{pmatrix} = \begin{pmatrix} n_{11} & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \boxed{n_{11}N'} & \\ 0 & & & \end{pmatrix}$$
$$= \mathrm{Diag}(n_{11}, k_1 n_{11}, k_1 k_2 n_{11}, \dots, k_1 \dots k_{n-1} n_{11}).$$

C'est le résultat recherché.

4. On garde les notations de la question précédente. Si on prend $A \in \mathrm{GL}_n(\mathbb{Z})$, alors $N \in \mathrm{GL}_n(\mathbb{Z})$ et $\det N = \pm 1$. Les coefficients n_{11} et les k_i sont forcément égaux à 1. La matrice A est donc équivalente à I_n et $A \in G$. On conclut que $\boxed{G = \mathrm{GL}_n(\mathbb{Z})}$. ◁

Puisque G *est égal à* $\mathrm{GL}_n(\mathbb{Z})$, *la relation* $\mathcal{R}$ *coïncide avec la notion de matrices équivalentes dans* $\mathcal{M}_n(\mathbb{Z})$. *La question 2 de l'exercice démontre que toute matrice est équivalente à une matrice de la forme* $\mathrm{Diag}(d_1, d_2, \dots, d_n)$, *où* $d_1, \dots, d_n$ *sont des entiers naturels tels que* $d_1 | d_2 | \cdots | d_n$. *Signalons qu'on a le résultat d'unicité suivant : soit* $(d_i)_{1 \leqslant i \leqslant n}$, $(d'_i)_{1 \leqslant i \leqslant n}$ *deux suites d'entiers naturels telles que* $d_1 | d_2 | \cdots | d_n$ *et* $d'_1 | d'_2 | \cdots | d'_n$. *Si les matrices* $\mathrm{Diag}(d_1, d_2, \dots, d_n)$ *et* $\mathrm{Diag}(d'_1, \dots, d'_n)$ *sont équivalentes, alors* $d_i = d'_i$ *pour tout* i. *Autrement dit, une classe d'équivalence est paramétrée par une suite de* n *entiers naturels* $(d_i)_{1 \leqslant i \leqslant n}$ *telle que* $d_1 | d_2 | \cdots | d_n$.

Le lecteur est invité à regarder comment ce résultat se généralise aux matrices à coefficients dans un anneau principal quelconque.

Pour terminer, on peut noter que l'exercice a aussi prouvé que l'ensemble formé des matrices de transvections $T_{ij}(a)$ *à coefficients entiers et des matrices* $I - 2E_{ii}$, *qui sont des matrices de symétries hyperplanes, engendre le groupe* $\mathrm{GL}_n(\mathbb{Z})$. *On se reportera au chapitre sur le groupe linéaire, dans le tome 2 d'algèbre, pour des compléments et des utilisations de ce résultat.*

On peut donner de l'exercice précédent une belle application : la classification des groupes abéliens finis. Cet exercice complète l'exercice 2.20

du chapitre sur les groupes, qui étudiait l'unicité de la décomposition en produit de groupes cycliques. On suppose connus les définitions et résultats de l'exercice 7.19.

7.22. Structure des groupes abéliens finis

On rappelle le résultat établi à l'exercice précédent : pour toute matrice A de $\mathcal{M}_n(\mathbb{Z})$, il existe P et Q dans $\mathrm{GL}_n(\mathbb{Z})$ et un n-uplet $(d_1,\dots,d_n)\in\mathbb{N}^n$ tels que

$$d_1|d_2|\cdots|d_n \quad\text{et}\quad \mathrm{PAQ}=\mathrm{Diag}(d_1,d_2,\dots,d_n).$$

1. Soit G un sous-groupe de $\mathbb{Z}^n$. Montrer qu'il existe une base $(\varepsilon_1,\dots,\varepsilon_n)$ de $\mathbb{Z}^n$, un entier $r\geqslant 0$ et $(d_1,\dots,d_r)\in(\mathbb{N}^*)^r$, vérifiant $d_1|d_2|\cdots|d_r$ et tels que $(d_1\varepsilon_1,d_2\varepsilon_2,\dots,d_r\varepsilon_r)$ soit une base de G.

2. On considère un groupe abélien fini H. Montrer qu'il existe un entier $r\geqslant 0$ et des entiers $d_1,\dots,d_r$, supérieurs à 2, avec $d_1|d_2|\cdots|d_r$ tels que

$$\mathrm{H}\simeq\mathbb{Z}/d_1\mathbb{Z}\times\mathbb{Z}/d_2\mathbb{Z}\times\cdots\times\mathbb{Z}/d_r\mathbb{Z}.$$

▷ **Solution.**

1. D'après l'exercice 7.19, il existe $r\geqslant 0$ et une base $(e_1,\dots,e_r)$ de G. Considérons la matrice A de $\mathcal{M}_n(\mathbb{Z})$ dont les vecteurs colonnes sont $e_1,\dots,e_r,0,\dots,0$. Le rang de A, vue comme matrice à coefficients rationnels, est r, car ses r premières colonnes sont $\mathbb{Z}$-libres et donc $\mathbb{Q}$-libres. D'après l'exercice précédent, il existe $(d_1,\dots,d_n)\in\mathbb{N}^n$, vérifiant $d_1|d_2|\cdots|d_n$ et P et Q dans $\mathrm{GL}_n(\mathbb{Z})$ tels que

$$\mathrm{PAQ}=\mathrm{Diag}(d_1,d_2,\dots,d_n)=\mathrm{D}.$$

Le rang de la matrice PAQ est r, comme celui de A. La matrice D a donc r termes non nuls sur sa diagonale. La condition $d_1|d_2|\cdots|d_r$ entraîne que $d_i\geqslant 1$ si $i\leqslant r$ et $d_{r+1}=0$ dans le cas où $r<n$.

Considérons les vecteurs-colonnes $\varepsilon_1,\dots,\varepsilon_n$ de P^{-1}. La matrice de passage de la base canonique $\mathcal{B}_0$ de $\mathbb{Z}^n$ (qui est celle de $\mathbb{Q}^n$) à $(\varepsilon_1,\dots,\varepsilon_n)$ est P^{-1}, qui appartient à $\mathrm{GL}_n(\mathbb{Z})$. Le système $(\varepsilon_1,\dots,\varepsilon_n)$ est donc une base de $\mathbb{Z}^n$, toujours d'après l'exercice 7.19. La matrice du système $(d_1\varepsilon_1,\dots,d_r\varepsilon_r,0,\dots,0)$, dans la base $\mathcal{B}_0$, est alors $\mathrm{P}^{-1}\mathrm{D}$. Notons $(\mathrm{C}_1,\dots,\mathrm{C}_n)$ les vecteurs-colonnes de $\mathrm{P}^{-1}\mathrm{D}$. On a ainsi

$$\mathrm{C}_1=d_1\varepsilon_1,\ \dots,\mathrm{C}_r=d_r\varepsilon_r,\ \mathrm{C}_{r+1}=\cdots=\mathrm{C}_n=0.$$

L'égalité $\mathrm{P}^{-1}\mathrm{D}=\mathrm{AQ}$ entraîne que $\mathrm{C}_1,\dots,\mathrm{C}_n$ sont des combinaisons linéaires à coefficients entiers de $e_1,\dots,e_r$ et appartiennent donc à G. Le

système $(C_1, \dots, C_r)$ est $\mathbb{Z}$-libre, car $\mathbb{Q}$-libre, puisque $\operatorname{rg} AQ = \operatorname{rg} A = r$. Enfin, on conclut de l'égalité $A = (C_1|\cdots|C_r|0|\cdots|0)Q^{-1}$, que les colonnes de A, c'est-à-dire les vecteurs $e_1, \dots, e_r$, sont combinaisons linéaires à coefficients entiers de $C_1, \dots, C_r$. Il en résulte que $C_1, \dots, C_r$ engendrent G tout entier. Le système $(C_i)_{1 \leqslant i \leqslant r} = (\delta_i \varepsilon_i)_{1 \leqslant i \leqslant r}$ constitue donc une base du groupe G.

Conclusion. On a trouvé une base $(\varepsilon_1, \dots, \varepsilon_n)$ de $\mathbb{Z}^n$, des entiers $d_i \geqslant 1$ tels que $d_1|d_2|\cdots|d_r$ et $(d_1\varepsilon_1, d_2\varepsilon_2, \dots, d_r\varepsilon_r)$ soit une base de G.

2. On note $x_1, \dots, x_n$ les éléments de H et on considère le morphisme de groupes

$$f : (k_1, \dots, k_n) \in \mathbb{Z}^n \longmapsto k_1 x_1 + \cdots + k_n x_n \in \mathrm{H}.$$

L'application f est surjective et d'après le théorème d'isomorphisme, on a $\mathbb{Z}^n / \operatorname{Ker} f \simeq \mathrm{H}$.

D'après la question précédente, il existe une base $(\varepsilon_1, \dots, \varepsilon_n)$ de $\mathbb{Z}$, des entiers non nuls $d_1, \dots, d_r$ tels que $d_1|d_2|\cdots|d_r$ et $(d_1\varepsilon_1, \dots, d_r\varepsilon_r)$ est une base de $\operatorname{Ker} f$. Considérons l'application φ qui, à $X \in \mathbb{Z}^n$, associe le système de coordonnées de X dans la base $(\varepsilon_1, \dots, \varepsilon_n)$. Il s'agit d'un automorphisme du groupe $\mathbb{Z}^n$, pour lequel on a $\varphi(\mathbb{Z}^n) = \mathbb{Z}^n$ et $\varphi(\operatorname{Ker} f) = d_1\mathbb{Z} \times \cdots \times d_r\mathbb{Z} \times \{0\}^{n-r}$.

Le lecteur se convaincra facilement que dans ces conditions, on obtient

$$\mathrm{H} \simeq \mathbb{Z}^n / \operatorname{Ker} f \simeq \varphi(\mathbb{Z}^n)/\varphi(\operatorname{Ker} f) = \mathbb{Z}^n / \left(d_1\mathbb{Z} \times \cdots \times d_r\mathbb{Z} \times \{0\}^{n-r}\right)$$

et donc

$$\mathrm{H} \simeq (\mathbb{Z}/d_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_r\mathbb{Z}) \times \mathbb{Z} \times \cdots \times \mathbb{Z}.$$

Comme H est fini, on a nécessairement $r = n$ et

$$\mathrm{H} \simeq (\mathbb{Z}/d_1\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_n\mathbb{Z}).$$

En notant s le plus petit entier tel que $d_s \geqslant 2$ (qui est défini dès que H n'est pas réduit à l'élément neutre), il vient

$$\boxed{\mathrm{H} \simeq (\mathbb{Z}/d_s\mathbb{Z}) \times \cdots \times (\mathbb{Z}/d_n\mathbb{Z})}$$

et la suite $(d_s, d_{s+1}, \dots, d_n)$ a les propriétés voulues. ◁

On se reportera à l'exercice 2.20 du chapitre 2 (Groupes) pour une démonstration de l'unicité des d_i.

Le résultat suivant tient plus de l'arithmétique que de la théorie des matrices. Il démontre que dans la résolution d'un système linéaire de m équations à n inconnues, avec $n > m$, et dont les coefficients sont entiers, on peut toujours trouver des solutions «pas trop grandes».

7.23. Lemme de Siegel

Soit n et m dans $\mathbb{N}^*$, avec $n > m$. $A = (a_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}} \in \mathcal{M}_{m,n}(\mathbb{Z})$. On suppose qu'il existe une entier $\alpha > 0$ tel que $|a_{i,j}| \leqslant \alpha$, pour tout $(i,j) \in [\![1,m]\!] \times [\![1,n]\!]$.

Montrer qu'il existe $X \subset \mathbb{Z}^n$, $X \neq 0$, tel que $AX = 0$ et tel que, si $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$, on ait, pour tout $j \in [\![1,n]\!]$

$$|x_j| \leqslant (n\alpha)^{\frac{m}{n-m}} + 1.$$

(École normale supérieure)

▷ **Solution.**

• Remarquons, pour commencer, que le système linéaire $AX = 0$ admet des solutions non nulles dans $\mathbb{Q}^n$, puisqu'il a m équations et n inconnues avec $m < n$. Si $X \in \mathbb{Q}^n$ est une solution non nulle, en multipliant X par un entier convenable (par exemple le produit des dénominateurs des coordonnées de X), on obtient une solution dans $\mathbb{Z}^n$. La question consiste à trouver une solution entière «assez petite».

• Si on trouve deux vecteurs distincts de $\mathbb{Z}^n$, X et X′ qui vérifient $AX = AX'$, on obtient un vecteur du noyau de A, à savoir $X - X'$. L'idée de la démonstration est la suivante : choisir des vecteurs de $\mathbb{Z}^n$ dans une partie bornée de $\mathbb{R}^n$ et en choisir suffisamment pour être sûr d'en trouver – par une application du principe des tiroirs de Dirichlet – deux qui ont la même image par A. On notera $\|\ \|$ la norme infinie, que ce soit dans $\mathbb{R}^n$ ou dans $\mathbb{R}^m$.

• On va commencer par estimer $\|AX\|$ en fonction de $\|X\|$. On note

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \quad \text{et} \quad Y = AX = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}.$$

On a pour tout $i \in [\![1,m]\!]$, $|y_j| = \left|\sum_{j=1}^{n} a_{ij}x_j\right| \leqslant \sum_{j=1}^{n} |a_{ij}||x_j| \leqslant n\alpha\|X\|$ et donc

$$\|AX\| \leqslant n\alpha\|X\|.$$

• Soit $M \in \mathbb{N}$. On note B_M la boule fermée de $\mathbb{R}^n$, centrée en 0, de rayon M pour la norme $\|\ \|$. Un vecteur $X \in \mathbb{Z}^n$ est dans B_M si et

seulement si, pour tout $1 \leqslant i \leqslant n$, $|x_i| \leqslant \mathrm{M}$. Pour chaque coordonnée, on a exactement $2\mathrm{M}+1$ possibilités de sorte que

$$\mathrm{Card}\,(\mathrm{B}_\mathrm{M} \cap \mathbb{Z}^n) = (2\mathrm{M}+1)^n.$$

D'après le point précédent, l'image par A de $\mathrm{B}_\mathrm{M} \cap \mathbb{Z}^n$ est incluse dans l'intersection de $\mathbb{Z}^m$ et de la boule fermée de $\mathbb{R}^m$ de rayon $n\alpha\mathrm{M}$. Cet ensemble est de cardinal $(1+2n\alpha\mathrm{M})^m$.

Par le lemme des tiroirs, on est donc assuré de trouver deux vecteurs distincts X et X′ de $\mathrm{B}_\mathrm{M} \cap \mathbb{Z}^n$, tels que $\mathrm{AX} = \mathrm{AX}'$, dès lors que

$$(2n\alpha\mathrm{M}+1)^m < (2\mathrm{M}+1)^n.$$

Pour cela, il suffit que $(2\mathrm{M}+1)^n \geqslant (n\alpha)^m(2\mathrm{M}+1)^m$, ce qui équivaut à

$$2\mathrm{M} \geqslant (n\alpha)^{\frac{m}{n-m}} - 1.$$

Pour M ainsi choisi, la solution $\mathrm{X}-\mathrm{X}'$ du système vérifie, par inégalité triangulaire, $\|\mathrm{X}-\mathrm{X}'\| \leqslant 2\mathrm{M}$. Il nous suffit donc de choisir M tel que

$$2\mathrm{M} \in \left[(n\alpha)^{\frac{m}{n-m}} - 1, (n\alpha)^{\frac{m}{n-m}} + 1\right]$$

pour conclure (c'est possible car un segment de longueur 2 contient toujours un entier pair). ◁

Siegel a utilisé ce résultat pour prouver la transcendance de certains nombres réels.

Index

A

algébrique (nombre), 108
algèbre, 293–298, 307–317
  simple, 294
anneau
  commutatif, 84–85
  des entiers de Gauss, 98–103
  des nombres décimaux, 90
  euclidien, 94, 98
  factoriel, 91, 255
  principal, 90–98, 255
  régulier, 85
  $\mathbb{Z}[X]$, 90
approximation par des rationnels, 247
arithmétique (fonction), 150–155
automorphismes
  de $K[X]$, 244
  de $\mathcal{L}(E)$, 293
  de $\mathbb{Q}(\sqrt{2})$, 106
  de S_n, 74
  intérieurs, 48, 74

B

base (d'un groupe abélien), 350
Beatty (théorème de), 33
Bell (nombres de), 14
Bernstein (inegalité de), 225, 241
Bezout (théorème de), 124, 255
Bohl-Sierpinski-Weyl (théorème de), 31
Bruhat (décomposition de), 347
Burnside (formule de), 289

C

carrés
  somme de deux, 100, 158–161
    dans $\mathbb{R}[X]$, 194
  somme de quatre, 162
  somme de trois, 164
Cauchy (lemme de), 46
centre d'un p-groupe, 47
commutant, 282
commutateur, 58
congruence, 144, 145
conjugaison (classe de), 47, 49, 53
contenu (d'un polynôme), 188
corps, 43, 106–115, 129
  fini, 16–20
  cyclotomique, 190
crible (formule du), 9, 156
cyclique
  groupe, 41, 43, 64, 129
cyclotomique
  corps, 190
  polynôme, 135, 190

D

Davenport-Cassels (lemme de), 164
dérangements, 7–14
dérivation, 311–315
Descartes (règle de), 237
diédral (groupe), 57
diophantienne (équation), 165
Dirichlet (théorème de la progression arithmétique de), 135
discriminant, 183
distingué (sous-groupe), 52
diviseur (d'un polynôme), 197
diviseur de zéro, 82
divisibilité, 120, 174, 175
drapeau, 263, 347
dualité, 298–302
  dans $\mathcal{M}_n(K)$, 329–331, 335
Ducci (problème de), 147

E

Eisenstein (critère de), 188–191
endomorphisme
  cyclique, 283
  nilpotent, 270–271
Eneström-Kakeya (théorème de), 233
équation aux classes, 46, 52, 54
Erdös, 153
espace vectoriel
  quotient, 260, 267, 269, 273
  sur un corps fini, 17–20
euclidien (anneau), 94, 98
Euler (indicatrice d'), 130, 135, 150
exposant d'un groupe abélien, 43
extension quadratique, 104

F

factoriel (anneau), 91, 255
factorisation, 266
  bilatère, 321
Fermat
  grand théorème, 166
  nombres de, 132
  petit théorème, 126–129
Fibonacci (suite de), 6
Fitting (décomposition de), 277

fonction
arithmétique, 150–155
de Möbius, 150, 156
multiplicative, 103, 153
polynomiale, 193–198, 200, 202–206
somme des diviseurs, 151–153
symétrique des racines, 206–214
zêta, 207
fractions rationnelles, 240–244
Frobenius (théorème de), 52
Frobenius-König (théorème de), 345

G

Gauss
anneau des entiers de, 98–103
lemme de, 256
pivot de, 347, 354
théorème de, 31
Gauss-Lucas (théorème de), 229, 232
générateurs
de $\mathcal{S}_n$, 67
du groupe modulaire, 60
groupe
multiplicatif d'un corps, 43, 129
abélien fini (théorème de structure), 64, 358
cyclique, 41, 43, 64, 129
dérivé, 57
des rotations, 54
diédral, 57
libre de type fini, 352
modulaire, 60
opérant sur un ensemble, 45–47, 289, 339
ordonné, 78
quotient, 39–41, 64
symétrique, 46, 66–77

H

Hilbert
matrice de, 242
polynômes, 198, 205
homothétie vectorielle, 268–270
Howie (théorème de), 27

I

idéal
de $\mathcal{L}(E)$, 295–298
maximal, 88
premier, 106
principal, 87
indicatrice d'Euler, 130, 135, 150
indice
de nilpotence, 270, 311
d'un sous-groupe, 46, 47, 50
inégalité
de Bernstein, 225
de Landau, 234
de Markov, 225
de Schur, 225
de Sylvester, 272
interpolation, 199–202
inversibilité, 82
involutions, 19, 125
irréductible
d'un anneau, 91, 98–103
fraction, 118
polynôme de $\mathbb{Z}[X]$, 188–191

K

Kronecker (théorème de), 213
Kurschak (théorème de), 138

L

Lagrange (interpolation de), 199–202
Laguerre (théorème de), 217
Landau (inégalité de), 234
Legendre (théorème de), 139
Lie (crochet de), 336
Liouville (théorème de), 179, 180
Lucas
congruences de, 144
théorème de, 144

M

Markov (inégalité de), 225
Maschke (théorème de), 292
Mason (théorème de), 180
matrice
de Cauchy, 242
de dilatation, 347, 354
de Hilbert, 242
de transvection, 347, 354
équivalente, 320–323
dans $\mathcal{M}_n(\mathbb{Z})$, 354
inversible dans $\mathcal{M}_n(\mathbb{Z})$, 352
monotone, 341
positive, 341
stochastique, 343
maximal (idéal), 88
Möbius (fonction de), 150, 156
modulaire (groupe), 60
monotone (matrice), 341
morphisme
de conjugaison, 48
de groupes, 38, 39, 69, 70

N

Newton (formules de), 209
nilpotent (endomorphisme), 270–271
nombre algébrique sur un corps, 108

P

Palfy-Erdös (théorème de), 140
partitions, 20
pivot de Gauss, 347, 354
point visible, 122
polynôme
  cyclotomique, 190
  de Hilbert, 198, 205
  de Tchebychev, 222–228
  interpolateur, 199–202
  positif, 194–197
  primitif, 188
  scindé, 211, 215–224
positive (matrice), 341
positivement génératrice (famille), 302–306
premier
  idéal, 106
  nombre, 130, 134, 137
principal
  anneau, 90–255
  idéal, 87
produits d'entiers consécutifs, 123
projecteur, 284–298
Prüfer (groupes de), 59
pseudo-inverse, 274

Q

quotient
  espace vectoriel, 260, 267, 269, 273
  groupe, 39–41, 64

R

racines d'un polynôme, 206, 214, 231
rang, 267, 320
  théorème du, 272–280
  endomorphismes de $\mathcal{M}_n(\mathbb{C})$ conservant le rang, 323–328
  matrices de rang 1, 325
relation d'équivalence, 16
résultant, 182
rotation, 54
Rough-Hurwitz (critère de), 235

S

Schur (inégalité de), 225
scindé (polynôme), 211, 215–224
série génératrice, 9, 14, 210
Siegel (lemme de), 360
Sophie Germain (théorème de), 167
sous-espace vectoriel
  de $\mathcal{M}_n(K)$, 331
  stable, 269, 286, 292
sous-groupe
  d'un groupe libre de type fini, 352
  d'indice infini, 50
  distingué, 52
  fini de $\mathbb{Q}/\mathbb{Z}$, 40
  fini de $SO_3(\mathbb{R})$, 54
stable
  partie de $\mathbb{N}$ (pour l'addition), 124
  sous-espace, 269, 286, 292
  supplémentaire, 292
stochastique (matrice), 343
Sturm (théorème de), 238
supplémentaire, 261
  stable, 292
Sylvester (inégalité de), 272
symétrique
  fonction des racines, 206–214
  groupe, 46, 66–77

T

Tchebychev (polynômes de), 222–228
théorème
  des deux carrés, 160
  des quatre carrés, 162
trace, 328–331, 339
transcendance de e, 249

V

valeurs absolues
  sur $\mathbb{C}[X]$, 112
  sur $\mathbb{Q}$, 114
valuation, 110
  p-adique, 138–139, 143

W

Wilson (théorème de), 128

Z

$\zeta(2)$, 207

IMPRIMÉ EN GRANDE-BRETAGNE
PAR CAMBRIDGE UNIVERSITY PRESS
DÉPÔT LÉGAL OCTOBRE 2007

Ce livre est le premier tome d'un recueil d'exercices résolus des oraux des Écoles normales supérieures et de l'École polytechnique, qui en comportera finalement six. L'ensemble de l'ouvrage offre un très large panorama des exercices posés ces dix dernières années.

Les auteurs ont tenu à présenter les solutions les plus pédagogiques possible, essayant « d'exposer clairement les idées et démarches de raisonnement, sans pour autant escamoter les détails ou calculs qui peuvent paraître évidents. »

À ce niveau les énoncés proposés aux candidats sont souvent des résultats intéressants par eux-mêmes. Les auteurs ont fait l'effort de les identifier et de les présenter, autant que possible, dans leur contexte naturel. C'est ainsi qu'à l'intérieur d'un chapitre, les exercices sont regroupés par thèmes, et que chaque chapitre est précédé d'une introduction historique. Tout au long du chapitre, les exercices sont insérés dans un texte de présentation qui tour à tour dégage des idées générales, apporte des prolongements ou effectue quelques rappels de cours.

Ce livre s'adresse naturellement aux élèves des classes préparatoires, mais il sera également très utile aux candidats à l'agrégation qui y trouveront de nombreux développements pour leur oral. Ces exercices constituent aussi un excellent complément à la préparation à l'écrit du CAPES. Ce premier tome, qui comprend de l'algèbre générale (combinatoire, groupes, anneaux et corps) et linéaire (espaces vectoriels, matrices), est accessible aux étudiants de première année.

Dans cette seconde édition du premier tome, qui voit le jour alors que les deux derniers (Algèbre 3 et Analyse 3) restent à paraître, figurent 24 exercices nouveaux.

Collection enseignement des mathématiques

ISBN 978-2-84225-132-1

Graphisme : Massin

exercices de mathématiques

oraux x-ens

algèbre 2

Serge Francinou
Hervé Gianella
Serge Nicolas

CASSINI

EXERCICES DE MATHÉMATIQUES
ORAUX X-ENS

1. J.-Y. Ouvrard, *Probabilités 1*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouvrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*
24. R. Sá Earp, F. Toubiana, *Introduction à la géométrie hyperbolique et aux surfaces de Riemann*
25. L. Di Menza, *Résolution numérique des équations aux dérivées partielles*
26. B. Candelpergher, *Calcul intégral*
27. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques, vol. 1*
28. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques, vol. 2*

SERGE FRANCINOU
HERVÉ GIANELLA
SERGE NICOLAS

Exercices de mathématiques des oraux de l'École polytechnique et des Écoles normales supérieures

Algèbre. Tome II

Deuxième édition, revue et augmentée

CASSINI

SERGE FRANCINOU, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Marcelin Berthelot.

HERVÉ GIANELLA, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Blaise Pascal à Orsay.

SERGE NICOLAS, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Henri IV.

ISBN 978-2-84225-142-0
(1re édition, 2003, ISBN-10 : 2-84225-091-5, ISBN-13 : 978-2-84225-091-1)

© Cassini, Paris, 2009.

Table des matières

Introduction 1

Chapitre 1. Formes multilinéaires et déterminants 5

1.1. Dimension de l'espace des formes n-linéaires alternées 5
1.2. Formes n-linéaires alternées 7
1.3. Positivité de $\det(A^p + B^p)$ 9
1.4. Borne supérieure du déterminant sur une boule unité 10
1.5. Théorème d'Auerbach 10
1.6. Calcul d'un déterminant (1) 12
1.7. Calcul d'un déterminant (2) 13
1.8. Calcul d'un déterminant (3) 14
1.9. Calcul d'un déterminant (4) 16
1.10. Déterminant de Vandermonde 16
1.11. Matrice de Vandermonde incomplète 18
1.12. Vandermonde généralisé (1) 19
1.13. Vandermonde généralisé (2) 20
1.14. Application du déterminant de Vandermonde (1) 21
1.15. Application du déterminant de Vandermonde (2) 22
1.16. Application du déterminant de Vandermonde (3) 23
1.17. Discriminant 26
1.18. Déterminant de Cauchy 29
1.19. Déterminant de Hürwitz 32
1.20. Liberté d'une famille de fonctions 33
1.21. Matrice circulante 34
1.22. Majoration du déterminant d'une matrice stochastique . . . 35
1.23. Calcul d'une trace 37
1.24. Spectre de la comatrice 38
1.25. La transposée de la comatrice de A est un polynôme en A . 40
1.26. Comatrice d'un produit 41
1.27. Équation matricielle faisant intervenir la comatrice 42
1.28. Expression de la transposée de la comatrice de $XI_n - A$ . . 44
1.29. Différentielle du déterminant 46
1.30. Formule de Cauchy-Binet 47
1.31. Déterminant d'une matrice par blocs (1) 48
1.32. Déterminant d'une matrice par blocs (2) : formule de Williamson 49
1.33. Déterminant d'une matrice par blocs (3) 51
1.34. Déterminant d'une matrice par blocs (4) 51
1.35. Déterminant d'un automorphisme intérieur 53

1.36. Déterminant d'une matrice d'incidence 54
1.37. Dérangements pairs et impairs 55
1.38. Déterminant de Smith (1875) 56
1.39. Première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$ 57
1.40. Opération de $\mathrm{GL}_n(\mathbb{Z})$ sur le réseau $\mathbb{Z}^n$ 59
1.41. Un problème de poids 60
1.42. Décomposition LU avec pivot de Gauss 61
1.43. Décomposition LU d'une matrice tridiagonale 65

Chapitre 2. Réduction **69**

2.1. Éléments propres d'un opérateur intégral 71
2.2. Endomorphisme d'un espace de suites (1) 72
2.3. Endomorphisme d'un espace de suites (2) 74
2.4. Valeur propre simple 76
2.5. Existence d'une valeur propre double 77
2.6. Détermination de spectre 77
2.7. Algorithme de Faddeev 79
2.8. Lemme d'Hadamard, disques de Gershgörin 80
2.9. Matrices stochastiques (1) 83
2.10. Matrices stochastiques (2) 84
2.11. Matrices stochastiques (3) 85
2.12. Matrices strictement stochastiques 86
2.13. Théorème de Perron-Frobenius (1907) 87
2.14. Polynôme annulateur 90
2.15. Noyaux et images de polynômes d'endomorphismes 91
2.16. Diagonalisabilité d'une matrice 92
2.17. Diagonalisabilité d'une matrice tridiagonale 93
2.18. Racine d'une matrice diagonalisable inversible 95
2.19. Diagonalisabilité de f dans le cas f^2 diagonalisable 95
2.20. Étude de diagonalisabilité 97
2.21. Diagonalisation simultanée 98
2.22. Matrices circulantes (1) 99
2.23. Matrices circulantes (2) 101
2.24. Diagonalisabilité d'une matrice par blocs (1) 103
2.25. Diagonalisabilité d'une matrice par blocs (2) 104
2.26. Diagonalisabilité d'un produit tensoriel de matrices 106
2.27. Théorème de Hoffman et Singleton (1960) 107
2.28. Caractérisation des matrices nilpotentes avec la trace . . . 111
2.29. Théorème de Kronecker 113
2.30. Trigonalisation simultanée (1) 114
2.31. Trigonalisation simultanée (2) 116
2.32. Trigonalisation simultanée (3) 117
2.33. Trigonalisation simultanée (4) 118
2.34. Vecteur propre commun à une famille de matrices 120

2.35. Étude d'un opérateur d'un espace de suites (1) 122
2.36. Étude d'un opérateur d'un espace de suites (2) 123
2.37. Étude d'un opérateur d'un espace de suites (3) 125
2.38. Ordre d'une valeur propre dans le polynôme minimal 128
2.39. Familles de matrices anticommutantes 130
2.40. Décomposition de Dunford 134
2.41. Matrices nilpotentes (1) 135
2.42. Matrices nilpotentes (2) 136
2.43. Application adjointe (1) 136
2.44. Application adjointe (2) 138
2.45. Application adjointe (3) 139
2.46. Sous-espaces stables par un endomorphisme nilpotent (1) . 141
2.47. Sous-espaces stables par un endomorphisme nilpotent (2) . 141
2.48. Endomorphisme nilpotent semi-simple 142
2.49. Existence d'un supplémentaire stable par un endomorphisme nilpotent 142
2.50. Endomorphismes semi-simples (1) 144
2.51. Endomorphismes semi-simples (2) 145
2.52. Endomorphismes cycliques (1) 147
2.53. Endomorphismes cycliques (2) 149
2.54. Endomorphismes simples 152
2.55. Nombre fini de sous-espaces stables 152
2.56. Polynôme minimal ponctuel (1) 154
2.57. Polynôme minimal ponctuel (2) 155
2.58. Rang d'un endomorphisme et degré du polynôme minimal . 156
2.59. Réduction d'un endomorphisme en somme d'endomorphismes cycliques 156
2.60. Équation du second degré matricielle 159
2.61. Commutant d'une matrice carrée de taille 2 159
2.62. Dimension du commutant 160
2.63. Équation avec un crochet de Lie (1) 163
2.64. Équation avec un crochet de Lie (2) 164
2.65. Équation de Sylvester (1) 165
2.66. Équation de Sylvester (2) 166
2.67. Équation de Sylvester (3) 168
2.68. Matrices possédant une racine cubique 169
2.69. Fonctions polynomiales Φ vérifiant $\Phi(AB) = \Phi(BA)$ 172
2.70. Invariance du polynôme minimal par extension de corps . . 174
2.71. Similitude et extension de corps 175

Chapitre 3. Le groupe linéaire 177

3.1. Génération du groupe linéaire 177
3.2. Groupe engendré par les matrices diagonalisables inversibles 179
3.3. Produits d'automorphismes de trace nulle 179

3.4. Élément d'ordre 5 de $\mathrm{GL}_2(\mathbb{Q})$ 180
3.5. Conjugaison dans $\mathrm{SL}_2(\mathbb{C})$ 181
3.6. Isomorphismes entre groupes linéaires 182
3.7. Sous-groupe de $\mathrm{GL}_n(\mathbb{R})$ 183
3.8. Un théorème de Burnside 185
3.9. Petits sous-groupes de $\mathrm{GL}_n(\mathbb{C})$ 186
3.10. Groupe dérivé de $\mathrm{GL}_n(\mathrm{K})$ 188
3.11. Commutateur égal à $-\mathrm{I}_2$ 189
3.12. Morphismes de $(\mathbb{R}^*, \times)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$ 190
3.13. Morphismes de $\mathrm{GL}_n(\mathrm{K})$ dans un groupe abélien fini 192
3.14. Génération de $\mathrm{SL}_2(\mathbb{Z})$ 195
3.15. Endomorphismes surjectifs de $\mathrm{SL}_n(\mathbb{Z})$ 197
3.16. Sous-groupes finis de $\mathrm{SL}_2(\mathbb{Z})$ 198
3.17. Ordre des éléments de $\mathrm{GL}_2(\mathbb{Z})$ 199
3.18. Ordre des éléments de $\mathrm{GL}_n(\mathbb{Z})$ 200
3.19. Éléments d'ordre 4 ou 3 dans $\mathrm{SL}_2(\mathbb{Z})$ 201
3.20. Sous-groupes finis de $\mathrm{GL}_n(\mathbb{Z})$ 205
3.21. Un calcul de signature 206
3.22. Étude de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ 207
3.23. Étude de $\mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$ 209
3.24. Réduction modulo n 211
3.25. Surjection de $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ dans $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ 215

Chapitre 4. Topologie et matrices, exponentielle **217**

4.1. Matrices de rang r 217
4.2. Matrices diagonalisables complexes 218
4.3. Matrices diagonalisables réelles 219
4.4. Matrices cycliques 220
4.5. Racines de l'identité 221
4.6. Point isolé dans l'ensemble des racines q-ièmes de l'identité . 223
4.7. Ensemble des symétries 224
4.8. Partie compacte stable par produit 226
4.9. Rayon spectral 226
4.10. Une preuve du théorème de Cayley-Hamilton 229
4.11. Suite des puissances bornée 230
4.12. Suite des traces des puissances itérées 232
4.13. Semi-normes invariantes par similitude 232
4.14. Condition pour que $\mathrm{rg}\,\mathrm{A} = \mathrm{rg}\,\mathrm{A}^2$ 234
4.15. Classes de similitude bornées 235
4.16. Classe de similitude d'une matrice diagonalisable 236
4.17. Classe de similitude d'une matrice nilpotente 237
4.18. Groupe linéaire 238
4.19. Connexité d'un ensemble de matrices 240
4.20. Chemin continu de projecteurs 240

4.21. Exponentielle d'une matrice à coefficients non diagonaux positifs 242
4.22. Exponentielle de matrices réelles diagonalisables 243
4.23. Exponentielle des matrices symétriques réelles 243
4.24. Image de l'exponentielle 244
4.25. Antécédent de l'identité par l'exponentielle 247
4.26. Sous-groupes à un paramètre de $\mathrm{GL}_n(\mathbb{C})$ 248
4.27. Groupes à un paramètre dans $\mathcal{M}_n(\mathbb{R})$ 249
4.28. Morphismes continus de S^1 dans $\mathrm{GL}_n(\mathbb{C})$ 249
4.29. Morphismes continus de S^1 dans $\mathrm{GL}_n(\mathbb{R})$ 251
4.30. Formules de Lie 253
4.31. Algèbre de Lie de $\mathrm{SO}_n(\mathbb{R})$ 255
4.32. Sous-groupe fermé de $\mathrm{SL}_2(\mathbb{R})$ 256

Table des matières **261**

Index **267**

Introduction

Cet ouvrage est le second tome d'algèbre d'un recueil d'exercices de mathématiques destiné à la préparation des oraux des concours d'entrée aux Écoles normales supérieures et à l'École polytechnique. Il comportera six tomes, trois d'algèbre et trois d'analyse.

La vocation première des Écoles normales est de former des chercheurs ou des enseignants-chercheurs. Le concours d'entrée vise donc à détecter les qualités scientifiques du candidat, son aptitude à la recherche. À l'oral, on jugera avant tout la capacité de prendre des initiatives, d'utiliser une indication, de mener à bien une démarche. On ne sera pas surpris que les exercices posés aient un contenu mathématique riche, qu'ils soient très éloignés du simple exercice technique, d'application du cours, qu'ils soient souvent difficiles. Ils visent la plupart du temps à la démonstration d'un résultat mathématique significatif. Ils pourraient apparaître excessivement difficiles si on perdait de vue le déroulement concret de l'épreuve. L'oral des ENS est un long dialogue (l'épreuve dure environ cinquante minutes, comme d'ailleurs à l'École polytechnique) entre le candidat et l'examinateur, qui tout au long de l'épreuve fournit des indications, quand c'est nécessaire, pour relancer la réflexion du candidat et tester ses réactions. Il est d'ailleurs impossible de rendre pleinement compte dans un recueil d'exercices du caractère oral de l'épreuve.

L'École polytechnique, quant à elle, est plus généraliste. Les exercices posés au concours sont de facture plus classique et, en règle générale, l'examinateur intervient moins. C'est au candidat de montrer sa maîtrise du programme dans la résolution d'un exercice dont la difficulté est cependant très variable. Certains sont proches des exercices d'ENS. Les énoncés circulent d'ailleurs d'un concours à l'autre, ou peuvent même être repris d'exercices d'Olympiades.

Les énoncés qui figurent dans ce recueil ont été donnés entre 1996 et 2008. Ils sont extraits pour l'essentiel des listes publiées chaque année par la RMS (*Revue des mathématiques de l'enseignement supérieur* aux éditions Vuibert jusqu'en 2003 et désormais *Revue de la filière Mathématiques* aux éditions e.net) dont nous remercions les auteurs pour l'aide précieuse qu'ils apportent ainsi aux élèves et aux professeurs des classes préparatoires. Il s'agit de versions communiquées par les étudiants, reflétant la compréhension que ceux-ci ont eue de l'exercice et le déroulement conjoncturel de leur oral, comme le montrent les variations d'une année à l'autre pour un même exercice. Nous n'avons pas hésité à les modifier, pour rectifier des erreurs, compléter un énoncé quand manifestement

l'exercice était arrêté avant que le résultat que l'examinateur avait en vue ne soit atteint, ou ajouter des indications.

Nous avons choisi de laisser quelques énoncés «bruts», ceux pour lesquels nous estimons qu'une démarche naturelle (qui peut être longue et ardue) permet de conduire à la solution. Pour d'autres exercices, nous avons pris la liberté de rajouter des questions intermédiaires, qui auraient pu être celles posées par l'examinateur. Quitte à perdre en concision, nous avons tenu à rédiger les solutions les plus pédagogiques possible, essayant d'exposer clairement les idées et démarches des raisonnements sans pour autant escamoter les détails ou calculs qui peuvent paraître évidents. On évite autant que possible l'introduction d'une astuce ou d'un objet *ad hoc* permettant d'atteindre rapidement la solution. S'il n'y a pas moyen d'expliquer l'origine de cette astuce, c'est que l'exercice est peu intéressant et que l'étudiant en tirera peu de profit.

À l'intérieur de chaque chapitre, les exercices ont été regroupés thématiquement, et à l'intérieur de chaque thème, souvent par ordre de difficulté croissante. Ainsi regroupés, ils apparaîtront plus accessibles, car plongés dans leur contexte mathématique, éclairés par d'autres exercices voisins. Les introductions historiques qui ouvrent chaque chapitre, outre leur intérêt propre, visent au même but. Enfin, nous avons agrémenté les énoncés de quelques remarques préliminaires. Sans faire de rappels de cours systématiques, nous avons énoncé, voire redémontré certains résultats : lemmes classiques, intervenant dans la résolution d'un grand nombre d'exercices, ou résultats au contraire à la lisière du programme, mais utiles, pour lesquels des éclaircissements étaient nécessaires. On trouvera aussi des remarques de synthèse ou des généralisations qui, nous l'espérons, pourront amener le candidat curieux à approfondir ses connaissances. Les quelques indications bibliographiques ont le même objectif.

Le lecteur ne tirera profit de ce livre d'exercices que s'il cherche des solutions personnelles avant d'en étudier les corrigés. Une bonne connaissance du cours est indispensable. En effet, les théorèmes du programme fournissent bon nombre de schémas de démonstration. Rappelons aussi quelques démarches générales qui peuvent faciliter l'appréhension des exercices difficiles :

▷ ne pas hésiter à faire les calculs ou à étudier le problème en dimension 2 ou 3 : par exemple pour l'étude d'un déterminant ou la réduction d'une matrice.

▷ le renforcement des hypothèses peut aboutir à un problème plus simple : cas où le corps est algébriquement clos, cas où les matrices sont diagonalisables ou inversibles (avec extension possible du résultat par un argument de densité).

▷ considérer des sous-espaces stables permet souvent d'envisager un raisonnement par récurrence sur la dimension.

Au-delà des étudiants en classe préparatoire, ces ouvrages intéresseront aussi les candidats au CAPES et à l'Agrégation, qui y trouveront matière à réviser les principales notions du programme, ainsi que des exemples pour nourrir un développement pour leur oral.

Voyons maintenant plus précisément le contenu de ce second tome d'algèbre. Le premier chapitre est consacré au déterminant et peut être abordé dès la première année en classe préparatoire. Le second chapitre sur la réduction des endomorphismes constitue le cœur du programme d'algèbre de seconde année et il est le plus riche des quatre. Le troisième chapitre, qui contient des exercices souvent difficiles, est dédié à l'étude du groupe linéaire. Enfin le dernier chapitre regroupe les énoncés sur la topologie des espaces matriciels ainsi que quelques questions sur l'exponentielle de matrice (d'autres énoncés sur ce thème seront présents dans le chapitre sur les équations différentielles linéaires du tome analyse 3).

Comme dans les autres tomes, les exercices sont classés par thème. La difficulté est toutefois plutôt croissante : les chapitres commencent par des questions techniques ou des savoir-faire indispensables (calculs de déterminants, recherche de valeurs propres, générateurs du groupe linéaire...) et se terminent souvent par des exercices plus théoriques.

Le troisième et dernier tome d'algèbre portera sur les espaces euclidiens, les espaces hermitiens, les formes quadratiques et la géométrie.

Nous remercions André et Catherine Bellaïche et nos élèves Nicolas Curien, Julien Elie, Joon Kwon et Baptiste Teyssier pour leur relecture approfondie de l'ouvrage et leurs nombreuses suggestions, tant sur le fond que sur la forme.

Enfin, si vous souhaitez nous contacter pour nous faire part de vos remarques, vous pouvez envoyer un courriel à l'adresse *fgn.cassini@free.fr*.

Chapitre 1

Formes multilinéaires et déterminants

L'utilisation des matrices et des déterminants trouve son origine dans l'étude systématique des systèmes linéaires menée à partir du XVIII^{e}*siècle. Alors que Leibniz et Mac Laurin avaient déjà introduit les notations à indices et résolu les systèmes à deux ou trois inconnues. Cramer, en 1754, comprend que les solutions d'un système linéaire s'expriment comme quotient de deux expressions polynomiales multilinéaires des coefficients du système. Ces expressions représentent des déterminants mais ces derniers, étudiés notamment par Vandermonde et Laplace ne sont définis alors que par récurrence sur la taille (autrement dit par le développement par rapport à une rangée). On doit également à Laplace l'interprétation du déterminant en termes de volume. Par la suite, au début du* XIX^{e}*siècle Gauss, dans ses recherches sur les formes quadratiques, représente les changements de base dans* $\mathbb{R}^3$ *à l'aide de tableaux de nombres (les matrices) et introduit le produit de deux de ces tableaux pour obtenir la composée de deux changements de bases. Cela devait suggérer en 1812 à Cauchy la règle générale du produit de deux déterminants et il lui revient d'imposer la terminologie moderne.*

Les premiers exercices portent sur les formes multilinéaires alternées. Si E *est un* K*-espace vectoriel de dimension* n*, l'espace des formes* n*-linéaires alternées sur* E *est une droite vectorielle. Si* B *est une base de* E*, l'application déterminant dans la base* B *est un élément non nul de cette droite. L'exercice suivant contient ce résultat et demande plus généralement la dimension de l'espace des formes* n*-linéaires alternées sur un espace de dimension* d *quelconque.*

1.1. Dimension de l'espace des formes n-linéaires alternées

Soit E un K-espace vectoriel de dimension finie $d \geqslant 1$. Pour $n \geqslant 1$, on note $\mathcal{A}_n(\mathrm{E})$ l'espace des formes n-linéaires alternées sur E. Calculer $\dim \mathcal{A}_n(\mathrm{E})$.

(École normale supérieure)

▷ **Solution.**

Soit $\mathcal{B} = (e_1, \ldots, e_d)$ une base de E. À tout $(x_1, \ldots, x_n) \in \mathrm{E}^n$, on associe la matrice $\mathrm{A} = (a_{ij})_{\substack{1 \leqslant i \leqslant d \\ 1 \leqslant j \leqslant n}} \in \mathcal{M}_{d,n}(\mathrm{K})$ des coordonnées des vec-

leurs x_j dans la base B :

$$\forall j \in [\![1, n]\!],\ \ x_j = a_{1j}e_1 + a_{2j}e_2 + \cdots + a_{dj}e_d.$$

Soit $f \in \mathcal{A}_n$. Développons $f(x_1, \ldots, x_n)$ en utilisant la multilinéarité : $f\left(\sum_{i_1=1}^{d} a_{i_1 1}e_{i_1}, \ldots, \sum_{i_n=1}^{d} a_{i_n n}e_{i_n}\right)$ est égal à

$$\sum_{(i_1,\ldots i_n)\in[\![1,d]\!]^n} a_{i_1 1}a_{i_2 2}\ldots a_{i_n n}f(e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

Dans la somme obtenue on peut se limiter aux indices $(i_1, \ldots, i_n)$ deux à deux distincts puisque f est alternée. En particulier, si $n > d$, f est nulle et $\dim \mathcal{A}_n(\mathrm{E}) = 0$. On supposera donc dans la suite $n \leqslant d$.

Comme f est alternée elle est aussi antisymétrique. Pour toute permutation $\sigma \in \mathcal{S}_n$ on a donc

$$f(e_{i_{\sigma(1)}}, \ldots, e_{i_{\sigma(n)}}) = \varepsilon(\sigma)f(e_{i_1}, e_{i_2}, \ldots e_{i_n}).$$

Il en résulte que $f(x_1, \ldots, x_n)$ vaut

$$\sum_{1\leqslant i_1<i_2<\cdots<i_n\leqslant d} \left(\sum_{\sigma\in\mathcal{S}_n} \varepsilon(\sigma)a_{i_{\sigma(1)}1}\ldots a_{i_{\sigma(n)}n}\right) f(e_{i_1}, e_{i_2}, \ldots, e_{i_n}).$$

Notons $\mathcal{P}_n([\![1, d]\!])$ l'ensemble des parties à n éléments de $[\![1, d]\!]$. Pour toute partie $\mathrm{I} = \{i_1 < i_2 < \cdots < i_n\} \in \mathcal{P}_n([\![1, d]\!])$, considérons la forme n-linéaire $g_\mathrm{I} : \mathrm{E}^n \to \mathrm{K}$ définie pour tout $(x_1, \ldots, x_n) \in \mathrm{E}^n$ par

$$g_\mathrm{I}(x_1, \ldots, x_n) = \sum_{\sigma\in\mathcal{S}_n} \varepsilon(\sigma)a_{i_{\sigma(1)}1}\ldots a_{i_{\sigma(n)}n}.$$

On vient de prouver que toute forme n-linéaire alternée f est combinaison linéaire des formes n-linéaires g_I. Par ailleurs, ces formes n-linéaires g_I sont bien alternées pour toute partie I : cela résulte du cours. En effet, $\mathrm{A} = (a_{ij})_{\substack{1\leqslant i\leqslant d\\ 1\leqslant j\leqslant n}}$ désignant toujours la matrice de la famille $(x_1, \ldots, x_n) \in \mathrm{E}^n$ dans la base $\mathcal{B}$, g_I est le déterminant de la sous-matrice de A obtenue en ne gardant que les lignes d'indice $i_1, i_2, \ldots, i_n$. Il est alors clair que si deux des x_i sont égaux, cette sous-matrice a un déterminant nul. Donc $(g_\mathrm{I})_{\mathrm{I}\in\mathcal{P}_n([\![1,d]\!])}$ est une partie génératrice de $\mathcal{A}_n(\mathrm{E})$.

Il ne reste plus qu'à vérifier que cette famille est libre. Supposons donc que

$$\sum_{\mathrm{I}\in\mathcal{P}_n([\![1,d]\!])} \lambda_\mathrm{I} g_\mathrm{I} = 0.$$

Si on applique cette égalité sur $(e_{i_1}, \ldots, e_{i_n})$ avec $i_1 < i_2 < \cdots < i_n$, la seule partie I telle que $g_\mathrm{I}(e_{i_1}, \ldots, e_{i_n})$ soit non nul est $\mathrm{I} = \{i_1, i_2, \ldots, i_n\}$

pour laquelle on trouve 1. Il en résulte que tous les λ_I sont nuls et la famille est bien libre. Ainsi, $(g_I)_{I\in\mathcal{P}_n([\![1,d]\!])}$ est une base de $\mathcal{A}_n(\mathrm{E})$ et

$$\boxed{\dim \mathcal{A}_n(\mathrm{E}) = \mathrm{Card}\,\mathcal{P}_n([\![1,d]\!]) = C_d^n}. \quad \lhd$$

Pour $n = 1$ on retrouve la dimension du dual de E *qui est celle de* E *et pour $n = d$ on obtient* 1.

Dans l'exercice suivant on montre que pour qu'une forme n-linéaire soit alternée, il suffit qu'elle s'annule sur tout n-uplet de vecteurs contenant deux vecteurs consécutifs égaux. L'objectif de l'exercice est de montrer qu'une famille de n vecteurs qui annule toute forme n-linéaire alternée est forcément liée (dans un espace E *de dimension quelconque).*

1.2. Formes n-linéaires alternées

Soit E un K-espace vectoriel, $\mathcal{A}_n(\mathrm{E})$ l'espace des formes n-linéaires alternées sur E.

1. Montrer que si φ est une forme n-linéaire vérifiant $\varphi(x_1,\dots,x_n) = 0$ dès qu'il existe $i \in [\![1, n-1]\!]$ tel que $x_i = x_{i+1}$, alors φ est alternée.

2. Soit $f \in \mathcal{A}_n(\mathrm{E})$ et μ une forme linéaire sur E. On pose

$$g(x_1,\dots,x_{n+1}) = \sum_{i=1}^{n+1}(-1)^i \mu(x_i) f(x_1,\dots,x_{i-1},x_{i+1},\dots,x_{n+1}).$$

Montrer que $g \in \mathcal{A}_{n+1}(\mathrm{E})$.

3. Soit $(x_1,\dots,x_n) \in \mathrm{E}^n$ tel que $g(x_1,\dots,x_n) = 0$ pour tout $g \in \mathcal{A}_n(\mathrm{E})$. Montrer que pour toute forme $f \in \mathcal{A}_{n-1}(\mathrm{E})$ on a

$$\sum_{i=1}^{n}(-1)^i f(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)x_i = 0.$$

Montrer que la famille $(x_1,\dots,x_n)$ est liée.

(École normale supérieure)

▷ **Solution.**

1. Soit φ une forme n-linéaire vérifiant $\varphi(x_1,\dots,x_n) = 0$ dès qu'il existe $i \in [\![1, n-1]\!]$ tel que $x_i = x_{i+1}$. Soit $(x_1,\dots,x_n) \in \mathrm{E}^n$ tel qu'il existe $i < j$ avec $x_i = x_j$. Montrons que $\varphi(x_1,\dots,x_n) = 0$. C'est le cas si $j = i+1$. Sinon, développons

$$\varphi(x_1,\dots,x_{i-1},x_i - x_{i+1},x_{i+1} - x_i,x_{i+2},\dots,x_n)$$

qui, par hypothèse, est nul :

$$0 = \varphi(x_1,\dots,x_i,x_{i+1},\dots,x_n) - \varphi(x_1,\dots,x_i,x_i,\dots,x_n) + \varphi(x_1,\dots,x_{i+1},x_i,\dots,x_n) - \varphi(x_1,\dots,x_{i+1},x_{i+1},\dots,x_n),$$

ce qui donne $\varphi(x_1,\dots,x_i,x_{i+1},\dots,x_n) = -\varphi(x_1,\dots,x_{i+1},x_i,\dots,x_n)$. En réitérant le procédé, $\varphi(x_1,\dots,x_i,x_{i+1},\dots,x_n)$ est égal à

$$(-1)^{j-i-1}\varphi(x_1,\dots,x_{i-1},x_{i+1},\dots,x_{j-1},x_i,x_j,\dots,x_n) = 0.$$

2. La $(n+1)$-linéarité de g est claire. Utilisons la question 1 pour montrer que g est alternée. Soit $(x_1,\dots,x_n) \in \mathrm{E}^n$ avec $x_i = x_{i+1}$ pour un certain $i \in [\![1, n-1]\!]$. Dans l'écriture de $g(x_1,\dots,x_n)$, il y a $n-2$ termes nuls et il reste

$$\begin{aligned} g(x_1,\dots,x_n) &= (-1)^i \mu(x_i) f(x_1,\dots,x_{i-1},x_{i+1},x_{i+2},\dots,x_n) \\ &\quad + (-1)^{i+1}\mu(x_{i+1}) f(x_1,\dots,x_{i-1},x_i,x_{i+2},\dots,x_n) \\ &= 0. \end{aligned}$$

Donc g est alternée.

3. Soit $f \in \mathcal{A}_{n-1}(\mathrm{E})$. L'image du vecteur

$$\sum_{i=1}^{n} (-1)^i f(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)x_i$$

par toute forme linéaire $\mu \in \mathrm{E}^*$ est nulle d'après la question 2. Il en résulte que ce vecteur est nul.

Pour montrer que la famille $(x_1,\dots,x_n)$ est liée, on raisonne par récurrence sur n, le cas $n = 1$ étant immédiat. Supposons le résultat acquis au rang $n-1$ et considérons $(x_1,\dots,x_n) \in \mathrm{E}^n$ annulant toute forme n-linéaire alternée. S'il existe une forme $(n-1)$-linéaire f telle que $f(x_1,\dots,x_{n-1}) \neq 0$ c'est terminé car la relation ci-dessus est une relation de liaison. Sinon, l'hypothèse de récurrence s'applique à $(x_1,\dots,x_{n-1})$ et le résultat en découle car si $(x_1,\dots,x_{n-1})$ est liée, $(x_1,\dots,x_n)$ l'est aussi. ◁

La forme $(n+1)$-linéaire alternée g définie dans la question 2 est le produit extérieur de la forme linéaire μ et de la forme n-linéaire alternée f. Cette notion de produit extérieur est plus générale et permet d'associer à une forme n-linéaire alternée f et une forme m-linéaire g, une forme $(n+m)$-linéaire alternée. Ce calcul extérieur se révèle très important dans l'étude des formes différentielles (voir par exemple CARTAN HENRI, *Cours de calcul différentiel,* Hermann*).*

L'exercice suivant utilise seulement la multiplicativité du déterminant et le fait que pour toute matrice A *de* $\mathcal{M}_n(\mathbb{C})$, $\overline{\det \mathrm{A}} = \det(\overline{\mathrm{A}})$, *ce qui résulte immédiatement de la définition.*

1.3. Positivité de $\det(A^p + B^p)$

Soient A, B dans $\mathcal{M}_n(\mathbb{R})$ avec $AB = BA$ et $\det(A + B) \geqslant 0$. Montrer que pour tout $p \in \mathbb{N}^*$ on a $\det(A^p + B^p) \geqslant 0$.
(École normale supérieure)

▷ **Solution.**

On utilise la factorisation de $A^p + B^p$ dans $\mathcal{M}_n(\mathbb{C})$. Traitons pour commencer le cas $p = 2$. Comme A et B commutent, on a

$$A^2 + B^2 = (A + iB)(A - iB).$$

On en déduit que

$$\begin{aligned}\det(A^2 + B^2) &= \det(A + iB)\det(A - iB) = \det(A + iB)\overline{\det(A + iB)} \\ &= |\det(A + iB)|^2 \geqslant 0.\end{aligned}$$

Dans le cas général, en notant $\omega_k = e^{\frac{i(2k+1)\pi}{p}}$ $(0 \leqslant k \leqslant p-1)$ les racines p-ièmes de -1, on a dans $\mathbb{C}[X, Y]$, $X^p + Y^p = \prod_{k=0}^{p-1}(X - \omega_k Y)$. Comme A et B commutent, on en déduit

$$A^p + B^p = \prod_{k=0}^{p-1}(A - \omega_k B)$$

et donc

$$\det(A^p + B^p) = \prod_{k=0}^{p-1}\det(A - \omega_k B).$$

On remarque que $\overline{\omega_k} = \omega_{p-k-1}$ et donc $\overline{A - \omega_k B} = A - \omega_{p-k-1}B$

Si $p = 2q$ est pair, les matrices $A - \omega_k B$ sont deux à deux conjuguées. Comme dans le cas $p = 2$ on obtient

$$\det(A^p + B^p) = \prod_{k=0}^{2q-1}\det(A - \omega_k B) = \prod_{k=0}^{q-1}|\det(A - \omega_k B)|^2 \geqslant 0.$$

Si $p = 2q + 1$ est impair on a $\omega_q = -1$ et on obtient

$$\det(A^p + B^p) = \det(A + B)\prod_{k=0}^{q-1}|\det(A - \omega_k B)|^2 \geqslant 0,$$

car $\det(A + B) \geqslant 0$, par hypothèse. ◁

La multilinéarité et le caractère alterné montrent que le déterminant d'une matrice ne change pas si on ajoute à une colonne (ou une ligne) une combinaison linéaire des autres colonnes (ou lignes). Cela est pratiquement utilisé dans tous les calculs.

1.4. Borne supérieure du déterminant sur une boule unité

On note S l'ensemble des matrices $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ telle que $|a_{ij}| \leqslant 1$ pour tout couple (i,j) et on considère $\alpha = \sup_{A \in S} \det A$.
Montrer que α est fini, puis que α est un entier divisible par 2^{n-1}.
(École polytechnique)

▷ **Solution.**

L'ensemble S est une partie compacte de $\mathcal{M}_n(\mathbb{R})$. Le déterminant étant continu, il est borné sur S et atteint sa borne supérieure. On peut donc affirmer que α est fini et qu'il est atteint. Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ dans S une matrice telle que $\det A = \alpha$. On va montrer qu'il est possible de choisir A avec tous ses coefficients dans $\{\pm 1\}$. Le déterminant est une fonction affine de chaque coefficient. Pour $(i,j) \in [\![1,n]\!]^2$, il existe donc des fonctions polynomiales U_{ij} et V_{ij} des coefficients m_{kl}, mais indépendantes de m_{ij}, telles que pour toute matrice $M = (m_{kl})_{1 \leqslant k,l \leqslant n}$ on ait $\det M = U_{ij}(M) m_{ij} + V_{ij}(M)$.

Si $U_{ij}(A) = 0$, rien de nous empêche de remplacer le coefficient a_{ij} par 1 : cela ne change pas le déterminant. Si en revanche, $U_{ij}(A)$ est non nul, alors a_{ij} est nécessairement égal à ± 1 car, sur le segment $[-1,1]$, une fonction affine non constante $ax+b$ atteint nécessairement son maximum en 1 ou en -1. Comme cela vaut pour tout couple (i,j), on a montré qu'il existe une matrice $A \in S$ telle que $\det A = \alpha$ et dont tous les coefficients valent ± 1. Le déterminant d'une matrice à coefficients dans $\mathbb{Z}$ est un entier donc α est entier. Il ne reste plus qu'à prouver qu'il est divisible par 2^{n-1}. Pour cela on utilise la multilinéarité. Ajoutons la première colonne de A à chacune des autres. On peut alors mettre 2 en facteur dans chaque colonne d'indice 2 à n. La multilinéarité du déterminant permet de conclure. ◁

L'exercice suivant montre l'existence de bases remarquables dans un espace normé réel de dimension finie qui jouent un peu le rôle des bases orthonormées d'un espace euclidien. Posé sans indication l'exercice est assez difficile et des versions plus détaillées ont été proposées aux oraux.

1.5. Théorème d'Auerbach

Soit V un espace vectoriel réel de dimension finie $n \geqslant 1$ muni d'une norme $\| \ \|$. Montrer qu'il existe une base de V formée de vecteurs unitaires, dont la base duale est elle aussi formée de vecteurs unitaires pour la norme induite par $\| \ \|$ sur V^*.
(École normale supérieure)

▷ **Solution.**

Soit $B = (e_1, e_2, \dots, e_n)$ une base unitaire de V (c'est-à-dire une base formée de vecteurs unitaires) et $B^* = (e_1^*, \dots, e_n^*)$ sa base duale. On commence par observer que comme $e_i^*(e_i) = 1$, on a $\|e_i^*\| \geqslant 1$, où $\|\ \|$ désigne la norme induite sur V^* par la norme $\|\ \|$. L'exercice consiste donc à montrer qu'on peut atteindre la situation minimale.

Pour avoir des idées sur la réponse à la question, on peut commencer par regarder un cas particulier. Prenons pour E un espace euclidien muni de sa norme euclidienne et $B = (e_1, \dots, e_n)$ une base orthonormée de E. Alors, on a pour tout i et tout $x \in E$, $e_i^*(x) = \langle e_i, x\rangle$. Il résulte alors de l'inégalité de Cauchy-Schwarz que $|e_i^*(x)| \leqslant \|e_i\|\|x\| = \|x\|$. On a donc $\|e_i^*\| = 1$ et dans ce cas la réponse est positive.

Pour avoir une idée de la démarche qui va suivre, essayons de caractériser les bases orthonormées d'un espace euclidien parmi les bases unitaires en termes uniquement de normes et de distances sans faire intervenir explicitement le produit scalaire. On peut penser à l'inégalité d'Hadamard (cf. exercice 1.21 du tome algèbre 3). Si $(e_1, \dots, e_n)$ est une base unitaire,

$$\big|[e_1, \dots, e_n]\big| \leqslant \|e_1\| \dots \|e_n\| = 1$$

où $[e_1, \dots, e_n]$ est le produit mixte, c'est-à-dire le déterminant dans une base orthonormée directe de E (supposé orienté) où encore le volume du parallélotope construit sur $(e_1, \dots, e_n)$. Et il y a égalité si et seulement si la base est orthonormée (ce qui est géométriquement assez évident).

Dans notre espace E, on peut donc s'intéresser aux bases unitaires de E qui maximisent le déterminant dans une base B_0 fixée. Plus précisément, soit S la sphère unité de E et $f : S^n \to \mathbb{R}^+$ l'application qui à $(x_1, \dots, x_n)$ associe $|\det_{B_0}(x_1, \dots, x_n)|$. Comme f est continue sur S^n qui est compacte, f atteint son maximum. Soit $B = (e_1, \dots, e_n) \in S^n$ un point où f est maximale. Comme f n'est pas nulle, $f(e_1, \dots, e_n) > 0$ de sorte que B est une base unitaire de E. On sait déjà que $\|e_i^*\| \geqslant 1$ pour tout i. Montrons pour terminer que les formes linéaires e_i^* sont de norme 1. Soit $i \in [\![1, n]\!]$ et $x \in S$. Comme $(e_1, \dots, e_{i-1}, x, e_{i+1}, \dots, e_n) \in S^n$ on a

$$\big|\det_{B_0}(e_1, \dots, e_{i-1}, x, e_{i+1}, \dots, e_n)\big| \leqslant f(e_1, \dots, e_n).$$

Or, $x = \displaystyle\sum_{j=1}^{n} e_j^*(x) e_j$. On remplace x par cette expression et on développe le déterminant par linéarité par rapport à la i-ième place. On a donc :

$$\begin{aligned}\det_{B_0}(e_1, \dots, e_{i-1}, x, \dots, e_n) &= \sum_{j=1}^{n} e_j^*(x) \det_{B_0}(e_1, \dots, e_{i-1}, e_j, \dots, e_n)\\ &= e_i^*(x) f(e_1, \dots, e_n)\end{aligned}$$

car les déterminants qui figurent dans la somme pour $j \neq i$ sont nuls. L'inégalité précédente conduit donc à $|e_i^*(x)| \leqslant 1$. Ceci vaut pour tout x unitaire, donc $\|e_i^*\| = 1$. On a donc trouvé une base répondant à la question de l'exercice. $\triangleleft$

Une telle base est appelée base d'Auerbach de E.

Les prochains exercices sont consacrés à des calculs de déterminants.

1.6. Calcul d'un déterminant (1)

Calculer

$$\det \begin{pmatrix} a_1+b_1 & b_1 & \dots & b_1 \\ b_2 & a_2+b_2 & \dots & b_2 \\ \vdots & & \ddots & \vdots \\ b_n & b_n & \dots & a_n+b_n \end{pmatrix}.$$

(École polytechnique)

$\triangleright$ **Solution.**

Notons $\mathrm{C} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$, $(e_1, e_2, \dots, e_n)$ la base canonique de K^n et

$$\mathrm{M} = \begin{pmatrix} a_1+b_1 & b_1 & \dots & b_1 \\ b_2 & a_2+b_2 & \dots & b_2 \\ \vdots & & \ddots & \vdots \\ b_n & b_n & \dots & a_n+b_n \end{pmatrix}.$$

La transposée de M admet pour i-ème colonne $b_i\mathrm{C} + a_i e_i$. Ainsi,

$$\det \mathrm{M} = \det({}^t\mathrm{M}) = \det(a_1e_1 + b_1\mathrm{C}, a_2e_2 + b_2\mathrm{C}, \dots, a_ne_n + b_n\mathrm{C}).$$

On développe cette expression en utilisant la multilinéarité du déterminant. Chaque terme où apparaissent $b_i\mathrm{C}$ et $b_j\mathrm{C}$ avec $i \neq j$ est nul, car le déterminant est une forme n-linéaire alternée. Il reste

$$\begin{aligned} \det \mathrm{M} &= \det(a_1e_1, \dots, a_ne_n) \\ &\quad + \sum_{k=1}^{n} \det(a_1e_1, \dots, a_{k-1}e_{k-1}, b_k\mathrm{C}, a_{k+1}e_{k+1}, \dots, a_ne_n) \\ &= a_1 \dots a_n + \sum_{k=1}^{n} b_k \left(\prod_{i \neq k} a_i \right) \det(e_1, \dots e_{k-1}, \mathrm{C}, e_{k+1}, \dots, e_n). \end{aligned}$$

Si à C on enlève $e_1 + \dots + e_{k-1} + e_{k+1} + \dots + e_n$ $(1 \leqslant k \leqslant n)$, il reste e_k. Comme det est une forme multilinéaire alternée, on a

$$\det(e_1,\dots e_{k-1},\mathrm{C},e_k,\dots,e_n)=\det(e_1\dots,e_{k-1},e_k.e_{k+1},\dots,e_n)=1$$

et finalement

$$\boxed{\det \mathrm{M}=\prod_{i=1}^{n}a_i+\sum_{k=1}^{n}b_k\prod_{i\neq k}a_i}.\lhd$$

L'exemple suivant est typique des calculs de déterminants : on effectue des opérations sur la matrice de manière à annuler certains coefficients, puis on développe selon une ligne ou une colonne pour se ramener à des déterminants de taille inférieure.

1.7. Calcul d'un déterminant (2)

Calculer le déterminant de la matrice $(\mathrm{C}_{i+j-2}^{j-1})_{1\leqslant i,j\leqslant n+1}$.
(École polytechnique)

$\rhd$ **Solution.**

Il s'agit de calculer $\mathrm{D}=\begin{vmatrix} \mathrm{C}_0^0 & \mathrm{C}_1^1 & \dots & \mathrm{C}_{n-1}^{n-1} & \mathrm{C}_n^n \\ \mathrm{C}_1^0 & \mathrm{C}_2^1 & \dots & \mathrm{C}_n^{n-1} & \mathrm{C}_{n+1}^n \\ \mathrm{C}_2^0 & \mathrm{C}_3^1 & \dots & \mathrm{C}_{n+1}^{n-1} & \mathrm{C}_{n+2}^n \\ \vdots & \vdots & & \vdots & \vdots \\ \mathrm{C}_n^0 & \mathrm{C}_{n+1}^1 & \dots & \mathrm{C}_{2n-1}^{n-1} & \mathrm{C}_{2n}^n \end{vmatrix}$.

On va effectuer sur les lignes de la matrice les opérations élémentaires suivantes : $\mathrm{L}_{n+1}\leftarrow \mathrm{L}_{n+1}-\mathrm{L}_n$, $\mathrm{L}_n\leftarrow \mathrm{L}_n-\mathrm{L}_{n-1}\dots,\mathrm{L}_2\leftarrow \mathrm{L}_2-\mathrm{L}_1$. Compte-tenu de la relation de Pascal, on a $\mathrm{C}_{i+j-2}^{j-1}-\mathrm{C}_{i+j-3}^{j-1}=\mathrm{C}_{i+j-3}^{j-2}$ pour tout $1\leqslant i\leqslant n+1$ et $2\leqslant j\leqslant n$, et il reste donc

$$\mathrm{D}=\begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ 0 & \mathrm{C}_1^0 & \mathrm{C}_2^1 & \dots & \mathrm{C}_n^{n-1} \\ 0 & \mathrm{C}_2^0 & \mathrm{C}_3^1 & \dots & \mathrm{C}_{n+1}^{n-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & \mathrm{C}_n^0 & \mathrm{C}_n^1 & \dots & \mathrm{C}_{2n-1}^{n-1} \end{vmatrix}=\begin{vmatrix} \mathrm{C}_1^0 & \mathrm{C}_2^1 & \dots & \mathrm{C}_n^{n-1} \\ \mathrm{C}_2^0 & \mathrm{C}_3^1 & \dots & \mathrm{C}_{n+1}^{n-1} \\ \vdots & \vdots & & \vdots \\ \mathrm{C}_n^0 & \mathrm{C}_n^1 & \dots & \mathrm{C}_{2n-1}^{n-1} \end{vmatrix}.$$

Sur ce dernier déterminant, on réitère la même suite d'opérations à savoir : $\mathrm{L}_n\leftarrow \mathrm{L}_n-\mathrm{L}_{n-1},\dots,\mathrm{L}_2\leftarrow \mathrm{L}_2-\mathrm{L}_1$:

$$\mathrm{D}=\begin{vmatrix} 1 & \mathrm{C}_2^1 & \dots & \mathrm{C}_n^{n-1} \\ 0 & \mathrm{C}_2^0 & \dots & \mathrm{C}_{n-2}^{n-1} \\ \vdots & \vdots & & \vdots \\ 0 & \mathrm{C}_{n-1}^0 & \dots & \mathrm{C}_{2n-2}^{n-2} \end{vmatrix}=\begin{vmatrix} \mathrm{C}_2^0 & \dots & \mathrm{C}_{n-2}^{n-1} \\ \vdots & & \vdots \\ \mathrm{C}_{n-1}^0 & \dots & \mathrm{C}_{2n-2}^{n-2} \end{vmatrix}.$$

Ainsi, de proche en proche, au bout de n séries d'opérations, il reste un déterminant de taille (1.1) :

$$\mathrm{D} = \left|\mathrm{C}_{2n-n}^{n-n}\right| = \left|\mathrm{C}_n^0\right| = 1. \lhd$$

Prévisons que la notation avec deux grandes barres verticales pour signifier un déterminant est due à Cayley (1841).

1.8. Calcul d'un déterminant (3)

Calculer le déterminant

$$\begin{vmatrix} a_1 & a_0 & 0 & \dots & 0 & 0 \\ a_2 & a_1 & a_0 & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & 0 \\ a_{n-1} & a_{n-2} & \dots & \dots & a_1 & a_0 \\ a_n & a_{n-1} & a_{n-2} & \dots & a_2 & a_1 \end{vmatrix}$$

avec $a_0 = 1$, $a_1 = \dfrac{1}{2}, \dots, a_n = \dfrac{1 \cdot 3 \cdot 5 \cdots (2n-1)}{2 \cdot 4 \cdot 6 \cdots (2n)}$.

(École polytechnique)

▷ **Solution.**

Appelons D_n le déterminant que nous devons calculer. On va essayer de trouver une relation de récurrence et pour cela on va développer D_n par rapport à la dernière ligne. Le cofacteur de a_k est

$$(-1)^{k+1}\begin{vmatrix} a_1 & a_0 & 0 & \dots & 0 & 0 & \dots & \dots & 0 \\ a_2 & a_1 & a_0 & & \vdots & \vdots & & & \vdots \\ \vdots & & & \ddots & \vdots & \vdots & & & \vdots \\ a_{n-k-1} & & & & a_0 & 0 & \dots & \dots & 0 \\ a_{n-k} & & & & a_1 & 0 & 0 & \dots & 0 \\ a_{n-k+1} & & & & a_2 & a_0 & 0 & \dots & 0 \\ \vdots & & & & \vdots & a_1 & a_0 & & 0 \\ \vdots & & & & \vdots & \vdots & & \ddots & \vdots \\ a_{n-1} & & & & a_k & a_{k-2} & & & a_0 \end{vmatrix}.$$

ce qui donne, d'après les règles de calcul des déterminants par blocs, $(-1)^{k+1}\mathrm{D}_{n-k}a_0^{k-1} = (-1)^{k+1}\mathrm{D}_{n-k}$. On en déduit que

$$\mathrm{D}_n = (-1)^{n+1}a_n\mathrm{D}_0 + (-1)^n a_{n-1}\mathrm{D}_1 + \cdots + (-1)^{k+1}a_k\mathrm{D}_{n-k} + \cdots + a_1\mathrm{D}_{n-1},$$

avec la convention $\mathrm{D}_0 = 1$. Autrement dit, si on pose

$$u_n = \sum_{k=0}^{n} (-1)^k a_k \mathrm{D}_{n-k},$$

on a $u_0 = 1$ et $u_n = 0$ pour $n \geqslant 1$.

On reconnaît dans la suite $(u_n)_{n\geqslant 0}$ le produit de Cauchy de la suite $((-1)^n a_n)_{n\geqslant 0}$ et de la suite $(\mathrm{D}_n)_{n\geqslant 0}$. Ainsi la suite $(\mathrm{D}_n)_{n\geqslant 0}$ est l'inverse pour le produit de convolution de la suite $((-1)^n a_n)_{n\geqslant 0}$. Pour expliciter cet inverse on considère les deux séries entières $f(x) = \sum_{n=0}^{+\infty} \mathrm{D}_n x^n$ et $g(x) = \sum_{n=0}^{+\infty} (-1)^n a_n x^n$. La seconde a un rayon de convergence égal à 1 et on reconnaît le développement en série entière de $(1+x)^{-1/2}$. En effet, rappelons que si $|x| < 1$ et α est réel, on a

$$(1+x)^\alpha = 1 + \sum_{n=1}^{+\infty} \frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} x^n.$$

En prenant $\alpha = -\dfrac{1}{2}$, on obtient

$$\frac{\alpha(\alpha-1)\dots(\alpha-n+1)}{n!} = \frac{(-1/2).(-3/2)\dots(1/2-n)}{n!} = (-1)^n a_n.$$

Or $\dfrac{1}{g(x)} = \sqrt{1+x}$ est aussi développable en série entière avec un rayon égal à 1. On peut écrire ce développement en prenant $\alpha = \dfrac{1}{2}$ dans la formule ci-dessus ou observer que $\dfrac{\mathrm{d}}{\mathrm{d}x}(\sqrt{1+x}) = \dfrac{1}{2\sqrt{1+x}} = \dfrac{1}{2} g(x)$. Par intégration on a

$$\sqrt{1+x} = \frac{1}{g(x)} = 1 + \sum_{n=1}^{+\infty} \frac{(-1)^{n-1} a_{n-1}}{2n} x^n.$$

Si on écrit pour $|x| < 1$, $\sqrt{1+x} = \sum_{n=0}^{+\infty} \mathrm{D}'_n x^n$, la règle du produit de Cauchy donne

$$\sum_{k=0}^{n} (-1)^k a_k \mathrm{D}'_{n-k} = \delta_{n,0}.$$

Comme $\mathrm{D}_0 = \mathrm{D}'_0 = 1$ une récurrence immédiate montre que $\mathrm{D}_n = \mathrm{D}'_n$ pour tout n. On a donc pour tout $n \geqslant 1$,

$$\boxed{\mathrm{D}_n = (-1)^{n+1} \frac{a_{n-1}}{2n}}.$$

Observons pour finir que a_n vaut aussi $\dfrac{(2n)!}{(2^n n!)^2} = \dfrac{1}{4^n} \mathrm{C}_{2n}^n$. ◁

L'expression polynomiale du déterminant sert rarement dans les calculs pratiques. C'est toutefois le cas dans l'exercice suivant.

1.9. Calcul d'un déterminant (4)

Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$ telle que :
(*i*) pour $i \neq j$, $a_{ij} \neq 0 \Longrightarrow a_{ji} = 0$;
(*ii*) ($a_{ij} \neq 0$ et $a_{jk} \neq 0$) $\Longrightarrow a_{ik} \neq 0$ dès que i, j, k sont deux à deux distincts.
Calculer $\det A$.

(École normale supérieure)

▷ **Solution.**
On va utiliser la définition du déterminant

$$\det A = \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) \prod_{i=1}^{n} a_{i\sigma(i)}.$$

Soit σ une permutation distincte de l'identité. Elle admet une décomposition en produit de cycles à supports disjoints. Soit $c = (i_1, i_2, \ldots, i_k)$, $k \geqslant 2$, l'un des cycles en question. Dans le produit $\prod_{i=1}^{n} a_{i\sigma(i)}$ intervient le facteur $a_{i_1 i_2} a_{i_2 i_3} \ldots a_{i_{k-1} i_k} a_{i_k i_1}$. Mais les propriétés (*i*) et (*ii*) impliquent la nullité d'au moins un terme de ce produit (s'ils sont tous non nuls, la transitivité conduit à $a_{i_1 i_k} \neq 0$ et alors $a_{i_k i_1} = 0$ par (*i*) : contradiction). La seule contribution au déterminant est donc donnée par l'identité et

$$\boxed{\det A = \prod_{i=1}^{n} a_{ii}}. \quad \triangleleft$$

Les exercices ci-après concernent des déterminants très classiques, à commencer par celui de Vandermonde.

1.10. Déterminant de Vandermonde

1. Calculer le déterminant

$$V(x_1, \ldots, x_n) = \begin{vmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{vmatrix}.$$

2. Montrer que, pour tout $(m_1, \ldots, m_n) \in \mathbb{Z}^n$, l'entier $\prod_{k=1}^{n-1} k!$ divise $\prod_{1 \leqslant i < j \leqslant n} (m_j - m_i)$.

(École polytechnique)

▷ **Solution.**

1. Si deux des x_i sont égaux, $\mathrm{V}(x_1, \ldots, x_n) = 0$ (deux lignes sont identiques). Supposons les x_i deux à deux distincts et posons

$$\mathrm{P}(\mathrm{X}) = \mathrm{V}(x_1, x_2, \ldots, x_{n-1}, \mathrm{X}).$$

C'est un polynôme de degré au plus $n-1$. Le développement suivant la dernière ligne nous persuade que le coefficient en X^{n-1} est $\mathrm{V}(x_1, \ldots, x_{n-1})$. De plus, si on substitue x_i à X $(1 \leqslant i \leqslant n-1)$, deux lignes sont identiques et $\mathrm{P}(x_i) = 0$. Les x_i étant deux à deux distincts, P est divisible par $\prod_{i=1}^{n-1} (\mathrm{X} - x_i)$, polynôme unitaire de degré $n-1$. Il en résulte que

$$\mathrm{P}(\mathrm{X}) = \mathrm{V}(x_1, \ldots, x_{n-1}) \prod_{i=1}^{n-1} (\mathrm{X} - x_i),$$

et en particulier $\mathrm{P}(x_n) = \mathrm{V}(x_1, \ldots, x_{n-1}) \prod_{i=1}^{n-1} (x_n - x_i)$. De proche en proche, on en tire la célèbre formule de Vandermonde

$$\boxed{\mathrm{V}(x_1, \ldots, x_n) = \prod_{1 \leqslant i < j \leqslant n} (x_j - x_i)}.$$

2. On reconnaît dans le second entier $\mathrm{V}(m_1, \ldots, m_n)$. Or, par des opérations sur les colonnes, il est clair que $\mathrm{V}(m_1, \ldots, m_n)$ vaut aussi

$$\begin{vmatrix} 1 & m_1 & m_1(m_1-1) & \ldots & m_1(m_1-1)\cdots(m_1-n+1) \\ 1 & m_2 & m_2(m_2-1) & \ldots & m_2(m_2-1)\cdots(m_2-n+1) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & m_n & m_n(m_n-1) & \ldots & m_n(m_n-1)\cdots(m_n-n+1) \end{vmatrix}.$$

Comme un produit de k entiers consécutifs est toujours divisible par $k!$ (cf. par exemple exercice 5.24 du tome 1 pour une preuve de ce résultat classique), les entiers de la $(k+1)$-ième colonne sont tous divisibles par $k!$. Par multilinéarité $\mathrm{V}(m_1, \ldots, m_n)$ est divisible par $\prod_{k=1}^{n-1} k!$. ◁

Le lecteur trouvera une seconde solution de la question 2 (moins élégante) dans l'exercice 5.57 du tome algèbre 1.

Les cinq exercices qui suivent supposent connu le déterminant de Vandermonde. On imagine que lors de l'oral le candidat a certainement été amené à rapidement retrouver ce résultat classique.

1.11. Matrice de Vandermonde incomplète

Pour $0 \leqslant k \leqslant n$, calculer le déterminant :

$$\mathrm{D}_k = \begin{vmatrix} 1 & x_1 & \dots & x_1^{k-1} & x_1^{k+1} & \dots & x_1^n \\ \vdots & \vdots & & \vdots & \vdots & & \vdots \\ 1 & x_n & \dots & x_n^{k-1} & x_n^{k+1} & \dots & x_n^n \end{vmatrix}.$$

(École polytechnique)

▷ **Solution.**

L'idée est simplement de se ramener à un déterminant de Vandermonde en rajoutant la colonne manquante, une ligne supplémentaire et d'interpréter le déterminant proposé comme un mineur de la matrice obtenue. On considère donc le polynôme

$$\mathrm{P}(\mathrm{X}) = \begin{vmatrix} 1 & x_1 & x_1^2 & \dots & x_1^{n-1} & x_1^n \\ 1 & x_2 & x_2^2 & \dots & x_2^{n-1} & x_2^n \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-1} & x_n^n \\ 1 & \mathrm{X} & \mathrm{X}^2 & \dots & \mathrm{X}^{n-1} & \mathrm{X}^n \end{vmatrix}.$$

Le déterminant de Vandermonde incomplet D_k s'interprète comme le mineur relatif à X^k. On a, d'après l'exercice précédent,

$$\begin{aligned} \mathrm{P}(\mathrm{X}) = \mathrm{V}(x_1, \dots, x_n, \mathrm{X}) &= \mathrm{V}(x_1, \dots, x_n) \prod_{k=1}^{n} (\mathrm{X} - x_k) \\ &= \mathrm{V}(x_1, \dots, x_n)(\mathrm{X}^n - \sigma_1 \mathrm{X}^{n-1} + \dots + (-1)^n \sigma_n) \end{aligned}$$

où les σ_k sont les fonctions symétriques élémentaires des x_k. Le développement par rapport à la dernière ligne nous assure que

$$(-1)^{n+1+k+1} \mathrm{D}_k = (-1)^{n-k} \sigma_{n-k} \mathrm{V}(x_1 \dots, x_n)$$

et finalement

$$\boxed{\mathrm{D}_k = \sigma_{n-k} \mathrm{V}(x_1, \dots, x_n)}. \quad \triangleleft$$

Dans l'exercice suivant on considère des déterminants qui ressemblent à celui de Vandermonde, mais les exposants sont des réels strictement positifs quelconques.

1.12. Vandermonde généralisé (1)

Soient $\alpha_1 < \cdots < \alpha_n$ des réels et $a_1, \ldots, a_n$ des réels non tous nuls.

1. Montrer que la fonction $f(x) = a_1 x^{\alpha_1} + \cdots + a_n x^{\alpha_n}$ admet au plus $n-1$ zéros distincts dans $\mathbb{R}_+^*$.

2. Soient $t_1, \ldots, t_n$ des réels tels que $0 < t_1 < \cdots < t_n$. Montrer que le déterminant de la matrice $(t_i^{\alpha_j})_{1 \leqslant i,j \leqslant n}$ est strictement positif.

(**École polytechnique**)

▷ **Solution.**

1. On procède par récurrence sur n, le résultat étant clair pour $n = 1$, puisque une fonction puissance $x \mapsto x^\alpha$ ne s'annule pas sur $\mathbb{R}_+^*$. Supposons le résultat vrai au rang $n-1$. Soient $\alpha_1 < \cdots < \alpha_n$ des réels et $f(x) = a_1 x^{\alpha_1} + \cdots + a_n x^{\alpha_n}$. On peut supposer qu'aucun a_i n'est nul sinon l'hypothèse de récurrence s'applique directement. Supposons par l'absurde que f admette n zéros distincts $0 < x_1 < x_2 < \cdots < x_n$. En divisant par x^{α_1}, on constate que la fonction g définie sur $\mathbb{R}_+^*$ par

$$g(x) = a_1 + a_2 x^{\alpha_2 - \alpha_1} + \cdots + a_n x^{\alpha_n - \alpha_1}$$

s'annule aussi en les x_k. Le théorème de Rolle appliqué sur chaque intervalle $[x_k, x_{k+1}]$ montre que g' admet au moins $n-1$ zéros deux à deux distincts. Il en est de même de la fonction

$$x \longmapsto x^{\alpha_1} g'(x) = \sum_{k=2}^{n} a_k(\alpha_k - \alpha_1) x^{\alpha_k}.$$

Cela contredit l'hypothèse de récurrence.

2. On va procéder par récurrence sur n. Pour $n = 1$ le déterminant vaut $t_1^{\alpha_1}$ et il est strictement positif. Supposons le résultat vrai au rang $n-1$ et soit $t_1, \ldots, t_n$ des réels tels que $0 < t_1 < \cdots < t_n$. On considère la fonction

$$f(t) = \begin{vmatrix} t_1^{\alpha_1} & t_1^{\alpha_2} & \ldots & t_1^{\alpha_n} \\ t_2^{\alpha_1} & t_2^{\alpha_2} & \ldots & t_2^{\alpha_n} \\ \vdots & \vdots & & \vdots \\ t^{\alpha_1} & t^{\alpha_2} & \ldots & t^{\alpha_n} \end{vmatrix}.$$

Le problème consiste à montrer que $f(t_n) > 0$. Or, si on développe selon la dernière ligne, on constate que f est de la forme

$$t \longmapsto a_1 t^{\alpha_1} + \cdots + a_n t^{\alpha_n}.$$

L'hypothèse de récurrence permet de dire que le coefficient a_n est strictement positif. La question 1 montre que f admet au plus $n-1$ zéros

distincts. Or, il est clair que $f(t_1) = f(t_2) = \cdots = f(t_{n-1}) = 0$ (car pour ces valeurs deux des lignes de la matrice sont égales). Il en résulte que f ne s'annule pas sur l'intervalle $]t_{n-1}, +\infty[$, et comme elle est continue, elle y garde un signe constant. Il suffit d'observer que $f(t)$ est équivalent à $a_n t^{\alpha_n}$ quand t tend vers $+\infty$ pour conclure que f est strictement positive sur $]t_{n-1}, +\infty[$, ce qui termine la récurrence. $\triangleleft$

Le lecteur pourra de la même manière traiter la variante suivante proposée aux oraux des écoles normales supérieures : si $a_1, \ldots, a_n$ sont des réels deux à deux distincts et $b_1, \ldots, b_n$ sont des réels deux à deux distincts, alors le déterminant de la matrice $(\exp(a_i b_j))_{1 \leqslant i,j \leqslant n}$ est non nul.

1.13. Vandermonde généralisé (2)

Soit $i_1, i_2 \ldots, i_k$ dans $\mathbb{N}^*$ et $n = i_1 + \cdots + i_k$. Pour $x \in \mathbb{R}$, on note $\mathrm{C}(x) = (1, x, x^2, \ldots, x^{n-1})$. On se donne $x_1, \ldots, x_k$ dans $\mathbb{R}$ et on considère la matrice M dont les lignes successives sont

$$\mathrm{C}(x_1), \frac{\mathrm{C}'(x_1)}{1!}, \ldots, \frac{\mathrm{C}^{(i_1-1)}(x_1)}{(i_1-1)!}, \ldots, \mathrm{C}(x_k), \frac{\mathrm{C}'(x_k)}{1!}, \ldots, \frac{\mathrm{C}^{(i_k-1)}(x_k)}{(i_k-1)!}.$$

Calculer $\det \mathrm{M}$.

(École polytechnique)

▷ **Solution.**

On considère un polynôme P réel unitaire de degré $n-1$ qu'on écrit $\mathrm{P} = a_0 + a_1 \mathrm{X} + \cdots + a_{n-2}\mathrm{X}^{n-2} + \mathrm{X}^{n-1}$. On ajoute à la dernière colonne de M la première multipliée par a_0, la seconde multipliée par a_1, jusqu'à l'avant dernière multipliée par a_{n-2}. Cela ne change pas le déterminant et fait apparaître dans la dernière colonne les coefficients

$$\mathrm{P}(x_1), \frac{\mathrm{P}'(x_1)}{1!}, \ldots, \frac{\mathrm{P}^{(i_1-1)}(x_1)}{(i_1-1)!}, \ldots, \mathrm{P}(x_k), \frac{\mathrm{P}'(x_x)}{1!}, \ldots, \frac{\mathrm{P}^{(i_k-1)}(x_k)}{(i_k-1)!}.$$

L'idée est de choisir le polynôme P de manière à avoir beaucoup de 0 dans cette dernière colonne. En fait, on peut s'arranger pour que $\mathrm{P}(x_1) = \mathrm{P}'(x_1) = \cdots = \mathrm{P}^{(i_1-1)}(x_1) = 0$: il suffit que x_1 soit racine d'ordre i_1 de P. De même pour les autres x_i. Toutefois le degré de P doit être égal à $n-1$ et P ne peut donc avoir que $n-1$ racines avec multiplicité. On prend alors

$$\mathrm{P}(\mathrm{X}) = (\mathrm{X} - x_1)^{i_1} \ldots (\mathrm{X} - x_{k-1})^{i_{k-1}} (\mathrm{X} - x_k)^{i_k - 1}.$$

Avec ce polynôme, le seul coefficient non nul dans la dernière colonne est $\frac{\mathrm{P}^{(i_k-1)}(x_k)}{(i_k-1)!}$. Or, il est aisé de voir que

$$\mathrm{P}^{(i_k-1)}(x_k) = (i_k-1)!(x_k-x_1)^{i_1}\dots(x_k-x_{k-1})^{i_{k-1}}.$$

En développant selon la dernière colonne, on voit que

$$\det \mathrm{M} = \frac{\mathrm{P}^{(i_k-1)}(x_k)}{(i_k-1)!} \times \det \mathrm{N} = \prod_{1\leqslant j\leqslant k-1} (x_k-x_j)^{i_j} \times \det \mathrm{N},$$

où N est la matrice de taille $n-1$ fabriquée tout comme M mais avec x_k à l'ordre i_k-1. Une récurrence aisée sur n, montre alors que

$$\boxed{\det \mathrm{M} = \prod_{j<k} (x_k-x_j)^{i_j i_k}}. \quad \lhd$$

Les deux exercices suivants utilisent des déterminants de Vandermonde.

1.14. Application du déterminant de Vandermonde (1)

Soit $\mathrm{A} = (a_{ij}) \in \mathcal{M}_n(\mathrm{K})$ avec, pour $1 \leqslant i,j \leqslant n$, $a_{ij} = 1 + x_i^j$ où $x_1,\dots,x_n$ sont des scalaires. Calculer le déterminant de A.

(École polytechnique)

▷ **Solution.**

En soustrayant à chaque colonne d'indice $j \in [\![2,n]\!]$ la colonne précédente, on obtient

$$\det \mathrm{A} = \begin{vmatrix} 1+x_1 & x_1(x_1-1) & \dots & x_1^{n-1}(x_1-1) \\ 1+x_2 & x_2(x_2-1) & \dots & x_2^{n-1}(x_2-1) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 1+x_n & x_n(x_n-1) & \dots & x_n^{n-1}(x_n-1) \end{vmatrix}.$$

En écrivant $1+x_i = 2x_i - (x_i-1)$ et en utilisant la linéarité du déterminant, on obtient

$$\det A = \begin{vmatrix} 2x_1 & x_1(x_1-1) & \dots & x_1^{n-1}(x_1-1) \\ 2x_2 & x_2(x_2-1) & \dots & x_2^{n-1}(x_2-1) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 2x_n & x_n(x_n-1) & \dots & x_n^{n-1}(x_n-1) \end{vmatrix}$$

$$- \begin{vmatrix} x_1-1 & x_1(x_1-1) & \dots & x_1^{n-1}(x_1-1) \\ x_2-1 & x_1(x_2-1) & \dots & x_2^{n-1}(x_2-1) \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ x_n-1 & x_n(x_n-1) & \dots & x_n^{n-1}(x_n-1) \end{vmatrix}.$$

Dans le premier déterminant, après avoir factorisé par 2, on additionne successivement chaque colonne à la suivante et on obtient

$$2 \begin{vmatrix} x_1 & x_1^2 & \dots & x_n \\ x_2 & x_2^2 & \dots & x_2^n \\ \vdots & & & \vdots \\ x_n & x_n^2 & \dots & x_n^n \end{vmatrix}$$

qui est égal à $2x_1 \dots x_n$ fois le déterminant de Vandermonde $V(x_1, \dots, x_n)$. Dans le second déterminant, après avoir mis en facteur x_1-1, x_2-1, $\dots$, x_n-1 dans les lignes 1 à n, on obtient de nouveau le déterminant de Vandermonde. On a donc

$$\begin{aligned} \det A &= 2x_1 \dots x_n V(x_1, \dots, x_n) - (x_1-1)\dots(x_n-1)V(x_1, \dots, x_n) \\ &= \prod_{i<j}(x_j - x_i)\left[2x_1x_2\dots x_n - (x_1-1)(x_2-1)\dots(x_n-1)\right]. \quad \lhd \end{aligned}$$

1.15. Application du déterminant de Vandermonde (2)

Soient $X_1, \dots, X_n, Y_1, \dots, Y_n$ des indéterminées. Calculer, pour tout $0 \leqslant k \leqslant n-1$, $\det\left(((X_i - Y_j)^k)_{1 \leqslant i,j \leqslant n}\right)$.

(École polytechnique)

$\rhd$ **Solution.**

Notons $K = \mathbb{R}(X_1, \dots, X_n, Y_1, \dots, Y_n)$ le corps des fractions rationnelles réelles en les indéterminées considérées et $A \in \mathcal{M}_n(K)$ la matrice de terme général $(X_i - Y_j)^k$. On a.

$$\forall (i,j) \in [\![1,n]\!]^2,\ a_{ij} = (X_i - Y_j)^k = \sum_{\ell=0}^{k} C_k^\ell X_i^\ell (-Y_j)^{k-\ell}.$$

On reconnaît un produit de matrices. En effet, si on considère les matrices B de $\mathcal{M}_{n,k+1}(\mathrm{K})$ et C de $\mathcal{M}_{k+1,n}(\mathrm{K})$ définies par

$$\forall (i,j) \in [\![1,n]\!] \times [\![1,k+1]\!],\quad b_{ij} = C_k^{j-1} X_i^{j-1},$$
$$\forall (i,j) \in [\![1,k+1]\!] \times [\![1,n]\!],\quad c_{ij} = (-Y_j)^{k-i+1},$$

on a $A = BC$. On en déduit $\mathrm{rg}(A) \leqslant \mathrm{rg}(B) \leqslant k+1$. Si $k < n-1$, on a $\mathrm{rg}\, A < n$ donc $\det A = 0$. Lorsque $k = n-1$, B et C sont carrées et $\det(A) = \det(B)\det(C)$. On a $\det(B) = \left(\prod_{\ell=0}^{k} C_k^\ell\right) V(X_1, \ldots, X_n)$ et comme

$$(-1)^{0+1+\cdots+(n-1)} \begin{vmatrix} Y_1^{n-1} & \cdots & \cdots & Y_n^{n-1} \\ \vdots & & & \vdots \\ Y_1 & \cdots & \cdots & Y_n \\ 1 & \cdots & \cdots & 1 \end{vmatrix} = \begin{vmatrix} 1 & \cdots & \cdots & 1 \\ Y_1 & \cdots & \cdots & Y_n \\ \vdots & & & \vdots \\ Y_1^{n-1} & \cdots & \cdots & Y_n^{n-1} \end{vmatrix}$$

on a $\det C = V(Y_1, \ldots, Y_n)$. On en déduit

$$\boxed{\det(A) = \left(\prod_{\ell=0}^{k} C_k^\ell\right) V(X_1, \ldots, X_n)\, V(Y_1, \ldots, Y_n).} \quad \lhd$$

1.16. Application du déterminant de Vandermonde (3)

Soit $m < n$ deux entiers naturels, $P_1, \ldots, P_n$ des polynômes de $\mathbb{R}_m[X]$ et $(a_1, \ldots, a_n)$ des réels.

1. Calculer

$$\begin{vmatrix} P_1(a_1) & \cdots & P_1(a_n) \\ P_2(a_1) & \cdots & P_2(a_n) \\ \vdots & & \vdots \\ P_n(a_1) & \cdots & P_n(a_n) \end{vmatrix}.$$

À quelle condition le déterminant est-il non nul ?

2. En déduire pour $m < n$, la valeur de

$$\begin{vmatrix} 1^m & 2^m & \dots & n^m \\ 2^m & 3^m & \dots & (n+1)^m \\ \vdots & \vdots & & \vdots \\ n^m & (n+1)^m & \dots & (2n-1)^m \end{vmatrix}.$$

(École polytechnique)

▷ **Solution.**

1. Si la famille $(P_1, \dots, P_n)$ est liée, le déterminant est nul. C'est le cas en particulier si $n > m+1 = \dim \mathbb{R}_m[X]$. Supposons donc $n = m+1$.

• La première méthode est inspirée du calcul du déterminant de Vandermonde : si deux des a_i sont égaux, le déterminant est nul. Supposons donc que les a_i sont deux à deux distincts.

Les polynômes $P_1, \dots, P_n$ étant fixés, notons $D(a_1, \dots a_n)$ le déterminant cherché et posons $P(X) = D(a_1, \dots, a_{n-1}, X)$. Le polynôme P est de degré inférieur ou égal à $m = n-1$ et possède $n-1$ racines distinctes $a_1, \dots, a_{n-1}$. Il s'écrit $D_1(a_1, \dots, a_{n-1}) \prod_{i=1}^{n-1} (X - a_i)$, où D_1 est une fonction polynomiale. En particulier, on obtient

$$D(a_1, \dots, a_n) = D_1(a_1, \dots, a_{n-1}) \prod_{i=1}^{n-1} (a_n - a_i).$$

En réitérant le raisonnement avec a_{n-1} au lieu de a_n, on montre que $D_1(a_1, \dots, a_{n-1}) = D_2(a_1, \dots, a_{n-2}) \prod_{i=1}^{n-2} (a_{n-1} - a_i)$ et finalement, on obtient que

$$D(a_1, \dots, a_n) = KV(a_1, \dots, a_n),$$

où K ne dépend pas des a_i mais seulement des polynômes P_j.

Posons désormais $K = S(P_1, \dots, P_n)$. Il est clair que S est une forme n-linéaire alternée. En notant det le déterminant dans la base canonique de $\mathbb{R}_{n-1}[X]$, on obtient qu'il existe $k \in \mathbb{R}$ tel que, pour tous polynômes $P_1, \dots, P_n$,

$$D(a_1, \dots, a_n) = kV(a_1, \dots, a_n) \det(P_1, \dots, P_n).$$

Si $(P_1, \dots, P_n)$ est la base canonique de $\mathbb{R}_{n-1}[X]$, on a alors $D(a_1, \dots, a_n) = V(a_1, \dots, a_n)$, ce qui montre que $k = 1$ et

$$D(a_1, \dots, a_n) = V(a_1, \dots, a_n) \det(P_1, \dots, P_n).$$

Cette formule reste vraie si les a_i ne sont pas distincts.

Cela montre que le déterminant est nul si, et seulement si, $n > m+1$ ou $n = m+1$ et ou bien deux des a_i sont égaux ou bien la famille $(P_1, \dots, P_n)$ est liée.

• La deuxième méthode consiste à écrire la matrice des $P_i(a_j)$ comme un produit : on se place comme précédemment dans le cas $m = n-1$. Si on écrit pour $1 \leqslant i \leqslant n$, $P_i = c_{i,0} + c_{i,1}X + \cdots + c_{i,n-1}X^{n-1}$, on a

$$P_i(a_j) = c_{i,0}a_j^0 + c_{i,1}a_j^1 + \cdots + c_{i,n-1}a_j^{n-1},$$

si bien que la matrice des $P_i(a_j)$ apparaît comme le produit de la matrice $(c_{i,j})_{\substack{1\leqslant i\leqslant n\\ 0\leqslant j\leqslant n-1}}$ (qui n'est rien d'autre que la transposée de matrice de $(P_1, \dots, P_n)$ dans la base $(1, X, \dots, X^{n-1})$) par la matrice de Vandermonde $(a_j^{i-1})_{1\leqslant i,j\leqslant n}$. On en déduit que le déterminant cherché est $\det(P_1, \dots, P_n)V(a_1, \dots, a_n)$.

2. En posant $P_1 = X^m$, $P_2 = (X+1)^m, \dots, P_n = (X+n-1)^m$ et $a_i = i$ pour $1 \leqslant i \leqslant n$, le déterminant à calculer, que l'on note D_n, est

$$\begin{vmatrix} P_1(a_1) & \dots & P_1(a_n) \\ P_2(a_1) & \dots & P_2(a_n) \\ \vdots & & \vdots \\ P_n(a_1) & \dots & P_n(a_n) \end{vmatrix}.$$

On applique le résultat de la question 1. Si $n > m+1$, $D_n = 0$. On suppose $n = m+1$. On calcule

$$V(a_1, \dots, a_n) = V(1, \dots, n) = \prod_{j=2}^{n}\prod_{i=1}^{j-1}(j-i) = \prod_{j=2}^{n}(j-1)! = \prod_{j=1}^{n-1} j!$$

et $\det(P_1, \dots, P_n)$. Comme $P_j = \sum_{i=0}^{n-1} C_{n-1}^i (j-1)^{n-1-i}X^i$, on obtient, en mettant C_{n-1}^i en facteur dans la $(i+1)$-ième ligne,

$$\det(P_1, \dots, P_n) = C_{n-1}^0 C_{n-1}^1 \cdots C_{n-1}^{n-1} \begin{vmatrix} 0 & 1 & 2^{n-1} & \dots & (n-1)^{n-1} \\ 0 & 1 & 2^{n-2} & \dots & (n-1)^{n-2} \\ \vdots & & & & \vdots \\ 0 & 1 & 2 & \dots & (n-1) \\ 1 & 1 & 1 & \dots & 1 \end{vmatrix}$$

$$= \frac{(n-1)!^n}{(0!1!\dots(n-1)!)^2}(-1)^{\frac{n(n-1)}{2}} V(0, 1, \dots, n-1),$$

car pour retrouver le déterminant de Vandermonde, il faut effectuer la permutation $k \mapsto n+1-k$ sur les lignes et cette permutation a pour

signature $(-1)^{\frac{n(n-1)}{2}}$. Comme $V(0,1,\dots,n-1) = \prod_{j=1}^{n-1} j!$, le déterminant cherché vaut

$$\begin{aligned} D_n &= \prod_{j=1}^{n-1} j! \cdot \frac{(n-1)!^n}{(0!1!\dots(n-1)!)^2}(-1)^{\frac{n(n-1)}{2}} \cdot \prod_{j=1}^{n-1} j! \\ &= (-1)^{\frac{n(n-1)}{2}}(n-1)!^n, \end{aligned}$$

ce qui peut se détailler ainsi

$$D_n = \begin{cases} (n-1)!^n & \text{si } n \equiv 0 \text{ ou } 1 \mod 4 \\ -(n-1)!^n & \text{si } n \equiv 2 \text{ ou } 3 \mod 4. \end{cases} \triangleleft$$

La notion de discriminant d'un polynôme P *peut se définir à partir du résultant (voir l'exercice 5.10 du tome 1 d'algèbre pour cette notion) : il s'agit du résultant des polynômes* P *et* P′. *Le résultant de deux polynômes de* $\mathbb{C}[X]$ *est nul si et seulement si ils ont une racine commune. Ainsi la nullité du discriminant équivaut à l'existence d'une racine multiple. L'énoncé suivant introduit directement le discriminant d'un polynôme* $P \in \mathbb{C}[X]$ *et démontre que celui-ci est un polynôme à coefficients entiers des coefficients de* P. *L'originalité de la méthode employée consiste à se ramener au cas où* P *est le polynôme caractéristique d'une matrice.*

1.17. Discriminant

Soit $n \in \mathbb{N}^*$ et $P \in \mathbb{C}[X]$ unitaire de degré n. On pose

$$P(X) = \prod_{i=1}^{n}(X - z_i) \quad \text{et} \quad \operatorname{disc}(P) = (-1)^{n(n-1)/2} \prod_{1 \leqslant i,j \leqslant n,\ i \neq j} (z_i - z_j).$$

1. Montrer que $\operatorname{disc}(P) = (-1)^{n(n-1)/2} \prod_{i=1}^{n} P'(z_i)$.

2. Calculer $\operatorname{disc}(P)$ si $P(X) = X^2 + aX + b$.

3. Soit $A \in \mathcal{M}_n(\mathbb{C})$. On pose $B = \left(\operatorname{Tr}\left(A^{i+j-2}\right)\right)_{1 \leqslant i,j \leqslant n}$.
Montrer que $\operatorname{disc}(\chi_A) = \det B$.

4. Soit $P = X^n + a_{n-1}X^{n-1} + \cdots + a_0 \in \mathbb{C}[X]$.
Montrer que $\operatorname{disc}(P) \in \mathbb{Z}[a_0, \dots, a_{n-1}]$.

(École normale supérieure)

Solution.

1. On remarque que, par définition, disc(P) s'annule si P possède une racine multiple. On a $P' = \sum_{i=1}^{n} \prod_{j \neq i} (X - z_j)$ donc $P'(z_i) = \prod_{j \neq i} (z_i - z_j)$ pour tout i et

$$\operatorname{disc}(P) = (-1)^{n(n-1)/2} \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} (z_i - z_j) = (-1)^{n(n-1)/2} \prod_{i=1}^{n} P'(z_i).$$

On peut donner une autre expression de disc(P) *faisant disparaître le facteur* $(-1)^{n(n-1)/2}$. *On a en effet*

$$\prod_{1 \leqslant i,j \leqslant n,\ i \neq j} (z_i - z_j) = \prod_{1 \leqslant i < j \leqslant n} (z_i - z_j) \prod_{1 \leqslant j < i \leqslant n} (z_i - z_j)$$

et

$$\prod_{1 \leqslant j < i \leqslant n} (z_i - z_j) = \prod_{1 \leqslant i < j \leqslant n} (z_j - z_i) = (-1)^{n(n-1)/2} \prod_{1 \leqslant i < j \leqslant n} (z_i - z_j),$$

ce qui donne

$$\operatorname{disc}(P) = \prod_{1 \leqslant i < j \leqslant n} (z_i - z_j)^2.$$

2. Dans ce cas, on retrouve l'expression habituelle du discriminant. En notant z_1 et z_2 les racines de P, on obtient en effet

$$\begin{aligned} \operatorname{disc}(P) &= -P'(z_1)P'(z_2) = -(2z_1 + a)(2z_2 + a) \\ &= -(4z_1z_2 + 2a(z_1 + z_2) + a^2) = -(4b - 2a^2 + a^2) = a^2 - 4b. \end{aligned}$$

3. On note $z_1, z_2, \ldots, z_n$ les valeurs propres de A prises avec multiplicité. Alors, pour tout entier naturel k, les valeurs propres de A^k sont $z_1^k, z_2^k, \ldots, z_n^k$ et on a donc $\operatorname{Tr}(A^k) = \sum_{\ell=1}^{n} z_\ell^k$. Le coefficient d'indice (i, j) de la matrice B est donc

$$b_{ij} = \sum_{\ell=1}^{n} z_\ell^{i+j-2}.$$

En notant, pour $1 \leqslant \ell \leqslant n$, $C_\ell = \begin{pmatrix} 1 \\ z_\ell \\ \vdots \\ z_\ell^{n-1} \end{pmatrix}$, on observe alors que la j-ième colonne de la matrice B est $\sum_{\ell=1}^{n} z_\ell^{j-1} C_\ell$. On a donc

$$\det(B) = \det(C_1 + \cdots + C_n, z_1 C_1 + \cdots + z_n C_n, \ldots, z_1^{n-1} C_1 + \cdots + z_n^{n-1} C_n).$$

Si les z_i ne sont pas distincts, les vecteurs $C_1, \dots, C_n$ forment une famille liée. Il en est de même des colonnes de B donc $\det B = 0$.

Si les z_i sont distincts, la famille $(C_1, \dots, C_n)$ est libre, car son déterminant est le déterminant de Vandermonde $V(z_1, \dots, z_n)$. En notant $\mathcal{B}$ la base canonique de $\mathbb{C}^n$ et $\mathcal{C}$ la base $(C_1, \dots, C_n)$, on obtient, par la formule de changement de base

$$\begin{aligned}\det B &= \det_{\mathcal{B}}(C_1 + \cdots + C_n, \dots, z_1^{n-1}C_1 + \cdots + z_n^{n-1}C_n)\\ &= \det_{\mathcal{B}}(\mathcal{C}) \det_{\mathcal{C}}(C_1 + \cdots + C_n, \dots, z_1^{n-1}C_1 + \cdots + z_n^{n-1}C_n)\\ &= V(z_1, \dots, z_n)V(z_1, \dots, z_n) = V(z_1, \dots, z_n)^2 = \prod_{1 \leqslant i < j \leqslant n} (z_i - z_j)^2.\end{aligned}$$

On remarque que cette formule reste valable si les z_i ne sont pas distincts. Compte tenu de la remarque qui suit la question 1, on conclut que

$$\det(B) = \operatorname{disc}(\chi_A).$$

4. Soit $A = \begin{pmatrix} 0 & \dots & \dots & \dots & 0 & -a_0 \\ 1 & \ddots & & & \vdots & -a_1 \\ 0 & \ddots & \ddots & & \vdots & -a_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & 0 & -a_{n-2} \\ 0 & \dots & \dots & 0 & 1 & -a_{n-1} \end{pmatrix}$ la matrice compagnon du polynôme P. On a $\chi_A = \begin{vmatrix} X & \dots & \dots & \dots & 0 & a_0 \\ -1 & \ddots & & & \vdots & a_1 \\ 0 & \ddots & \ddots & & \vdots & a_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & X & a_{n-2} \\ 0 & \dots & \dots & 0 & -1 & a_{n-1} \end{vmatrix}$. En multipliant, pour $2 \leqslant i \leqslant n$, la i-ième ligne par X^{i-1} et en l'additionnant à la première, on obtient

$$\chi_A = \begin{vmatrix} 0 & \dots & \dots & \dots & 0 & P \\ -1 & \ddots & & & \vdots & a_1 \\ 0 & \ddots & \ddots & & \vdots & a_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ \vdots & & \ddots & \ddots & X & a_{n-2} \\ 0 & \dots & \dots & 0 & -1 & a_{n-1} \end{vmatrix} = (-1)^{n+1} P \begin{vmatrix} -1 & X & \dots & \dots & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & X \\ 0 & \dots & \dots & 0 & -1 \end{vmatrix}$$

soit $\chi_A = P$. La matrice A est à coefficients dans $\mathbb{Z}[a_0, a_1, \ldots, a_{n-1}]$. Il en est de même de A^k pour tout $k \in \mathbb{N}$ et donc de la matrice B associée à A comme dans la question précédente. Ainsi son déterminant, qui est disc(P), appartient à $\mathbb{Z}[a_0, a_1, \ldots, a_{n-1}]$. ◁

Le déterminant de Cauchy qui suit est utilisé dans des problèmes d'approximation. En effet, si on munit l'espace des fonctions continues sur $[0,1]$ *du produit scalaire* $\langle f, g\rangle = \int_0^1 f(t)g(t)\,dt$, *la matrice de Gram associée à une famille de fonctions puissances conduit à une matrice de Cauchy.*

1.18. Déterminant de Cauchy

Soient $(\alpha_1, \ldots, \alpha_n) \in \mathbb{C}^n$ et $(\beta_1, \ldots, \beta_n) \in \mathbb{C}^n$ tels que pour tout $(i,j) \in [\![1,n]\!]^2$, $\alpha_i + \beta_j \neq 0$. Montrer que

$$\det\left(\frac{1}{\alpha_i + \beta_j}\right)_{1\leqslant i,j\leqslant n} = \frac{\displaystyle\prod_{1\leqslant i<j\leqslant n}(\alpha_j - \alpha_i)(\beta_j - \beta_i)}{\displaystyle\prod_{1\leqslant i,j\leqslant n}(\alpha_i + \beta_j)}.$$

(École polytechnique)

▷ **Solution.**

On note D le déterminant à calculer. On multiplie chaque colonne C_j par $\alpha_n + \beta_j$, $1 \leqslant j \leqslant n$. On obtient

$$D = \frac{1}{(\alpha_n + \beta_1)\ldots(\alpha_n + \beta_n)} \begin{vmatrix} \dfrac{\alpha_n + \beta_1}{\alpha_1 + \beta_1} & \cdots & \cdots & \dfrac{\alpha_n + \beta_n}{\alpha_1 + \beta_n} \\ \vdots & & & \vdots \\ \dfrac{\alpha_n + \beta_1}{\alpha_{n-1} + \beta_1} & \cdots & \cdots & \dfrac{\alpha_n + \beta_n}{\alpha_{n-1} + \beta_n} \\ 1 & \cdots & \cdots & 1 \end{vmatrix}.$$

Dans ce dernier déterminant, le terme d'indice (i,j) est, pour $i < n$

$$\frac{\alpha_n + \beta_j}{\alpha_i + \beta_j} = 1 + \frac{\alpha_n - \alpha_i}{\alpha_i + \beta_j}.$$

En retranchant la dernière ligne à toutes les autres on obtient donc

$$\mathrm{D} = \frac{1}{(\alpha_n+\beta_1)\dots(\alpha_n+\beta_n)} \begin{vmatrix} \frac{\alpha_n-\alpha_1}{\alpha_1+\beta_1} & \dots & \dots & \frac{\alpha_n-\alpha_1}{\alpha_1+\beta_n} \\ \vdots & & & \vdots \\ \frac{\alpha_n-\alpha_{n-1}}{\alpha_{n-1}+\beta_1} & \dots & \dots & \frac{\alpha_n-\alpha_{n-1}}{\alpha_{n-1}+\beta_n} \\ 1 & \dots & \dots & 1 \end{vmatrix}$$

$$= \frac{(\alpha_n-\alpha_1)\dots(\alpha_n-\alpha_{n-1})}{(\alpha_n+\beta_1)\dots(\alpha_n+\beta_n)} \begin{vmatrix} \frac{1}{\alpha_1+\beta_1} & \dots & \dots & \frac{1}{\alpha_1+\beta_n} \\ \vdots & & & \vdots \\ \frac{1}{\alpha_{n-1}+\beta_1} & \dots & \dots & \frac{1}{\alpha_{n-1}+\beta_n} \\ 1 & \dots & \dots & 1 \end{vmatrix}$$

car $\alpha_n - \alpha_i$ est facteur de tous les coefficients de la i-ème ligne. Retranchons maintenant la dernière colonne à toutes les autres, on obtient

$$\mathrm{D} = \frac{(\alpha_n-\alpha_1)\dots(\alpha_n-\alpha_{n-1})}{(\alpha_n+\beta_1)\dots(\alpha_n+\beta_n)} \begin{vmatrix} \boxed{\left(\frac{\beta_n-\beta_j}{(\alpha_i+\beta_j)(\alpha_i+\beta_n)}\right)_{1\leqslant i,j\leqslant n-1}} & \times \\ 0 \quad \dots \quad 0 & 1 \end{vmatrix}$$

et en développant selon la dernière ligne

$$\mathrm{D} = \frac{(\alpha_n-\alpha_1)\dots(\alpha_n-\alpha_{n-1})}{(\alpha_n+\beta_1)\dots(\alpha_n+\beta_n)} \det\left(\frac{\beta_n-\beta_j}{(\alpha_i+\beta_j)(\alpha_i+\beta_n)}\right)_{1\leqslant i,j\leqslant n-1}.$$

Comme $(\beta_n - \beta_j)$ est en facteur dans la j-ième colonne et $(\alpha_i + \beta_n)$ dans la i-ième ligne, on a

$$\mathrm{D} = \frac{(\alpha_n-\alpha_1)\dots(\alpha_n-\alpha_{n-1})(\beta_n-\beta_1)\dots(\beta_n-\beta_{n-1})}{(\alpha_n+\beta_1)\dots(\alpha_n+\beta_n)(\alpha_1+\beta_n)\dots(\alpha_{n-1}+\beta_n)} \det\left(\frac{1}{\alpha_i+\beta_j}\right)_{1\leqslant i,j\leqslant n-1}$$

Le résultat découle alors d'une récurrence immédiate sur n.

Nous présentons maintenant une autre méthode : il s'agit de mettre en action les idées du calcul du déterminant de Vandermonde. Tout d'abord le déterminant est nul dès que deux α_i sont égaux (resp. deux β_j) puisque deux lignes (resp. deux colonnes) sont alors identiques. Nous supposerons maintenant les α_i (resp. les β_j) deux à deux distincts.

Posons

$$\mathrm{F}(\mathrm{X}) = \begin{vmatrix} \frac{1}{\alpha_1+\beta_1} & \frac{1}{\alpha_1+\beta_2} & \dots & \frac{1}{\alpha_1+\beta_n} \\ \vdots & \vdots & & \vdots \\ \frac{1}{\alpha_{n-1}+\beta_1} & \frac{1}{\alpha_{n-1}+\beta_2} & \dots & \frac{1}{\alpha_{n-1}+\beta_n} \\ \frac{1}{\mathrm{X}+\beta_1} & \frac{1}{\mathrm{X}+\beta_2} & \dots & \frac{1}{\mathrm{X}+\beta_n} \end{vmatrix} \in \mathbb{C}(\mathrm{X}).$$

En imaginant le développement de ce déterminant selon la dernière ligne, on convient que F est la somme de n fractions nulles ou de degré -1 et donc $\deg \mathrm{F} \leqslant -1$. Toujours avec ce développement, on peut réduire au même dénominateur pour mettre F sous la forme

$$\mathrm{F} = \frac{\mathrm{P}(\mathrm{X})}{(\mathrm{X}+\beta_1)(\mathrm{X}+\beta_2)\dots(\mathrm{X}+\beta_n)}.$$

On a donc $\deg \mathrm{P} \leqslant n-1$. Les α_i $(1 \leqslant i \leqslant n-1)$ ne sont pas pôles de F et $\mathrm{F}(\alpha_i) = 0$ (en substituant α_i à X, le déterminant présente deux lignes identiques). Donc $\mathrm{P}(\alpha_i) = 0$. Les α_i étant deux à deux distincts et $\deg \mathrm{P} \leqslant n-1$, il existe $\lambda \in \mathbb{R}$ avec

$$\mathrm{P} = \lambda(\mathrm{X}-\alpha_1)(\mathrm{X}-\alpha_2)\dots(\mathrm{X}-\alpha_{n-1}).$$

Il s'agit de calculer λ. Multiplions F par $\mathrm{X}+\beta_n$:

$$(\mathrm{X}+\beta_n)\mathrm{F} = \begin{vmatrix} \dfrac{1}{\alpha_1+\beta_1} & \cdots & \cdots & \dfrac{1}{\alpha_1+\beta_n} \\ \vdots & \vdots & & \vdots \\ \dfrac{1}{\alpha_{n-1}+\beta_1} & \cdots & \cdots & \dfrac{1}{\alpha_{n-1}+\beta_n} \\ \dfrac{\mathrm{X}+\beta_n}{\mathrm{X}+\beta_1} & \cdots & \dfrac{\mathrm{X}+\beta_n}{\mathrm{X}+\beta_{n-1}} & 1 \end{vmatrix}$$

$$= \lambda \frac{(\mathrm{X}-\alpha_1)\dots(\mathrm{X}-\alpha_{n-1})}{(\mathrm{X}+\beta_1)\dots(\mathrm{X}+\beta_{n-1})}$$

et évaluons en $-\beta_n$:

$$\begin{vmatrix} \dfrac{1}{\alpha_1+\beta_1} & \cdots & \dfrac{1}{\alpha_1+\beta_{n-1}} & \times \\ \vdots & & \vdots & \vdots \\ \dfrac{1}{\alpha_{n-1}+\beta_1} & \cdots & \dfrac{1}{\alpha_{n-1}+\beta_{n-1}} & \times \\ 0 & \cdots & 0 & 1 \end{vmatrix} = \lambda \frac{(-1)^{n-1}(\beta_n+\alpha_1)\dots(\beta_n+\alpha_{n-1})}{(-1)^{n-1}(\beta_n-\beta_1)\dots(\beta_n-\beta_{n-1})}.$$

En développant ce déterminant par rapport à la dernière ligne, il vient

$$1 \times \begin{vmatrix} \dfrac{1}{\alpha_1+\beta_1} & \cdots & \dfrac{1}{\alpha_1+\beta_{n-1}} \\ \vdots & & \vdots \\ \dfrac{1}{\alpha_{n-1}+\beta_1} & \cdots & \dfrac{1}{\alpha_{n-1}+\beta_{n-1}} \end{vmatrix} = \lambda \frac{(\beta_n+\alpha_1)\dots(\beta_n+\alpha_{n-1})}{(\beta_n-\beta_1)\dots(\beta_n-\beta_{n-1})}$$

d'où

$$\lambda = \begin{vmatrix} \dfrac{1}{\alpha_1+\beta_1} & \cdots & \dfrac{1}{\alpha_1+\beta_{n-1}} \\ \vdots & & \vdots \\ \dfrac{1}{\alpha_{n-1}+\beta_1} & \cdots & \dfrac{1}{\alpha_{n-1}+\beta_{n-1}} \end{vmatrix} \frac{(\beta_n-\beta_1)\dots(\beta_n-\beta_{n-1})}{(\beta_n+\alpha_1)\dots(\beta_n+\alpha_{n-1})}.$$

Le déterminant cherché, $\mathrm{F}(\alpha_n)$, est donc égal à

$$\begin{vmatrix} \frac{1}{\alpha_1+\beta_1} & \cdots & \frac{1}{\alpha_1+\beta_{n-1}} \\ \vdots & & \vdots \\ \frac{1}{\alpha_{n-1}+\beta_1} & \cdots & \frac{1}{\alpha_{n-1}+\beta_{n-1}} \end{vmatrix} \frac{\prod_{i<n}(\beta_n-\beta_i)\prod_{i<n}(\alpha_n-\alpha_i)}{\prod_{i<n}(\beta_n+\alpha_i)\prod_{i\leqslant n}(\alpha_i+\beta_n)}.$$

On établit par récurrence la formule du déterminant de Cauchy. ◁

1.19. Déterminant de Hürwitz

Calculer le déterminant

$$\begin{vmatrix} r_1 & a & \dots & a \\ b & r_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & a \\ b & \dots & b & r_n \end{vmatrix}$$

dans $\mathbb{R}$, puis dans un corps commutatif quelconque.

(École normale supérieure)

▷ **Solution.**

Notre démarche va consister à rajouter une indéterminée aux coefficients de la matrice. Posons

$$\mathrm{P}(\mathrm{X}) = \begin{vmatrix} r_1+\mathrm{X} & a+\mathrm{X} & \dots & a+\mathrm{X} \\ b+\mathrm{X} & r_2+\mathrm{X} & \ddots & \vdots \\ \vdots & \ddots & \ddots & a+\mathrm{X} \\ b+\mathrm{X} & \dots & b+\mathrm{X} & r_n+\mathrm{X} \end{vmatrix}.$$

Il s'agit d'un polynôme de degré inférieur ou égal à 1. Pour s'en convaincre, il suffit de retrancher la première colonne à toutes les autres avant de développer par rapport à cette colonne. Il suffit donc de connaître les valeurs de P en deux points distincts.

Supposons dans un premier temps $a \neq b$. On a on a $\mathrm{P}(-a) = \mathrm{Q}(a)$ et $\mathrm{P}(-b) = \mathrm{Q}(b)$, en posant $\mathrm{Q} = (r_1 - \mathrm{X})\dots(r_n - \mathrm{X})$. La valeur qui nous intéresse est $\mathrm{P}(0)$. Si on écrit $\mathrm{P}(\mathrm{X}) = \alpha\mathrm{X}+\beta$, on obtient $\mathrm{P}(-a) = -\alpha a+\beta$ et $\mathrm{P}(-b) = -\alpha b + \beta$ de sorte que

$$\mathrm{P}(0) = \beta = \frac{b\mathrm{P}(-a) - a\mathrm{P}(-b)}{b-a} = \boxed{\frac{b\mathrm{Q}(a) - a\mathrm{Q}(b)}{b-a}}$$

Il reste à traiter le cas où $a = b$. Dans le cas où le corps de base est $\mathbb{R}$ il suffit de faire tendre b vers a dans l'expression précédente. On obtient

$$\lim_{b\to a} \frac{b\mathrm{Q}(a) - a\mathrm{Q}(b)}{b-a} = \lim_{b\to a} \mathrm{Q}(a) - a\frac{\mathrm{Q}(b)-\mathrm{Q}(a)}{b-a} = \boxed{\mathrm{Q}(a) - a\mathrm{Q}'(a)}.$$

En fait ce résultat reste encore valable avec un corps quelconque : on retrouve un cas particulier du déterminant de l'exercice 1.6 (avec les notations de cet exercice il suffit en effet de prendre les b_k tous égaux à a et $a_k = r_k - a$ pour tout k). ◁

Le calcul de ce cas particulier a aussi été détaillé dans l'exercice 1.12 du tome algèbre 1 où la matrice de Hürwitz intervenait dans un problème combinatoire.

L'exercice suivant reprend avec une autre argumentation le résultat de l'exercice 6.28 du premier tome d'algèbre. Ce résultat est important et a de nombreuses applications. Outre celle que le lecteur trouvera dans l'exercice 6.28 précité, on peut par exemple en déduire que si V *est un sous-espace de dimension finie de* $\mathcal{C}^0(\mathbb{R},\mathbb{R})$*, alors la convergence simple d'une suite de* V *implique la convergence uniforme de cette suite sur tout segment et la limite est dans* V.

1.20. Liberté d'une famille de fonctions

Soit X un ensemble, K un corps commutatif et n fonctions $f_1, f_2, \ldots, f_n$ de X dans K. Montrer que les applications $f_1, f_2, \ldots, f_n$ sont libres dans $\mathcal{F}(\mathrm{X},\mathrm{K})$ si, et seulement si, il existe des éléments $x_1, x_2, \ldots, x_n$ de X tel que le déterminant de $(f_i(x_j))_{1\leqslant i,j\leqslant n}$ est non nul.

(École polytechnique)

▷ **Solution.**

Si on suppose la famille $(f_1, \ldots, f_n)$ liée, il existe des scalaires $\lambda_1, \lambda_2, \ldots, \lambda_n$ non tous nuls tels que $\lambda_1 f_1 + \lambda_2 f_2 + \cdots + \lambda_n f_n = 0$. Si on évalue en x_j, il vient $\lambda_1 f_1(x_j) + \lambda_2 f_2(x_j) + \cdots + \lambda_n f_n(x_j) = 0$, ce qui donne une relation de liaison sur les lignes de la matrice $(f_i(x_j))_{1\leqslant i,j\leqslant n}$. Son déterminant est donc nul.

Montrons maintenant par récurrence sur $n \geqslant 1$ que si $(f_1, f_2, \ldots, f_n)$ est libre, il existe $x_1, x_2, \ldots, x_n \in \mathrm{X}$ tels que le déterminant de $(f_i(x_j))_{1\leqslant i,j\leqslant n}$ soit non nul. Si $n = 1$ et (f_1) est libre, $f_1 \neq 0$ et il existe donc x_1 tel que $f_1(x_1) \neq 0$. Supposons $n \geqslant 2$ et les f_i libres. La sous-famille $(f_1, \ldots, f_{n-1})$ est donc libre et par hypothèse de récurrence, il existe $(x_1, \ldots, x_{n-1}) \in \mathrm{X}^{n-1}$ tel que $\det(f_i(x_j))_{1\leqslant i,j\leqslant n-1}$ est non nul. Considérons, pour $x \in \mathrm{X}$, le déterminant

$$\Phi(x) = \begin{vmatrix} f_1(x_1) & \cdots & f_1(x_{n-1}) & f_1(x) \\ \vdots & & & \vdots \\ f_{n-1}(x_1) & \cdots & f_{n-1}(x_{n-1}) & f_{n-1}(x) \\ f_n(x_1) & \cdots & f_n(x_{n-1}) & f_n(x) \end{vmatrix}.$$

En développant $\Phi(x)$ par rapport à la dernière colonne, on obtient

$$\Phi(x) = \mathrm{D}_1 f_1(x) + \cdots + \mathrm{D}_{n-1} f_{n-1}(x) + \mathrm{D}_n f_n(x),$$

avec D_i cofacteur de $f_i(x)$. La fonction Φ s'écrit

$$\Phi = \mathrm{D}_1 f_1 + \cdots + \mathrm{D}_{n-1} f_{n-1} + \mathrm{D}_n f_n.$$

Comme $\mathrm{D}_n = \det(f_i(x_j))_{1 \leqslant i,j \leqslant n-1} \neq 0$ et les f_i sont libres, Φ ne peut être identiquement nulle (car sinon on aurait une relation de liaison). Il existe donc x_n tel que $\Phi(x_n) \neq 0$. Le n-uplet $(x_1, \ldots, x_{n-1}, x_n)$ répond à la question. ◁

Deux matrices semblables ont le même déterminant. La réduction est donc un outil puissant qui peut intervenir dans le calcul des déterminants.

1.21. Matrice circulante

Soit p premier et $(a_0, a_1, \ldots, a_{p-1}) \in \mathbb{Z}^p$. Montrer que

$$\begin{vmatrix} a_0 & a_1 & a_2 & \ldots & a_{p-1} \\ a_{p-1} & a_0 & a_1 & \ldots & a_{p-2} \\ a_{p-2} & a_{p-1} & a_0 & \ldots & a_{p-3} \\ \vdots & & & \ddots & \vdots \\ a_1 & a_2 & a_3 & \ldots & a_0 \end{vmatrix} \equiv a_0 + a_1 + \cdots + a_{p-1} \quad [p].$$

(École polytechnique)

▷ **Solution.**

La disposition des a_k dans la matrice proposée invite à introduire la matrice de permutation J correspondant au p-cycle $(p, p-1, \ldots, 1)$:

$$\mathrm{J} = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & & \vdots \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & 0 & 1 \\ 1 & 0 & \ldots & 0 & 0 \end{pmatrix}.$$

On a $\mathrm{J}^p = \mathrm{I}$ et $\mathrm{A} = a_0\mathrm{I} + a_1\mathrm{J} + a_2\mathrm{J}^2 + \cdots + a_{p-1}\mathrm{J}^{p-1}$. Notons $\overline{\mathrm{A}}$ la matrice de $\mathcal{M}_p(\mathbb{Z}/p\mathbb{Z})$ obtenue à partir de A en remplaçant les a_i par leurs classes

modulo p. Le déterminant de $\overline{A}$ est alors égal à la classe modulo p du déterminant de A. On est donc ramené à prouver que $\det \overline{A} = \sum_{i=0}^{p-1} \overline{a_i}$.

On a $\overline{A} = \sum_{k=0}^{p-1} \overline{a_k}\overline{J}^k$. On va trigonaliser la matrice $\overline{J}$. Son polynôme caractéristique s'obtient en développant selon la première colonne,

$$\chi_{\overline{J}}(X) = (X^p - 1) = (X-1)^p \in \mathbb{Z}/p\mathbb{Z}[X]$$

car il est bien connu que p divise C_p^k pour $1 \leqslant k \leqslant p-1$ (cf. exercice 4.25 du tome algèbre 1). Comme $\chi_{\overline{J}}$ est scindé, $\overline{J}$ est trigonalisable, semblable à une matrice triangulaire supérieure dont les termes diagonaux sont tous égaux à $\overline{1}$. Il en résulte que $\overline{A}$ est semblable à une matrice triangulaire supérieure ayant pour coefficients diagonaux $\sum_{k=0}^{p-1} \overline{a_k}$. Ainsi,

$$\det \overline{A} = \left(\sum_{k=0}^{p-1} \overline{a_k}\right)^p = \sum_{k=0}^{p-1} \overline{a_k}^p = \sum_{k=0}^{p-1} \overline{a_k}$$

la dernière égalité résultant du petit théorème de Fermat. D'où le résultat demandé. ◁

On s'intéressera de manière plus générale à la réduction des matrices circulantes dans le second chapitre (exercices 2.22 et 2.23).

Le thème des matrices stochastiques est un sujet riche qui sera largement développé dans le chapitre réduction (exercices 2.9 et suivants). L'énoncé qui vient propose d'étudier la borne supérieure du déterminant sur l'ensemble des matrices stochastiques.

1.22. Majoration du déterminant d'une matrice stochastique

Soit $M = (m_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ telle que $m_{ij} \geqslant 0$ pour tout couple (i,j) et $\sum_{j=1}^{n} m_{ij} = 1$ pour tout $i \in [\![1,n]\!]$.

1. Montrer que $|\det M| \leqslant 1$.

2. Étudier le cas d'égalité.

(École polytechnique)

▷ **Solution.**

1. Plusieurs approches sont possibles pour cette question. On peut par exemple montrer que toutes les valeurs propres complexes d'une matrice stochastique sont de module $\leqslant 1$ comme cela est fait dans l'exercice

2.9 du chapitre diagonalisation. Cela implique le résultat puisque le déterminant est le produit des valeurs propres. On va ici montrer le résultat par récurrence sur n, le cas $n = 1$ étant évident. Soit $M = (m_{ij})_{1 \leqslant i,j \leqslant n}$ dans $\mathcal{M}_n(\mathbb{R})$ vérifiant les hypothèses. Pour se ramener à des matrices de taille $n-1$ on développe le déterminant de M selon la première ligne. On a

$$\det M = \sum_{j=1}^{n} m_{1j}(-1)^{1+j} \det M_{1j}$$

où M_{1j} est extraite de M en ôtant la première ligne et la j-ième colonne. Les coefficients de ces matrices M_{1j} sont bien positifs mais la somme des coefficients de chaque ligne ne vaut plus nécessairement 1 : ces sommes sont toutes $\leqslant 1$. Ce petit problème nous empêche d'appliquer l'hypothèse de récurrence à ces matrices. En fait, on s'en sort très facilement en démontrant un résultat plus fort ! On prend comme propriété :

(H_n) : *toute matrice* $M \in \mathcal{M}_n(\mathbb{R})$ *dont les coefficients sont positifs ou nuls et telle que la somme des coefficients de chaque ligne est inférieure ou égale à* 1 *vérifie* $|\det M| \leqslant 1$.

L'assertion (H_1) est toujours claire. Si on suppose (H_{n-1}) vraie, et si $M = (m_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ vérifie les hypothèses de (H_n), on a, par le développement ci-dessus,

$$|\det M| \leqslant \sum_{j=1}^{n} m_{1j} |\det M_{1j}| \leqslant \sum_{j=1}^{n} m_{1j} \leqslant 1$$

car toutes les matrices extraites M_{1j} vérifient les hypothèses de (H_{n-1}). Le résultat est donc prouvé.

2. Étudions le cas d'égalité.

Pour $n = 1$ la seule matrice qui réalise l'égalité est la matrice identité.

On observe dans la majoration réalisée ci-dessus que, s'il y a égalité, tous les cofacteurs des coefficients m_{1j} non nuls doivent valoir 1 en valeur absolue et que la somme des coefficients de la première ligne doit valoir 1. En développant selon une autre ligne, on a plus généralement le cofacteur de chaque coefficient non nul qui vaut 1 en valeur absolue, et la somme des coefficients de chaque ligne doit faire 1, c'est-à-dire que la matrice est stochastique.

Regardons alors ce qui se passe pour une matrice $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ de taille $(2,2)$. Si $a > 0$ on a donc $|d| = 1$ et donc $d = 1$ puisque les coefficients sont positifs. On a alors $c = 0$ car $c + d = 1$. De $|\det M| = 1$, on tire $a = 1$ puis $b = 0$: la matrice M est la matrice identité. Lorsque $a = 0$, on voit de même que $M = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. On a alors l'intuition du résultat suivant : les seules matrices stochastiques M telles que $|\det M| = 1$ sont les matrices de permutation (*qui conviennent clairement*).

Supposons cela vrai au rang $n-1$ et soit $\mathrm{M} = (m_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{R})$ une matrice qui réalisé l'égalité au rang n. L'un au moins des coefficients de la première ligne est non nul. Quitte à permuter les colonnes de M on peut supposer sans perte de généralité que $m_{11} > 0$. La matrice extraite $\mathrm{M}' = (m_{ij})_{2\leqslant i,j\leqslant n}$ est de taille $n-1$ et doit vérifier $|\det \mathrm{M}'| = 1$ d'après ce qui est dit plus haut. Par hypothèse de récurrence il s'agit d'une matrice de permutation. Comme la somme des coefficients de chaque ligne de M vaut 1 on a nécessairement $m_{21} = m_{31} = \cdots = m_{n1} = 0$. Ainsi, si on développe selon la première colonne, $\det \mathrm{M} = m_{11} \det \mathrm{M}'$. Cela impose donc $m_{11} = 1$ et par suite $m_{12} = \cdots = m_{1n} = 0$. La matrice M est une matrice de permutation. Cela termine la récurrence et prouve le résultat annoncé. ◁

La comatrice est le thème des exercices suivants. Le seul résultat vraiment important à retenir est la relation

$$^t\mathrm{Com}\,\mathrm{A}\,\mathrm{A} = \mathrm{A}\,{}^t\mathrm{Com}\,\mathrm{A} = (\det \mathrm{A})\mathrm{I}_n$$

valable pour toute matrice $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$. *On en déduit notamment le rang de la comatrice en fonction du rang de* A. *Si* A *est inversible i.e. de rang* n, *on a* $\det \mathrm{A} \neq 0$ *et la comatrice de* A *est inversible. Si* $\mathrm{rg}\,\mathrm{A} < n-1$, *la comatrice de* A *est nulle, car tous les mineurs de taille* $n-1$ *extraits de* A *sont nuls. Dans le cas où* $\mathrm{rg}\,\mathrm{A} = n-1$, *la comatrice n'est pas nulle. Mais la relation* $^t\mathrm{Com}\,\mathrm{A}\,\mathrm{A} = 0$ *montre que l'image de* A *est incluse dans le noyau de* $^t\mathrm{Com}\,\mathrm{A}$. *On a donc dans ce cas* $\mathrm{rg}\,{}^t\mathrm{Com}\,\mathrm{A} = \mathrm{rg}\,\mathrm{Com}\,\mathrm{A} = 1$. *On se rendra compte à travers les exercices suivants qu'on est souvent amené à discuter selon ces trois cas.*

1.23. Calcul d'une trace

Soit A, B dans $\mathcal{M}_n(\mathbb{R})$ et X un vecteur non nul de $\mathbb{R}^n$. On suppose que $\mathrm{AX} = 0$ et qu'il existe $\mathrm{Y} \in \mathbb{R}^n$ tel que $\mathrm{AY} = \mathrm{BX}$. On note A_j la matrice obtenue en remplaçant la j-ième colonne de A par celle de B. Montrer que $\sum_{j=1}^{n} \det \mathrm{A}_j = 0$.

(École normale supérieure)

▷ **Solution.**

Interprétons d'abord le déterminant de A_j. Notons Δ_{ij} les cofacteurs de la matrice A et (b_{ij}) les coefficients de B. Si on développe le déterminant de A_j par rapport à la j-ième colonne on obtient

$$\det \mathrm{A}_j = \sum_{i=1}^{n} b_{ij}\Delta_{ij}.$$

Cette somme s'interprète comme le coefficient d'indice (j,j) de la matrice produit $\widetilde{A}B$ où $\widetilde{A}$ désigne la transposée de la comatrice de A. La somme que l'on demande d'estimer n'est donc autre que $\mathrm{Tr}(\widetilde{A}B)$.

La matrice A n'est pas inversible par hypothèse donc $\mathrm{rg}(A) \leqslant n-1$. Si $\mathrm{rg}(A) < n-1$, le rang étant la taille maximale des sous-matrices de A inversibles, tous les mineurs d'ordre $n-1$ sont nuls et la comatrice est nulle. Le résultat s'en déduit. Supposons donc que $\mathrm{rg}\,A = n-1$. Son noyau est alors la droite engendrée par X. On sait que $A\widetilde{A} = (\det A)I_n = 0$. En particulier $\mathrm{Im}\,\widetilde{A} \subset \mathrm{Ker}\,A$ de sorte que $\widetilde{A}$ est de rang au plus 1 (en fait de rang 1 exactement car A admet au moins un mineur de taille $n-1$ non nul).

Posons $M = \widetilde{A}B$. On a $\mathrm{rg}(M) \leqslant 1$ et $\mathrm{Im}(M) \subset \mathrm{Ker}\,A = \mathbb{R}.X$ vu ce qui précède. Si $M = 0$ on a fini. Sinon, on prend une base $(e_1, \ldots, e_{n-1})$ du noyau de M que l'on complète par e_n en une base de $\mathbb{R}^n$. La matrice de l'endomorphisme canoniquement associé à M dans la base $(e_1, \ldots, e_n)$ est de la forme

$$\begin{pmatrix} 0 & 0 & \ldots & 0 & \times \\ 0 & 0 & & 0 & \times \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & & 0 & \times \\ 0 & 0 & \ldots & 0 & d \end{pmatrix}.$$

On en déduit que $d = \mathrm{Tr}\,M = 0$ si, et seulement si, $\mathrm{Im}\,M \subset \mathrm{Ker}\,M$. Or justement, BX est dans l'image de A donc dans le noyau de $\widetilde{A}$ et $MX = 0$. D'où le résultat. ◁

Dans les deux exercices suivants et dans bien d'autres de cet ouvrage, on utilise la densité des matrices inversibles dans $\mathcal{M}_n(\mathbb{K})$ lorsque $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$: pour $A \in \mathcal{M}_n(\mathbb{K})$, la fonction polynôme $\lambda \longmapsto \det(A - \lambda I_n)$ n'est pas identiquement nulle donc a des zéros isolés. Pour λ assez petit non nul, $A - \lambda I_n$ est inversible. Pour $k \in \mathbb{N}^*$ assez grand, la matrice $A_k = A - \dfrac{1}{k} I_n$ est inversible et la suite (A_k) converge vers A.

La densité du groupe linéaire permet d'étendre par continuité à toute matrice de $\mathcal{M}_n(\mathbb{K})$ des propriétés valables pour les matrices inversibles. Le lecteur pourra se reporter au dernier chapitre sur la topologie des espaces de matrices pour d'autres résultats de même nature.

1.24. Spectre de la comatrice

Soit $A \in \mathcal{M}_n(\mathbb{C})$. On note $\widetilde{A}$ la transposée de la comatrice de A. Quelles sont les valeurs propres de $\widetilde{A}$?

(École polytechnique)

▷ **Solution.**

On distingue trois cas selon le rang de A.

• Si A est inversible, la relation $\widetilde{A}A = A\widetilde{A} = (\det A)I_n$ montre que $\widetilde{A} = (\det A)A^{-1}$. Le spectre de la $\widetilde{A}$ est donc

$$\left\{\frac{\det A}{\lambda}, \lambda \in \mathrm{Sp}(A)\right\}.$$

• Si $\mathrm{rg}\, A < n-1$, tous les mineurs d'ordre $n-1$ de A sont nuls et la comatrice de A est nulle. Le spectre de $\widetilde{A}$ est donc réduit à $\{0\}$.

• Il reste le cas $\mathrm{rg}\, A = n-1$: il existe alors un mineur d'ordre $n-1$ non nul, puisque le rang est la taille maximale des sous-matrices carrées inversibles et donc $\mathrm{rg}\, A \geqslant 1$. D'autre part, $A\widetilde{A} = (\det A)I_n = 0$. Donc $\mathrm{Im}\, \widetilde{A} \subset \mathrm{Ker}\, A$ qui, d'après le théorème du rang, est une droite de $\mathbb{C}^n$. Il s'ensuit que $\mathrm{Im}\, \widetilde{A} = \mathrm{Ker}\, A$ et donc $\dim \mathrm{Ker}\, \widetilde{A} = n-1$ (toujours en vertu du théorème du rang). Donc 0 est valeur propre de $\widetilde{A}$ avec une multiplicité au moins égale à $n-1$. Nous savons que la somme des valeurs propres comptées avec multiplicité est égale à la trace. On en déduit donc que le spectre de la $\widetilde{A}$ est $\{0, d\}$ avec $d = \mathrm{Tr}\, \widetilde{A}$. Intéressons-nous à $\mathrm{Tr}\, \widetilde{A}$.

Revenons au cas A inversible : notons $\lambda_1, \ldots, \lambda_n$ les racines du polynôme caractéristique qui sont donc non nulles. Les scalaires $\dfrac{\det A}{\lambda_i}$ sont les valeurs propres distinctes ou confondues de $\widetilde{A}$. Donc sa trace est

$$\mathrm{Tr}\, \widetilde{A} = \det A \sum_{i=1}^{n} \frac{1}{\lambda_i} = \sum_{i=1}^{n} \prod_{j \neq i} \lambda_j = \sigma_{n-1}(A)$$

où $\sigma_{n-1}(A)$ désigne l'expression symétrique élémentaire des produits de $n-1$ facteurs parmi les racines distinctes ou confondues de χ_A. En posant $\chi_A = X^n + a_{n-1}X^{n-1} + \cdots + a_1 X + a_0$, on obtient $\sigma_{n-1}(A) = (-1)^{n-1}a_1$. Puisque $\mathrm{GL}_n(\mathbb{C})$ est dense dans $\mathcal{M}_n(\mathbb{C})$, que $A \longmapsto a_1$ est continue, car a_1 est une fonction polynomiale des coefficients de A, et que l'application $A \longmapsto \mathrm{Tr}\, \widetilde{A}$ est également continue, la relation $\mathrm{Tr}\, \widetilde{A} = (-1)^{n-1}a_1$ reste valable pour une matrice non inversible.

Dans le cas où A est de rang $n-1$, le spectre de la comatrice est $\{0, (-1)^{n-1}a_1\}$ où a_1 est le coefficient de X dans χ_A. Si 0 est une racine simple de χ_A, il s'agit simplement du produit des $n-1$ autres racines. Sinon $\widetilde{A}$ ne possède qu'une valeur propre 0. ◁

Voici une autre solution possible. Il découle de l'exercice 1.26 que si A *et* B *sont deux matrices semblables il en est de même de* Com A *et* Com B. *Il suffit donc de trigonaliser* A *et de calculer la comatrice d'une matrice triangulaire supérieure pour retrouver le fait que si* $\lambda_1, \ldots, \lambda_n$ *sont les valeurs propres de* A *prises avec multiplicité, alors les valeurs propres de* Com A *sont les* $\mu_k = \prod_{i \neq k} \lambda_i$ *pour* $1 \leqslant k \leqslant n$.

1.25. La transposée de la comatrice de A est un polynôme en A

Soit $A \in \mathcal{M}_n(K)$. On note $\widetilde{A}$ la transposée de la comatrice de A. Montrer que $\widetilde{A}$ est un polynôme en A. On traitera successivement les cas $K = \mathbb{R}$, K infini, K quelconque.

(École normale supérieure)

▷ **Solution.**

• Comme nous y invite l'énoncé, prenons d'abord $K = \mathbb{R}$. D'après le théorème de Cayley-Hamilton, si $\chi_A = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, on a

$$\chi_A(A) = A^n + a_{n-1}A^{n-1} + \cdots + a_1A + a_0I_n = 0.$$

Supposons A inversible. Alors $\widetilde{A} = \det(A)A^{-1}$. Comme

$$A(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n) = -a_0I_n = -\det(A)I_n,$$

il vient

$$\widetilde{A} = \det(A)A^{-1} = -(A^{n-1} + a_{n-1}A^{n-2} + \cdots + a_2A + a_1I_n) \in \mathbb{R}[A].$$

Le résultat est donc vrai pour une matrice inversible.

Nous allons étendre la propriété sur $\mathcal{M}_n(\mathbb{R})$ par densité du groupe linéaire $GL_n(\mathbb{R})$. Soit $(A_p)_{p\in\mathbb{N}}$ une suite de matrices inversibles qui converge vers A. Pour $M \in \mathcal{M}_n(\mathbb{R})$, écrivons

$$\chi_M(X) = X^n + a_{n-1}(M)X^{n-1} + \cdots + a_1(M)X + a_0(M).$$

Les a_i apparaissent comme des fonctions polynomiales des coefficients de M. En particulier, ce sont des fonctions continues et par conséquent $\lim_{p\to+\infty} a_i(A_p) = a_i(A)$. D'autre part, $M \longmapsto \widetilde{M}$ est continue car polynomiale. On a, pour tout p,

$$\widetilde{A_p} = -(A_p^{n-1} + a_{n-1}(A_p)A_p^{n-2} + \cdots + a_2(A_p)A_p + a_1(A_p)I_n)$$

ce qui donne en faisant tendre p vers l'infini :

$$\widetilde{A} = -(A^{n-1} + a_{n-1}(A)A^{n-2} + \cdots + a_2(A)A_p + a_1(A)I_n) \in \mathbb{R}[A].$$

Le résultat est donc prouvé pour $K = \mathbb{R}$.

• Supposons maintenant K infini. Comme précédemment, pour M dans $\mathcal{M}_n(\mathbb{R})$, écrivons

$$\chi_M(X) = X^n + a_{n-1}(M)X^{n-1} + \cdots + a_1(M)X + a_0(M)$$

avec les a_i fonctions polynomiales des coefficients de M. Soit $\lambda \in K$. Si λ n'est pas valeur propre de A, $A - \lambda I_n$ est inversible et le même calcul que dans le cas réel assure :

$$\widetilde{(A - \lambda I_n)} = -((A-\lambda I)^{n-1} + a_{n-1}(A-\lambda I)(A-\lambda I)^{n-2} + \cdots + a_1(A-\lambda I)I).$$

Il s'agit là d'une identité polynomiale en λ, vraie pour λ en dehors du spectre de A *i.e.* pour une infinité de valeurs. Le polynôme

$$\widetilde{(A - XI_n)} + (A - XI)^{n-1} + a_{n-1}(A - XI)(A - XI)^{n-2} + \cdots + a_1(A - XI)I$$

est donc nul. En $X = 0$, cela donne

$$\widetilde{A} = -(A^{n-1} + a_{n-1}(A)A^{n-2} + \cdots + a_2(A)A_p + a_1(A)I_n) \in K[A].$$

Le résultat est donc vrai pour K infini.

• Reste à traiter le cas où le corps K est fini. Nous avons démontré que lorsque K est infini, $\widetilde{A}$ est combinaison linéaire des A^i avec des coefficients qui sont au signe près les coefficients du polynôme caractéristique de A. Si le corps K est fini, on le plonge dans un surcorps infini, par exemple $L = K(Y)$. Alors $\widetilde{A}$ est combinaison linéaire des A^i à coefficients dans L. Plus précisément, ces coefficients sont au signe près les coefficients du polynôme caractéristique de A, donc des éléments de K. La transposée de la comatrice est bien un polynôme en A à coefficients dans K. ◁

En particulier, toute matrice qui commute avec A *commute avec* $\widetilde{A}$. *Cette question a été posée indépendamment lors d'un autre oral à l'École normale supérieure.*

1.26. Comatrice d'un produit

Soit R un anneau de caractéristique différente de 2. Pour A dans $\mathcal{M}_n(R)$ on note Com A la comatrice de A.

Montrer que Com(AB) = Com A Com B. Avant d'aborder le cas général, on traitera successivement les cas suivants :

1. R est le corps des réels ou des complexes ;

2. R est un corps quelconque ;

3. R est un anneau intègre.

(École normale supérieure)

▷ **Solution.**

1. On sait que, si A est inversible et si R est un corps, on a l'égalité $\text{Com A} = (\det A)\,{}^tA^{-1}$. Ainsi dans le cas où A et B sont toutes deux inversibles,

$$\begin{aligned}\operatorname{Com}(AB) &= \det(AB)\,{}^t(AB)^{-1} = (\det A \det B)({}^tB\,{}^tA)^{-1}\\ &= \det A({}^tA)^{-1}\det B({}^tB)^{-1} = \operatorname{Com} A \operatorname{Com} B.\end{aligned}$$

Dans le cas $R = \mathbb{R}$ (resp. $R = \mathbb{C}$), comme $GL_n(\mathbb{R})$ (resp. $GL_n(\mathbb{C})$) est dense dans $\mathcal{M}_n(\mathbb{R})$ (resp. $\mathcal{M}_n(\mathbb{C})$) et que l'application qui à A associe Com A est continue, le résultat demeure pour toutes matrices A et B de $\mathcal{M}_n(R)$, par passage à la limite.

2. Si R est un corps quelconque on ne peut plus invoquer un argument topologique pour étendre la formule aux matrices non inversibles. On plonge alors $\mathcal{M}_n(R)$ dans $\mathcal{M}_n(R(X))$, où R(X) est le corps des fractions rationnelles à une indéterminée sur R. Les matrices $A - XI$ et $B - XI$ sont inversibles dans $\mathcal{M}_n(R(X))$ (car $\det(A - XI) = \chi_A(X) \neq 0$). D'après la question précédente,

$$\operatorname{Com}((A - XI)(B - XI)) = \operatorname{Com}(A - XI)\operatorname{Com}(B - XI)$$

Cette égalité matricielle équivaut à l'égalité de n^2 polynômes formels en X. Si on évalue ces polynômes en 0 on obtient l'égalité de tous les coefficients des matrices Com(AB) et Com A Com B.

3. Si R est un anneau intègre, on le plonge dans son corps des fractions pour se ramener à la question précédente. ◁

Si R *est un anneau commutatif quelconque, le résultat est encore vrai. Soit* $A = (a_{ij})$ *et* $B = (b_{ij})$ *deux matrices de* $\mathcal{M}_n(R)$. *Soient* $(X_{ij})_{1\leqslant i,j\leqslant n}$ *et* $(Y_{ij})_{1\leqslant i,j\leqslant n}$ $2n^2$ *indéterminées. Les matrices* $M = (X_{ij})_{1\leqslant i,j\leqslant n}$ *et* $N = (Y_{ij})_{1\leqslant i,j\leqslant n}$ *sont à coefficients dans l'anneau intègre* R′ *des polynômes en* X_{ij} *et* Y_{ij} *à coefficients dans* $\mathbb{Z}$. *D'après ce qui précède on a* $\operatorname{Com}(MN) - \operatorname{Com} M \operatorname{Com} N = 0$. *Par la propriété universelle de* R′ *il existe un unique morphisme d'anneau* ψ *de* R′ *dans* R *tel que* $\psi(X_{ij}) = a_{ij}$ *et* $\psi(Y_{ij}) = b_{ij}$ *pour tout* (i, j). *L'identité* $\operatorname{Com}(MN) - \operatorname{Com} M \operatorname{Com} N = 0$ *équivaut à la nullité de* n^2 *polynômes de* R′. *Les images de ces polynômes par* ψ *sont également nulles, ce qui conduit à la nullité de tous les coefficients de la matrice* $\operatorname{Com}(AB) - \operatorname{Com} A \operatorname{Com} B$.

On pourra utiliser le résultat qui vient d'être démontré dans l'exercice suivant.

1.27. Équation matricielle faisant intervenir la comatrice

Soit $n \geqslant 2$ un entier. Déterminer les matrices $A \in \mathcal{M}_n(\mathbb{C})$ telles que $A + \widetilde{A}$ soit scalaire, où $\widetilde{A}$ est la transposée de la comatrice de A.
(**École normale supérieure, École polytechnique**)

▷ **Solution.**

Notons pour tout $\lambda \in \mathbb{C}$, E_λ l'ensemble des matrices $A \in \mathcal{M}_n(\mathbb{C})$ telles que $A + \widetilde{A} = \lambda I_n$. Commençons par une observation fondamentale : si A est dans E_λ, toute la classe de similitude de A est incluse dans E_λ. En effet, si $B = P^{-1}AP$ avec $P \in GL_n(\mathbb{C})$, on a, d'après l'exercice 1.26,

$$\begin{aligned}\widetilde{B} &= {}^t\mathrm{Com}(P^{-1}AP) = {}^t(\mathrm{Com}\,P^{-1}\ \mathrm{Com}\,A\ \mathrm{Com}\,P)\\ &= \widetilde{P}\widetilde{A}\widetilde{P^{-1}} = (\det P)P^{-1}\widetilde{A}(\det P^{-1})P\\ &= P^{-1}\widetilde{A}P.\end{aligned}$$

Il en résulte donc que $B + \widetilde{B} = P^{-1}(A + \widetilde{A})P = \lambda I_n$, ce qui prouve le résultat annoncé.

On se donne dans la suite une matrice $A \in \mathcal{M}_n(\mathbb{C})$ et un scalaire $\lambda \in \mathbb{C}$ tels que $A + \widetilde{A} = \lambda I_n$. Observons qu'en multipliant cette égalité par A, il vient $A^2 - \lambda A + \det(A)I = 0$. On va mener l'analyse en discutant selon le rang de A. On fera la synthèse directement dans chaque cas.

• Supposons $\mathrm{rg}(A) < n-1$. Alors tous les mineurs de taille $n-1$ de A sont nuls de sorte que $\widetilde{A} = 0$. La matrice A est alors scalaire et donc nulle puisque son rang n'est pas égal à n. Réciproquement, la matrice nulle est bien dans E_0.

• Supposons que $\mathrm{rg}(A) = n-1$ et que le scalaire λ n'est pas nul. On a alors $A^2 = \lambda A$ de sorte que la matrice A est annulée par le polynôme scindé à racines simples $X(X-\lambda)$. Elle donc diagonalisable. Comme son rang est égal à $n-1$, elle est semblable à la matrice diagonale $D = \mathrm{Diag}(\lambda, \lambda, \ldots, \lambda, 0)$. Un calcul immédiat montre que $\widetilde{D} = \mathrm{Diag}(0, 0, \ldots, 0, \lambda^{n-1})$. La matrice D est donc dans E_λ si et seulement si $\lambda = \lambda^{n-1}$ c'est-à-dire si et seulement si λ est une racine $(n-2)$-ième de l'unité (on a supposé ici $\lambda \neq 0$).

Réciproquement, pour toute racine $(n-2)$-ième de l'unité λ la classe de similitude de la matrice $\mathrm{Diag}(\lambda, \lambda, \ldots, \lambda, 0)$ est incluse dans E_λ. On obtient donc $n-2$ classes de similitude de ce type.

• Supposons toujours que $\mathrm{rg}(A) = n-1$ mais avec $\lambda = 0$. On a alors $A^2 = 0$, c'est-à-dire $\mathrm{Im}\,A \subset \mathrm{Ker}\,A$. Cela impose $2\,\mathrm{rg}(A) \leqslant n$, c'est-à-dire $2(n-1) \leqslant n$, soit encore $n \leqslant 2$. Donc E_0 ne contient pas de matrice de rang $n-1$ lorsque $n > 2$. Pour $n = 2$, on a, en posant $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, $\widetilde{A} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ de sorte que $A + \widetilde{A} = \mathrm{Tr}(A)I_n$. Ainsi, pour $n = 2$ le problème est entièrement résolu : toutes les matrices sont solutions, A étant dans $E_{\mathrm{Tr}\,A}$. Dans la suite de la discussion on supposera donc $n \geqslant 3$.

• Il reste pour finir le cas où A est inversible. Notons Δ le discriminant du polynôme $P(X) = X^2 - \lambda X + \det A$.

∗ Si $\Delta \neq 0$, P est scindé à racines simples et la matrice A est diagonalisable. Notons $\lambda_1 \neq \lambda_2$ les deux racines de P et $p \in [\![1, n-1]\!]$ la

dimension du sous-espace propre associé à λ_1. La matrice A est semblable à $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2)$, où λ_1 apparaît p fois et λ_2 apparaît $n-p$ fois. On a les relations $\lambda_1 + \lambda_2 = \lambda$ et $\lambda_1\lambda_2 = \det A$ (relation coefficients-racines pour l'équation du second degré). Par ailleurs on a aussi $\det A = \det D = \lambda_1^p \lambda_2^{n-p}$. Il en résulte que λ_1 et λ_2 doivent vérifier la relation $\lambda_1^{1-p} = \lambda_2^{n-p-1}$.

Étudions la réciproque. Soit λ_1 un nombre complexe non nul quelconque et p un entier naturel compris entre 1 et $n-1$. Soit λ_2 un nombre complexe distinct de λ_1 vérifiant $\lambda_2^{n-p-1} = \lambda_1^{1-p}$. Si $p \neq n-1$, il y a $n-p-1$ possibilités pour λ_2 sauf lorsque λ_1 est une racine $(n-2)$-ième de l'unité où il y en a une de moins. Lorsque $p = n-1$, λ_1 doit nécessairement être une racine $(n-1)$-ième de 1 et λ_2 est un complexe non nul quelconque, différent de λ_1. Posons $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2)$ où λ_1 apparaît p fois et λ_2 apparaît $n-p$ fois. On a, grâce à la relation vérifiée par λ_1 et λ_2,

$$\widetilde{D} = (\det D)D^{-1} = \lambda_1\lambda_2 D^{-1} = \mathrm{Diag}(\underbrace{\lambda_2, \ldots, \lambda_2}_{p \text{ fois}}, \underbrace{\lambda_1, \ldots, \lambda_1}_{n-p \text{ fois}})$$

Et donc $D + \widetilde{D} = (\lambda_1 + \lambda_2)I_n = \lambda I_n$ avec $\lambda = \lambda_1 + \lambda_2$. Toute la classe de similitude de D est incluse dans E_λ.

$*$ Il reste à étudier le cas où $\Delta = 0$, c'est-à-dire le cas où $\lambda^2 = 4\det A$. Le polynôme du second degré P admet alors $\mu = \dfrac{\lambda}{2}$ comme racine double et $(A - \mu I_n)^2 = 0$. Mais alors μ est l'unique valeur propre de A et on a donc $\det A = \mu^n$. Comme $\mu^2 = \det A$, μ doit être une racine $(n-2)$-ième de l'unité ($\mu \neq 0$ car $\det A \neq 0$). Réciproquement, soit μ une racine $(n-2)$-ième de l'unité et B une matrice quelconque vérifiant $B^2 = 0$. La matrice $A = \mu I_n + B$ est inversible, d'inverse $A^{-1} = \dfrac{1}{\mu} I_n - \dfrac{1}{\mu^2} B$. On a $\det A = \mu^n = \mu^2$. Donc, $\widetilde{A} = (\det A)A^{-1} = \mu I_n - B$ et $\widetilde{A} + A = 2\mu I_n$ est bien scalaire. $\lhd$

1.28. Expression de la transposée de la comatrice de $XI_n - A$

Soit K un corps commutatif et $A \in \mathcal{M}_n(K)$. On note ${}^t\mathrm{Com}(XI_n - A)$ la matrice transposée de la matrice des cofacteurs de $XI_n - A$. Montrer qu'il existe une base $(U_0, U_1, \ldots, U_{n-1})$ de $K_{n-1}[X]$ telle que ${}^t\mathrm{Com}(XI_n - A) = \sum_{i=0}^{n-1} U_i(X)A^i$.

(École normale supérieure)

$\rhd$ **Solution.**

Chaque cofacteur de la matrice $XI_n - A$ est un polynôme en X de

degré au plus $n-1$, ce qui montre l'existence de matrices $B_0, \dots, B_{n-1}$ telles que

$$ {}^t\mathrm{Com}(XI_n - A) = B_0X^{n-1} + B_1X^{n-2} + \cdots + B_{n-1}. $$

Cette décomposition est unique car, si on écrit $B_k = (b_{ij}^k)$, pour tout $(i,j) \in [\![1,n]\!]^2$, la décomposition dans la base canonique de $K_{n-1}[X]$ du coefficient d'indice (i,j) de ${}^t\mathrm{Com}(XI_n - A)$ est $b_{ij}^0X^{n-1} + \cdots + b_{ij}^{n-1}$.

Nous allons montrer que les matrices B_k sont des polynômes en A en utilisant la relation fondamentale vérifiée par la comatrice. Écrivons $\chi_A = a_nX^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$ le polynôme caractéristique de A, avec $a_n = 1$. La relation

$$ {}^t\mathrm{Com}(XI_n - A)(XI_n - A) = \det(XI_n - A)I_n $$

s'écrit

$$ \begin{aligned} &B_0X^n + (B_1 - B_0A)X^{n-1} + \cdots + (B_{n-1} - B_{n-2}A)X - B_{n-1}A \\ &= \chi_A(X)I_n = a_nX^nI_n + a_{n-1}X^{n-1}I_n + \cdots + a_0I_n, \end{aligned} $$

ce qui par unicité de la décomposition donne

$$ B_0 = a_nI_n \quad \text{et} \quad \text{pour } 1 \leqslant k \leqslant n-1,\ B_k - B_{k-1}A = a_{n-k}I_n. $$

On en déduit $B_0 = a_nI_n$, $B_1 = a_nA + a_{n-1}I_n$ et, plus généralement, par une récurrence immédiate, pour $0 \leqslant k \leqslant n-1$,

$$ B_k = a_nA^k + a_{n-1}A^{k-1} + \cdots + a_{n-k+1}A + a_{n-k}I_n = \sum_{i=0}^{k} a_{n-k+i}A^i. $$

On obtient donc

$$ \begin{aligned} {}^t\mathrm{Com}(XI_n - A) &= \sum_{k=0}^{n-1} B_{n-1-k}X^k = \sum_{k=0}^{n-1} \sum_{i=0}^{n-k-1} a_{k+i+1}A^iX^k \\ &= \sum_{i=0}^{n-1} \left(\sum_{k=0}^{n-i-1} a_{k+i+1}X^k \right) A^i. \end{aligned} $$

C'est la décomposition cherchée avec $U_i(X) = \sum_{k=0}^{n-i-1} a_{k+i+1}X^k$. Chaque polynôme U_i est de degré $n-i-1$ car le coefficient de X^{n-i-1} est $a_n = 1$. Ces n polynômes étant échelonnés en degré, ils forment une famille libre donc une base de $K_{n-1}[X]$. ◁

Dans l'exercice suivant, on détermine la différentielle du déterminant. La relation fondamentale vérifiée par la comatrice est l'outil essentiel.

1.29. Différentielle du déterminant

1. Prouver que l'application $\varphi : \mathrm{GL}_n(\mathbb{R}) \to \mathcal{M}_n(\mathbb{R})$ qui à X associe $(\det \mathrm{X})\mathrm{X}^{-1}$ admet un et un seul prolongement continu $\overline{\varphi}$ à $\mathcal{M}_n(\mathbb{R})$.

2. Soient A, B dans $\mathcal{M}_n(\mathbb{R})$. Prouver que

$$\left(\frac{d}{dt}(\det(\mathrm{A}+t\mathrm{B}))\right)_0 = \mathrm{Tr}(\overline{\varphi}(\mathrm{A})\mathrm{B}).$$

(École polytechnique)

▷ **Solution.**

1. Pour tout $\mathrm{X} \in \mathcal{M}_n(\mathbb{R})$, notons $\widetilde{\mathrm{X}}$ la transposée de la comatrice de X. On a la relation $\widetilde{\mathrm{X}}\mathrm{X} = \mathrm{X}\widetilde{\mathrm{X}} = (\det \mathrm{X})\mathrm{I}_n$. En particulier, si X est inversible, alors $\widetilde{\mathrm{X}} = (\det \mathrm{X})\mathrm{X}^{-1} = \varphi(\mathrm{X})$. L'application $\overline{\varphi} : \mathrm{X} \longmapsto \widetilde{\mathrm{X}}$ est donc un prolongement de φ à $\mathcal{M}_n(\mathbb{R})$. Elle est continue, car les coefficients de $\widetilde{\mathrm{X}}$ sont des fonctions polynomiales donc continues des coefficients de X.

L'application $\overline{\varphi}$ est le seul prolongement continu de φ à $\mathcal{M}_n(\mathbb{R})$, car $\mathrm{GL}_n(\mathbb{R})$ est dense dans $\mathcal{M}_n(\mathbb{R})$.

2. Considérons la fonction $f : t \in \mathbb{R} \longmapsto \det(\mathrm{A}+t\mathrm{B}) \in \mathbb{R}$. Il s'agit de déterminer $f'(0)$.

Notons $\mathrm{C}_1, \mathrm{C}_2, \ldots, \mathrm{C}_n$ (respectivement $\mathrm{C}'_1, \mathrm{C}'_2, \ldots, \mathrm{C}'_n$) les vecteurs colonnes de A (respectivement de B). On a, pour tout réel t,

$$f(t) = \det(\mathrm{C}_1 + t\mathrm{C}'_1, \ldots, \mathrm{C}_n + t\mathrm{C}'_n).$$

Par multilinéarité du déterminant, f est une fonction polynôme. Son nombre dérivée en 0 est égal au coefficient du terme de degré 1. Par multilinéarité, on trouve

$$f'(0) = \sum_{j=1}^{n} \det(\mathrm{C}_1, \ldots, \mathrm{C}_{j-1}, \mathrm{C}'_j, \mathrm{C}_{j+1}, \ldots, \mathrm{C}_n).$$

Calculons $\det(\mathrm{C}_1, \ldots, \mathrm{C}_{j-1}, \mathrm{C}'_j, \mathrm{C}_{j+1}, \ldots, \mathrm{C}_n)$ en le développant selon la j-ième colonne. On obtient $\sum_{i=1}^{n} b_{ij}\Delta_{ij}$, où Δ_{ij} est le cofacteur du coefficient a_{ij} de la matrice A. On a donc $f'(0) = \sum_{1 \leqslant i,j \leqslant n} \Delta_{ij} b_{ij}$. Comme Δ_{ij} est le terme d'indice (j,i) de $\overline{\varphi}(\mathrm{A})$, $\sum_{i=1}^{n} \Delta_{ij} b_{ij}$ est le terme d'indice jj de $\overline{\varphi}(\mathrm{A})\mathrm{B}$ et donc $f'(0) = \mathrm{Tr}(\overline{\varphi}(\mathrm{A})\mathrm{B})$. On conclut :

$$\boxed{\left(\frac{d}{dt}(\det(\mathrm{A}+t\mathrm{B}))\right)_0 = \mathrm{Tr}(\overline{\varphi}(\mathrm{A})\mathrm{B})}. \quad \triangleleft$$

On vient donc de prouver que la différentielle de l'application déterminant en un point A *est la forme linéaire* $M \mapsto \mathrm{Tr}(\overline{\varphi}(A)M)$.

Une propriété fondamentale du déterminant est la multiplicativité : $\det(AB) = \det A \det B$ *si* A *et* B *sont deux matrices carrées de même taille. Toutefois, le produit* AB *peut être une matrice carrée sans que* A *et* B *soient carrées : il suffit que* A *soit de taille* (n, n_1) *et* B *de taille* (m, n). *Que dire du déterminant de* AB *dans ce cas ? La formule de Cauchy-Binet répond à cette question.*

1.30. Formule de Cauchy-Binet

Pour une matrice A et deux p-uplets d'indices $i_1 < i_2 < \cdots < i_p$ et $k_1 < k_2 < \cdots < k_p$ on note $A\begin{pmatrix} i_1 i_2 \dots i_p \\ k_1 k_2 \dots k_p \end{pmatrix}$ le mineur de A obtenu en prenant les lignes $i_1, \dots, i_p$ et les colonnes $k_1, \dots, k_p$. Soit $A \in \mathcal{M}_{n,m}(K)$, $B \in \mathcal{M}_{m,n}(K)$ et $C = AB$. Montrer que si $n > m$ alors $\det C = 0$. Si $n \leqslant m$ prouver que

$$\det C = \sum_{1 \leqslant k_1 < k_2 < \cdots < k_n \leqslant m} A\begin{pmatrix} 1\ 2\ \dots n \\ k_1 k_2 \dots k_n \end{pmatrix} B\begin{pmatrix} k_1 k_2 \dots k_n \\ 1\ 2\ \dots n \end{pmatrix}$$

(École normale supérieure)

▷ **Solution.**

La matrice C est carrée de taille (n, n) : parler de son déterminant a bien un sens. Si $n > m$, $\mathrm{rg}(C) \leqslant \mathrm{rg}(A) \leqslant m < n$ donc $\det C = 0$. Dans la suite on suppose $n \leqslant m$ et on note $A_1, \dots, A_m$ les colonnes de la matrice A et b_{ij} les coefficients de B. On a

$$\begin{aligned}
\det C &= \det\left(\sum_{k_1=1}^{m} b_{k_1 1} A_{k_1}, \dots, \sum_{k_n=1}^{m} b_{k_n n} A_{k_n}\right) \\
&= \sum_{(k_1,\dots,k_n) \in \llbracket 1,m \rrbracket^n} A\begin{pmatrix} 1\ 2\ \dots n \\ k_1 k_2 \dots k_n \end{pmatrix} b_{k_1 1} b_{k_2 2} \dots b_{k_n n} \\
&= \sum_{k_1 < \cdots < k_n} \sum_{\sigma \in \mathcal{S}_n} \underbrace{A\begin{pmatrix} 1\ 2\ \dots n \\ k_{\sigma(1)} k_{\sigma(2)} \dots k_{\sigma(n)} \end{pmatrix}}_{= \varepsilon(\sigma) A\left(\begin{smallmatrix} 1\ 2\ \dots n \\ k_1 k_2 \dots k_n \end{smallmatrix}\right)} b_{k_{\sigma(1)} 1} b_{k_{\sigma(2)} 2} \dots b_{k_{\sigma(n)} n} \\
&= \sum_{k_1 < \cdots < k_n} A\begin{pmatrix} 1\ 2\ \dots n \\ k_1 k_2 \dots k_n \end{pmatrix} \sum_{\sigma \in \mathcal{S}_n} \varepsilon(\sigma) b_{k_{\sigma(1)} 1} b_{k_{\sigma(2)} 2} \dots b_{k_{\sigma(n)} n}
\end{aligned}$$

$$= \sum_{1 \leqslant k_1 < k_2 < \cdots < k_n \leqslant m} \mathrm{A}\begin{pmatrix} 1 & 2 & \ldots & n \\ k_1 & k_2 & \ldots & k_n \end{pmatrix} \mathrm{B}\begin{pmatrix} k_1 & k_2 & \ldots & k_n \\ 1 & 2 & \ldots & n \end{pmatrix}.$$

D'où le résultat. ◁

On a regroupé ci-après les calculs de déterminants par blocs. Le seul résultat à connaître concerne les matrices triangulaires par blocs : le déterminant est alors égal au produit des déterminants des blocs diagonaux. Dans la plupart des cas on essaie de se ramener à cette situation.

1.31. Déterminant d'une matrice par blocs (1)

Soient A, B, C et D quatre matrices de $\mathcal{M}_n(\mathbb{C})$ telles que D et C commutent. Calculer $\det\begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix}$.

(École polytechnique)

▷ **Solution.**

L'idée est d'écrire M comme produit de deux matrices triangulaires par blocs (l'une supérieure et l'autre inférieure) dont on pourra calculer facilement le déterminant. On peut essayer par exemple de trouver des blocs U, V, W de taille n tels que

$$\mathrm{M} = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D} \end{pmatrix} = \begin{pmatrix} \mathrm{U} & \mathrm{V} \\ 0 & \mathrm{D} \end{pmatrix}\begin{pmatrix} \mathrm{I} & 0 \\ \mathrm{W} & \mathrm{I} \end{pmatrix}.$$

Pour cela, il suffit que $\mathrm{U} + \mathrm{VW} = \mathrm{A}$, $\mathrm{V} = \mathrm{B}$ et $\mathrm{DW} = \mathrm{C}$. Supposons d'abord que D est inversible. Il suffit de prendre $\mathrm{W} = \mathrm{D}^{-1}\mathrm{C}$, $\mathrm{V} = \mathrm{B}$ et $\mathrm{U} = \mathrm{A} - \mathrm{BD}^{-1}\mathrm{C}$ pour que la décomposition soit valide et dans ces conditions

$$\det \mathrm{M} = \det(\mathrm{A} - \mathrm{BD}^{-1}\mathrm{C}) \det \mathrm{D} = \det(\mathrm{AD} - \mathrm{BD}^{-1}\mathrm{CD}) = \det(\mathrm{AD} - \mathrm{BC})$$

car D et C commutent.

Pour étendre le résultat au cas où D n'est pas inversible nous proposons trois méthodes :

• Première méthode : par passage à la limite. On se donne une suite (D_k) formée de matrices inversibles commutant avec C et qui converge vers D. On peut prendre par exemple $\mathrm{D}_k = \mathrm{D} - \dfrac{1}{k}\mathrm{I}_n$: la matrice D_k est inversible si $\dfrac{1}{k}$ est inférieur à la plus petite valeur propre réelle > 0 de D (s'il en existe). Posons $\mathrm{M}_k = \begin{pmatrix} \mathrm{A} & \mathrm{B} \\ \mathrm{C} & \mathrm{D}_k \end{pmatrix}$ pour tout k. La suite $\det \mathrm{M}_k$ converge vers $\det \mathrm{M}$ par continuité du déterminant. Or d'après ce qui précède,

$$\det \mathrm{M}_k = \det(\mathrm{AD}_k - \mathrm{BC}) \xrightarrow[k \to \infty]{} \det(\mathrm{AD} - \mathrm{BC}).$$

Par unicité de la limite, on a bien $\det M = \det(AD - BC)$.

• Deuxième méthode : cette méthode n'est pas très éloignée de la précédente, mais elle n'utilise pas de passage à la limite. Pour tout $\lambda \in \mathbb{C}$, $D - \lambda I$ commute avec C et si λ est en dehors du spectre de D, cette matrice est inversible. Dans ce cas, on a

$$\det\begin{pmatrix} A & B \\ C & D - \lambda I \end{pmatrix} = \det(A(D - \lambda I) - BC)$$

Puisque le spectre de D est fini, le polynôme

$$P = \det\begin{pmatrix} A & B \\ C & D - XI \end{pmatrix} - \det(A(D - XI) - BC)$$

s'annule en une infinité de valeurs, il est donc nul. L'identité $P(0) = 0$ donne l'identité recherchée. Cette méthode est encore valable lorsqu'on remplace $\mathbb{C}$ par n'importe quel corps commutatif infini.

• Troisième méthode : par plongement de $\mathcal{M}_{2n}(\mathbb{C})$ dans $\mathcal{M}_{2n}(\mathbb{C}(X))$. On pose $M(X) = \begin{pmatrix} A - XI_n & B \\ C & D - XI_n \end{pmatrix}$: c'est une matrice de taille $2n$ à coefficients dans le corps $\mathbb{C}(X)$. La matrice $D - XI_n$ commute encore avec C. D'autre part, $\det(D - XI_n) = (-1)^n \chi_{D(X)} \neq 0$ puisque $\chi_{D(X)}$ est de degré n unitaire. En particulier, $D - XI_n$ est inversible, c'est-à-dire $D - XI_n \in GL_n(\mathbb{C}(X))$. Par conséquent

$$\det M(X) = \det((A - XI_n)(D - XI_n) - BC).$$

En faisant $X = 0$, on trouve $\det M = \det M(0) = \det(AD - BC)$. L'avantage de cette solution est qu'elle reste valable lorsqu'on remplace $\mathbb{C}$ par n'importe quel corps commutatif. ◁

Voici une formule qui généralise le résultat de l'exercice précédent.

1.32. Déterminant d'une matrice par blocs (2) : formule de Williamson

Soit $(A_{ij})_{1 \leqslant i,j \leqslant m}$ une famille de matrices de $\mathcal{M}_n(\mathbb{C})$, qui commutent deux à deux. On considère les matrices

$$A = \begin{pmatrix} A_{11} & \dots & A_{1m} \\ \vdots & & \vdots \\ A_{m1} & \dots & A_{mm} \end{pmatrix} \in \mathcal{M}_{nm}(\mathbb{C}) \quad \text{et} \quad B = \sum_{\sigma \in \mathcal{S}_m} \varepsilon(\sigma) \prod_{i=1}^{m} A_{\sigma(i)i}.$$

Montrer que $\det A = \det B$.

(École polytechnique)

▷ **Solution.**

On va procéder par récurrence sur m le résultat étant évident pour $m = 1$ puisqu'on a alors $A = B = A_{11}$. Supposons donc le résultat vrai au rang $m - 1$ avec $m \geqslant 2$. Comme d'habitude avec des déterminants par blocs on va essayer de se ramener à une matrice triangulaire par blocs pour utiliser ensuite l'hypothèse de récurrence.

Supposons tout d'abord que A_{mm} est inversible. En s'inspirant des étapes élémentaires dans l'algorithme du pivot de Gauss, il est facile d'annuler tous les blocs de la dernière colonne de A excepté le bloc diagonal. Il suffit pour cela de multiplier A à gauche par la matrice

$$M = \begin{pmatrix} I_n & 0 & \dots & 0 & -A_{1m}A_{mm}^{-1} \\ 0 & I_n & 0 & \dots & -A_{2m}A_{mm}^{-1} \\ \vdots & & \ddots & & \vdots \\ 0 & & & \ddots & -A_{m-1,m}A_{mm}^{-1} \\ 0 & \dots & & 0 & I_n \end{pmatrix}.$$

Notons C_{ij} les blocs qui interviennent dans la matrice $C = MA$: on a $C_{im} = 0$ pour tout $i < m$, $C_{mj} = A_{mj}$ pour tout $j \in [\![1, m]\!]$ et enfin $C_{ij} = A_{ij} - A_{im}A_{mm}^{-1}A_{mj}$ pour $(i, j) \in [\![1, m-1]\!]^2$. Comme M est de déterminant 1, on a

$$\det A = \det C = \det A_{mm} \det(C_{ij})_{1 \leqslant i,j \leqslant m-1}.$$

Or les matrices C_{ij} commutent deux à deux et il est donc possible d'appliquer l'hypothèse de récurrence à la matrice C. On a donc

$$\det A = \det A_{mm} \det \left(\sum_{\sigma \in \mathcal{S}_{m-1}} \varepsilon(\sigma) \prod_{i=1}^{m-1} C_{\sigma(i)i} \right).$$

Notons R le sous-anneau de $\mathcal{M}_n(\mathbb{C})$ engendré par les matrices A_{ij}. C'est un anneau commutatif (pas forcément intègre). La matrice B représente en fait le déterminant de la matrice A regardée comme une matrice de taille m à coefficients dans l'anneau R. Si on regarde le calcul précédent comme étant effectué dans $\mathcal{M}_m(R)$, on constate que

$$A_{mm} \sum_{\sigma \in \mathcal{S}_{m-1}} \varepsilon(\sigma) \prod_{i=1}^{m-1} C_{\sigma(i)i} = \sum_{\sigma \in \mathcal{S}_m} \varepsilon(\sigma) \prod_{i=1}^{m} A_{\sigma(i)i} = B.$$

On a donc bien $\det A = \det B$.

Pour le cas général on peut conclure en utilisant la densité du groupe linéaire puisque les deux termes de l'égalité dépendent continûment de la matrice A_{mm}. ◁

1.33. Déterminant d'une matrice par blocs (3)

Soit $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ une matrice carrée complexe inversible avec A et D elles aussi carrées. On écrit $M^{-1} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}$ avec le même découpage. *Trouver une relation entre* $\det M$, $\det D$ *et* $\det A'$.

(École normale supérieure)

▷ **Solution.**

On notera m la taille de A et n celle de D. Pour avoir une idée du résultat regardons le cas particulier où $C = 0$. La matrice M est alors triangulaire par blocs et il vient $\det M = \det A \det D$. Nécessairement A est inversible et d'inverse A'. On a donc $\det D = \det M \det A'$. Nous allons tâcher de montrer ce résultat pour C quelconque.

Comme $M^{-1}M = I_{m+n}$, on peut écrire

$$\begin{pmatrix} I_m & 0 \\ 0 & I_n \end{pmatrix} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \left(\begin{array}{c|c} A'A + B'C & A'B + B'D \\ \hline C'A + D'C & C'B + D'D \end{array}\right).$$

On a donc, en particulier, $A'B + B'D = 0$ et $C'B + D'D = I_n$.

Compte-tenu de ces relations, on a

$$M^{-1}\begin{pmatrix} I_m & B \\ 0 & D \end{pmatrix} = \begin{pmatrix} A' & B' \\ C' & D' \end{pmatrix}\begin{pmatrix} I_m & B \\ 0 & D \end{pmatrix} = \left(\begin{array}{c|c} A' & A'B + B'D \\ \hline C' & C'B + D'D \end{array}\right)$$
$$= \begin{pmatrix} A' & 0 \\ C' & I_n \end{pmatrix}.$$

En passant aux déterminants on a

$$\det\left(M^{-1} \times \begin{pmatrix} I_m & B \\ 0 & D \end{pmatrix}\right) = \det M^{-1} \det D = \det A' \det I_n,$$

d'où finalement $\det M \det A' = \det D$. ◁

1.34. Déterminant d'une matrice par blocs (4)

Soit $n \geqslant 3$, $(a, b, c, d) \in \mathbb{C}^4$ et $N \in \mathcal{M}_{n-2}(\mathbb{C})$. On considère la matrice $M = \begin{pmatrix} a & * & b \\ * & N & * \\ c & * & d \end{pmatrix} \in \mathcal{M}_n(\mathbb{C})$, ainsi que $A = \begin{pmatrix} a & * \\ * & N \end{pmatrix}$, $B = \begin{pmatrix} * & b \\ N & * \end{pmatrix}$, $C = \begin{pmatrix} * & N \\ c & * \end{pmatrix}$ et $D = \begin{pmatrix} N & * \\ * & d \end{pmatrix}$.

1. Montrer que $\det A = \det B = 0$ implique $\det M \det N = 0$.

2. Montrer que $\det \mathrm{M} \det \mathrm{N} = \det \mathrm{A} \det \mathrm{D} - \det \mathrm{B} \det \mathrm{C}$.
(École normale supérieure)

▷ **Solution.**

1. Supposons que $\det \mathrm{A} = \det \mathrm{B} = 0$ et que N est inversible. Dans ce cas la matrice A n'est pas inversible, donc de rang $\leqslant n-2$, mais contient une sous-matrice de taille $n-2$ inversible à savoir N. On a donc $\mathrm{rg}(\mathrm{A}) = n-2$ et si on note $\mathrm{A}_1, \ldots, \mathrm{A}_{n-1}$ les colonnes de A, alors les colonnes $\mathrm{A}_2, \ldots, \mathrm{A}_{n-1}$ engendrent l'image de A (on confond la matrice A et l'endomorphisme de $\mathbb{C}^{n-1}$ qui lui est canoniquement associé). En particulier A_1 peut s'écrire sous la forme $\mathrm{A}_1 = \lambda_2 \mathrm{A}_2 + \cdots + \lambda_{n-1}\mathrm{A}_{n-1}$. On calcule alors le déterminant de M en effectuant sur les colonnes la manipulation $\mathrm{C}_1 \leftarrow \mathrm{C}_1 - \lambda_2 \mathrm{C}_2 - \cdots - \lambda_{n-1}\mathrm{C}_{n-1}$ puis en développant selon la première colonne. Il vient :

$$\det \mathrm{M} = \begin{vmatrix} 0 & * & b \\ 0 & \mathrm{N} & * \\ c' & * & d \end{vmatrix} = (-1)^{n+1} c' \det \mathrm{B} = 0.$$

2. Supposons dans un premier temps que N est inversible. Le vecteur colonne de $\mathbb{C}^{n-2}$ qui apparaît dans la première colonne de M entre les coefficients a et c peut s'écrire comme combinaison linéaire des colonnes de N. En effectuant la manipulation correspondante sur les colonnes de M, on a

$$\det \mathrm{M} = \begin{vmatrix} a' & * & b \\ 0 & \mathrm{N} & * \\ c' & * & d \end{vmatrix}.$$

De la même manière, on peut faire apparaître des zéros entre b et d :

$$\det \mathrm{M} = \begin{vmatrix} a' & * & b' \\ 0 & \mathrm{N} & 0 \\ c' & * & d' \end{vmatrix}.$$

En manipulant sur les lignes on peut également remplacer les $*$ par des zéros :

$$\det \mathrm{M} = \begin{vmatrix} a' & 0 & b' \\ 0 & \mathrm{N} & 0 \\ c' & 0 & d' \end{vmatrix}$$

(les coefficients a', b', c', d' ne sont pas affectés par ces dernières opérations). On développe maintenant par rapport à la première colonne :

$$\begin{aligned} \det \mathrm{M} &= a' \begin{vmatrix} \mathrm{N} & 0 \\ 0 & d' \end{vmatrix} + (-1)^{n+1} c' \begin{vmatrix} 0 & b' \\ \mathrm{N} & 0 \end{vmatrix} \\ &= a'd' \det \mathrm{N} + (-1)^{n+1}(-1)^n b'c' \det \mathrm{N}. \end{aligned}$$

Mais on a $a' \det \mathrm{N} = \det \mathrm{A}$, $b' \det \mathrm{N} = \det \mathrm{B}$, $c' \det \mathrm{N} = \det \mathrm{C}$ et $d' \det \mathrm{N} = \det \mathrm{D}$, car les manipulations précédentes n'ont pas modifié les déterminants des matrices extraites A, B, C, D. En multipliant l'égalité ci-dessus par $\det \mathrm{N}$ on obtient donc

$$\boxed{\det \mathrm{M} \det \mathrm{N} = \det \mathrm{A} \det \mathrm{D} - \det \mathrm{B} \det \mathrm{C}}.$$

Par densité de $\mathrm{GL}_{n-2}(\mathbb{C})$ dans $\mathcal{M}_{n-2}(\mathbb{C})$, le résultat demeure vrai pour N quelconque. $\lhd$

Dans l'exercice suivant on calcule le déterminant d'une application linéaire sur $\mathcal{L}(\mathrm{E})$ qui fait intervenir des matrices par blocs.

1.35. Déterminant d'un automorphisme intérieur

Soit E un K-espace vectoriel de dimension finie et $g \in \mathrm{GL}(\mathrm{E})$. On considère $\varphi : u \longmapsto gug^{-1}$. Déterminer $\det \varphi$ et $\mathrm{Tr}\, \varphi$.

(**École polytechnique**)

$\rhd$ **Solution.**

On va commencer par le déterminant. Le problème devient plus simple si on le généralise. En effet, multiplier à gauche et à droite sont deux opérations indépendantes : si on note φ_1 (resp. φ_2) l'application $u \longmapsto gu$ (resp. $u \longmapsto ug^{-1}$), on a $\varphi = \varphi_1 \circ \varphi_2 = \varphi_2 \circ \varphi_1$ et par conséquent $\det \varphi = \det \varphi_1 \det \varphi_2$.

On regarde alors le problème matriciellement. Si A est une matrice de $\mathcal{M}_n(\mathrm{K})$, on notera $\mathrm{G_A} : \mathrm{M} \longmapsto \mathrm{AM}$ (resp. $\mathrm{D_A} : \mathrm{M} \longmapsto \mathrm{MA}$) l'application consistant à multiplier à gauche (resp. à droite) par A. Considérons la base canonique $(\mathrm{E}_{ij})_{1 \leqslant i,j \leqslant n}$ de $\mathcal{M}_n(\mathrm{K})$ ordonnée de la manière suivante :

$$(\mathrm{E}_{11}, \mathrm{E}_{21}, \ldots, \mathrm{E}_{n1}, \mathrm{E}_{12}, \mathrm{E}_{22}, \ldots, \mathrm{E}_{n2}, \ldots, \mathrm{E}_{1n}, \mathrm{E}_{2n}, \ldots, \mathrm{E}_{nn}).$$

Si $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n}$, on a :

$$\mathrm{G_A}(\mathrm{E}_{ij}) = \mathrm{AE}_{ij} = a_{1i}\mathrm{E}_{1j} + \cdots + a_{ni}\mathrm{E}_{nj}.$$

Il en résulte que la matrice de $\mathrm{G_A}$ est une matrice diagonale par blocs formée de n matrices A. On a donc $\det \mathrm{G_A} = (\det \mathrm{A})^n$.

Pour la multiplication à droite, il est inutile de faire une étude supplémentaire. En effet, notons $\mathrm{T} : \mathrm{M} \longmapsto {}^t\mathrm{M}$ la transposition. C'est un isomorphisme involutif de $\mathcal{M}_n(\mathrm{K})$. Pour toutes matrices A et M on a $\mathrm{T}(\mathrm{AM}) = \mathrm{T}(\mathrm{M})\mathrm{T}(\mathrm{A})$, c'est-à-dire que $\mathrm{T} \circ \mathrm{G_A} = \mathrm{D}_{{}^t\mathrm{A}} \circ \mathrm{T}$. En passant au déterminant on a donc $\det \mathrm{G_A} = \det \mathrm{D}_{{}^t\mathrm{A}}$ (car $\det \mathrm{T} \neq 0$). Une matrice et sa transposée ayant même déterminant, on en déduit que $\det \mathrm{D_A} = (\det \mathrm{A})^n$.

Dans le cas qui nous intéresse, on choisit une base de E et on on note A la matrice de g dans une base. Le déterminant de φ est celui de $G_A \circ D_{A^{-1}}$. On a donc

$$\det \varphi = \det G_A \det D_{A^{-1}} = (\det A)^n (\det A^{-1})^n = 1.$$

Passons à la trace. Ici la décomposition de φ en $G_A \circ D_{A^{-1}}$ ne sert à rien (la trace n'est pas multiplicative !). Il n'y a guère d'autre choix que de regarder directement la matrice de φ dans la base $(E_{ij})_{1 \leqslant i,j \leqslant n}$. On pose $A^{-1} = (b_{ij})_{1 \leqslant i,j \leqslant n}$. On a

$$\varphi(E_{ij}) = A E_{ij} A^{-1} = \sum_{1 \leqslant r,s \leqslant n} a_{ri} b_{js} E_{rs}.$$

La coordonnée de $\varphi(E_{ij})$ selon E_{ij} vaut donc $a_{ri} b_{jj}$. Il en résulte que

$$\operatorname{Tr} \varphi = \sum_{1 \leqslant i,j \leqslant n} a_{ri} b_{jj} = \operatorname{Tr} A \operatorname{Tr} A^{-1} = \boxed{\operatorname{Tr} g \operatorname{Tr} g^{-1}}. \quad \triangleleft$$

Les exercices suivants concernent des matrices à coefficients entiers. Le premier se rattache à la théorie des ensembles.

1.36. Déterminant d'une matrice d'incidence

Soit $I = \{1, 2, \ldots, n\}$, $(A_i)_{1 \leqslant i \leqslant 2^n - 1}$ les parties non vides de I numérotées dans un ordre quelconque et $B \in \mathcal{M}_{2^n-1}(\mathbb{R})$ la matrice définie par $b_{ij} = 1$ si $A_i \cap A_j \neq \emptyset$ et $b_{ij} = 0$ sinon.

Montrer que $\det B$ ne dépend pas de la numérotation choisie et le calculer.

▷ **Solution.**

Changer la numérotation des parties non vides de I revient à considérer une permutation σ de $\{1, \ldots, 2^n - 1\}$ et à noter $(A_{\sigma(i)})_{1 \leqslant i \leqslant 2^n - 1}$ les parties non vides de I. On considère alors la matrice $B' \in \mathcal{M}_{2^n-1}(\mathbb{R})$ définie par $b'_{ij} = 1$ si $A_{\sigma(i)} \cap A_{\sigma(j)} \neq \emptyset$ et $b'_{ij} = 0$ sinon. On a donc $b'_{ij} = b_{\sigma(i),\sigma(j)}$. La matrice de permutation $P_\sigma = (p_{ij}) \in \mathcal{M}_{2^n-1}(\mathbb{R})$ définie par $p_{ij} = 1$ si $\sigma(i) = j$ et 0 sinon, vérifie $(P_\sigma)^{-1} = P_{\sigma^{-1}} = {}^t P_\sigma$. On en déduit que

$$b'_{ij} = b_{\sigma(i),\sigma(j)} = \sum_{1 \leqslant k,l \leqslant 2^n - 1} p_{ik} b_{kl} p_{jl},$$

ce qui montre que $B' = P_\sigma B (P_\sigma)^{-1}$. Les matrices B et B' sont semblables, donc ont même déterminant.

On note désormais B_n une des matrices précédentes et D_n son déterminant. On le calcule en raisonnant par récurrence. On a clairement $D_1 = 1$. Pour un entier $n \geqslant 2$, on numérote les parties non vides de I de la façon suivante : d'abord les parties ne contenant pas n, puis les parties contenant n, la dernière étant $\{n\}$. On suppose de plus que, pour $1 \leqslant i \leqslant 2^{n-1} - 1$, on a $A_{i+2^{n-1}-1} = A_i \cup \{n\}$. On obtient

$$B_n = \left(\begin{array}{ccc|c} & B_{n-1} & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline & B_{n-1} & & 1 \\ & & & \vdots \\ & & & 1 \\ \hline 0 & \cdots & 0 \;|\; 1 & \cdots \quad 1 & 1 \end{array}\right),$$

où $J \in M_{2^{n-1}-1}(\mathbb{R})$ est la matrice dont tous les coefficients sont égaux à 1.

En soustrayant la dernière colonne aux $2^{n-1} - 1$ précédentes, il vient

$$\begin{aligned} D_n &= \det \left(\begin{array}{ccc|c} & B_{n-1} & & 0 \\ & & & \vdots \\ & & & 0 \\ \hline & B_{n-1} & & 1 \\ & & & \vdots \\ & & & 1 \\ \hline 0 & \cdots & 0 \;|\; 0 & \cdots \quad 0 & 1 \end{array}\right) \\ &= \det \left(\begin{array}{c|c} B_{n-1} & B_{n-1} \\ \hline B_{n-1} & 0 \end{array}\right) = \det \left(\begin{array}{c|c} B_{n-1} & B_{n-1} \\ \hline 0 & -B_{n-1} \end{array}\right) \\ &= \det B_{n-1} \det(-B_{n-1}) = -(\det B_{n-1})^2 = -D_{n-1}^2, \end{aligned}$$

puisque la taille de B_{n-1} est impaire. Compte tenu de la valeur de D_1, on obtient, pour tout $n \geqslant 2$, $D_n = -1$. ◁

L'exercice qui suit utilise un déterminant pour résoudre un problème de dénombrement.

1.37. Dérangements pairs et impairs

Un dérangement est une permutation sans point fixe. Y a-t-il plus de dérangements impairs ou de dérangements pairs dans $\mathcal{S}_n$?

(École polytechnique)

▷ **Solution.**

En notant D_n l'ensemble des dérangements d'ordre n et Δ_n la différence entre le nombre de dérangements pairs et le nombre de dérangements impairs, on obtient $\Delta_n = \sum_{\sigma \in D_n} \varepsilon(\sigma)$. Considérons la matrice $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ dont tous les coefficients sont égaux à 1 sauf ceux de la diagonale qui sont nuls. On écrit $\det A = \sum_{\sigma \in S_n} \varepsilon(\sigma) a_{1\sigma(1)} \dots a_{n\sigma(n)}$. Le produit $a_{1\sigma(1)} \dots a_{n\sigma(n)}$ est nul s'il existe $i \in [\![1, n]\!]$ tel que $\sigma(i) = i$, c'est-à-dire si σ n'est pas un dérangement. Sinon, $a_{1\sigma(1)} \dots a_{n\sigma(n)} = 1$. On obtient donc

$$\det A = \sum_{\sigma \in D_n} \varepsilon(\sigma) = \Delta_n.$$

Reste à calculer le déterminant. C'est un cas particulier du calcul effectué dans l'exercice 1.6. Si on note $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$ et u le vecteur $(1, \dots, 1)$, on obtient

$$\begin{aligned} \Delta_n = \det(u - e_1, \dots, u - e_n) &= \det(-e_1, \dots, -e_n) \\ &\quad + \sum_{i=1}^{n} \det(-e_1, \dots, -e_{i-1}, u, -e_{i+1}, \dots, -e_n) \\ &= (-1)^n + n(-1)^{n-1} = (-1)^n (n-1). \end{aligned}$$

Il y a donc plus de dérangements pairs si $n - 1$ est pair, *i.e.* si n est impair. ◁

L'énoncé qui suit présente un calcul particulièrement astucieux.

1.38. Déterminant de Smith (1875)

Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ où $a_{ij} = \gcd(i, j)$. Calculer $\det A$. On utilisera la relation $n = \sum_{d|n} \varphi(d)$ où φ est l'indicatrice d'Euler.

(École normale supérieure)

▷ **Solution.**

Notons que le lecteur trouvera une preuve de l'identité vérifiée par l'indicatrice d'Euler dans l'exercice 1.18 du tome algèbre 1. L'idée du calcul de $\det A$ est vraiment très astucieuse : il s'agit de décomposer A en un produit de matrices plus simples, en utilisant justement la relation $n = \sum_{d|n} \varphi(d)$. On a effectivement

$$a_{ij} = \operatorname{pgcd}(i,j) = \sum_{d \mid \operatorname{pgcd}(i,j)} \varphi(d) = \sum_{d \mid i \ d \mid j} \varphi(d) = \sum_{d=1}^{n} \varphi(d) b_{di} b_{dj},$$

avec $b_{kl} = 1$ si k divise l et $b_{kl} = 0$ sinon. La matrice $\mathrm{B} = (b_{ij})$ de $\mathcal{M}_n(\mathbb{R})$ est la matrice d'incidence de la relation de divisibilité. C'est une matrice triangulaire supérieure unipotente (*i.e.* avec des 1 sur la diagonale) et la relation ci-dessus équivaut à l'identité matricielle

$$\mathrm{A} = {}^t\mathrm{B}\,\mathrm{Diag}(\varphi(1), \ldots, \varphi(n))\mathrm{B}.$$

En passant au déterminant, on obtient donc le joli résultat suivant :

$$\boxed{\det \mathrm{A} = \prod_{k=1}^{n} \varphi(k)}. \quad \triangleleft$$

L'introduction de la matrice B *permet aussi de calculer le déterminant de la matrice dont le terme* (i,j) *est le nombre de diviseurs communs à* i *et* j.

Dans l'énoncé suivant on étudie à quelle condition un vecteur de $\mathbb{Z}^n$ *peut être complété pour former une base du réseau* $\mathbb{Z}^n$. *Le lecteur trouvera une autre approche de ce problème (qui n'utilise pas les déterminants) dans l'exercice 7.20 du tome 1.*

1.39. Première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$

1. À quelle condition une matrice de $\mathcal{M}_n(\mathbb{Z})$ est-elle inversible dans $\mathcal{M}_n(\mathbb{Z})$?

2. À quelle condition un vecteur de $\mathbb{Z}^n$ est-il la première colonne d'une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$?

(École normale supérieure)

▷ **Solution.**

1. Soit $\mathrm{M} \in \mathcal{M}_n(\mathbb{Z})$.

• Supposons M inversible dans $\mathcal{M}_n(\mathbb{Z})$ *i.e.* $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$. Alors M^{-1} est à coefficients entiers. On a $\det \mathrm{M} \det \mathrm{M}^{-1} = \det \mathrm{I}_n = 1$. Or $\det \mathrm{M}$ et $\det \mathrm{M}^{-1}$ sont des entiers. Donc $\det \mathrm{M}$ est inversible dans $\mathbb{Z}$. Par conséquent, $\det \mathrm{M} = \pm 1$.

• Réciproquement, supposons que $\det \mathrm{M} = \varepsilon = \pm 1$. Notons N la transposée de la comatrice de M. C'est une matrice à coefficients entiers. On a la célèbre relation : $\mathrm{MN} = \mathrm{NM} = \varepsilon \mathrm{I}_n$ et $\mathrm{M}(\varepsilon \mathrm{N}) = (\varepsilon \mathrm{N})\mathrm{M} = \mathrm{I}_n$. Comme $\varepsilon \mathrm{N} \in \mathcal{M}_n(\mathbb{Z})$, M est inversible dans $\mathcal{M}_n(\mathbb{Z})$. On conclut :

$$\boxed{\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z}) \iff \det \mathrm{M} = \pm 1}$$

Plus généralement, si A *est un anneau commutatif et* $\mathrm{M} \in \mathcal{M}_n(\mathrm{A})$, M *est une matrice inversible de* $\mathcal{M}_n(\mathrm{A})$ *si et seulement si* $\det \mathrm{M}$ *est un inversible de* A.

2. Soit $\mathrm{X} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{Z}^n$. Montrons qu'il existe une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$ dont la première colonne est X si et seulement si le pgcd des a_i est égal à 1.

• La condition est nécessaire : soit A une matrice inversible de $\mathcal{M}_n(\mathbb{Z})$ dont la première colonne est X ; on calcule le déterminant de A en développant selon la première colonne : $\det \mathrm{A} = \sum_{i=1}^{n} a_i \mathrm{A}_i$, où A_i est le cofacteur de a_i. Par hypothèse, les A_i sont entiers et $\det \mathrm{A} = \pm 1$ d'après la première question, donc les a_i sont premiers entre eux d'après le théorème de Bezout.

• Montrons que la condition est suffisante en raisonnant par récurrence sur n. C'est évident pour $n = 1$: $a_1 = \pm 1$. Supposons la propriété vérifiée au rang $n-1$ et considérons $\mathrm{X} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in \mathbb{Z}^n$ tel que $\mathrm{pgcd}(a_1, a_2, \ldots, a_n) = 1$. Posons $d = \mathrm{pgcd}(a_1, a_2, \ldots, a_{n-1})$ et, pour $1 \leqslant k \leqslant n-1$, $a'_k = \dfrac{a_k}{d}$. D'après l'hypothèse de récurrence, il existe une matrice inversible B′ de $\mathcal{M}_{n-1}(\mathbb{Z})$ dont la première colonne est $\begin{pmatrix} a'_1 \\ a'_2 \\ \vdots \\ a'_{n-1} \end{pmatrix}$. En multipliant sa première colonne par d, on obtient une matrice B de première colonne $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-1} \end{pmatrix}$, de taille $n-1$ et de déterminant $\pm d$.

On a par hypothèse $\mathrm{pgcd}(d, a_n) = 1$ et donc $\mathrm{pgcd}(\det \mathrm{B}, a_n) = 1$. Il existe $(u, v) \in \mathbb{Z}^2$ tel que $u \det \mathrm{B} + v a_n = 1$. Cherchons une matrice A de taille n répondant au problème, de la forme

$$\mathrm{A} = \left(\begin{array}{ccc|c} & \mathrm{B} & & x_1 \\ & & & \vdots \\ & & & x_{n-1} \\ \hline a_n & 0 & \dots\ 0 & u \end{array}\right) \quad \text{où } (x_1, x_2, \dots, x_{n-1}) \in \mathbb{Z}^{n-1} \text{ est à choisir.}$$

Développons le déterminant de A selon la dernière ligne. Pour cela, posons :

$$\mathrm{B} = \left(\begin{array}{c|c} a_1 & \\ \vdots & \mathrm{C} \\ a_{n-1} & \end{array}\right) \quad \text{et} \quad \mathrm{D} = \left(\begin{array}{c|c} \mathrm{C} & \begin{array}{c} x_1 \\ \vdots \\ x_{n-1} \end{array} \end{array}\right).$$

Dans ces conditions $\det \mathrm{A} = u \det \mathrm{B} + (-1)^{n+1} a_n \det \mathrm{D}$. Pour avoir $\det \mathrm{A} = 1$, il suffit d'avoir $\det \mathrm{D} = (-1)^{n+1} v$. Si $(x_1, \dots, x_{n-1}) = (a'_1, \dots, a'_{n-1})$, alors la matrice D est obtenue à partir de B′ en faisant une permutation circulaire des colonnes. Alors $\det \mathrm{D} = \pm \det \mathrm{B}' = \pm 1$.

En prenant $(x_1, \dots, x_{n-1}) = \varepsilon v(a'_1, \dots, a'_{n-1})$, avec $\varepsilon = \pm 1$, on peut choisir ε pour que $\det \mathrm{D} = (-1)^{n+1} v$ et donc $\det \mathrm{A} = 1$. ◁

1.40. Opération de $\mathrm{GL}_n(\mathbb{Z})$ sur le réseau $\mathbb{Z}^n$

Pour toute matrice $\mathrm{X} \in \mathcal{M}_{n,1}(\mathbb{Z}) = \mathbb{Z}^n$ on note $\Lambda(\mathrm{X})$ le pgcd des coefficients de X. Pour $\mathrm{A} \in \mathcal{M}_n(\mathbb{Z})$ montrer qu'il y a équivalence entre :

(*i*) $\det \mathrm{A} = \pm 1$;
(*ii*) $\forall \mathrm{X} \in \mathbb{Z}^n, \quad \Lambda(\mathrm{AX}) = \Lambda(\mathrm{X})$.

(École normale supérieure)

▷ **Solution.**

• (*i*) $\Longrightarrow$ (*ii*). Par homogénéité du pgcd, il suffit de montrer la relation (*ii*) lorsque $\Lambda(\mathrm{X}) = 1$. Dans ce cas, le théorème de Bezout donne l'existence de $\mathrm{U} \in \mathbb{Z}^n$ tel que ${}^t\mathrm{UX} = 1$. On sait que A est inversible et que A^{-1} est dans $\mathcal{M}_n(\mathbb{Z})$ car $\det \mathrm{A} = \pm 1$ (voir l'exercice précédent). Posons $\mathrm{V} = {}^t\mathrm{A}^{-1}\mathrm{U} \in \mathbb{Z}^n$. On a alors ${}^t\mathrm{VAX} = 1$, ce qui prouve, par la réciproque du théorème de Bezout, que $\Lambda(\mathrm{AX}) = 1$.

• (*ii*) $\Longrightarrow$ (*i*). Montrons que la matrice A est inversible. En effet, si ce n'est pas le cas on peut trouver $\mathrm{X} \in \mathbb{Z}^n$ non nul tel que $\mathrm{AX} = 0$ ce qui contredit (*ii*). Donc $\det \mathrm{A} \in \mathbb{Z}^*$.

Supposons par l'absurde $|\det \mathrm{A}| \neq 1$ et soit p premier divisant $\det \mathrm{A}$. On note $\widetilde{\mathrm{A}} \in \mathcal{M}_n(\mathbb{Z}/p\mathbb{Z})$ la matrice obtenue à partir de A par réduction modulo p. Comme $\det \widetilde{\mathrm{A}} = 0$, cette matrice n'est pas inversible. Il existe donc $\mathrm{Y} \in \mathcal{M}_{n,1}(\mathbb{Z}/p\mathbb{Z}) = (\mathbb{Z}/p\mathbb{Z})^n$ non nul tel que $\widetilde{\mathrm{A}}\mathrm{Y} = 0$. On peut alors

écrire $Y = \begin{pmatrix} \overline{x_1} \\ \overline{x_2} \\ \vdots \\ \overline{x_n} \end{pmatrix}$ avec les x_i entiers. Si $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{Z}^n$, le nombre premier p divise $\Lambda(AX)$ mais sans diviser $\Lambda(X)$. Contradiction. ◁

L'implication (i) $\Longrightarrow$ (ii) montre que si on fait opérer le groupe $\mathrm{GL}_n(\mathbb{Z})$ sur l'ensemble $\mathbb{Z}^n$ de manière naturelle, deux points X et Y qui sont dans la même orbite vérifient $\Lambda(Y) = \Lambda(X)$. Il est naturel de se demander si cela caractérise les orbites. C'est bien le cas. En effet, si $\Lambda(X) = 1$, X est dans l'orbite du vecteur $(1, 0, \ldots, 0)$ car il existe une matrice M de $\mathrm{GL}_n(\mathbb{Z})$ dont la première colonne est X (cf. exercice 1.39). Par suite, pour tout $d \in \mathbb{N}$, les éléments X de $\mathbb{Z}^n$ vérifiant $\Lambda(X) = d$ sont dans l'orbite de $(d, 0, \ldots, 0)$ (il suffit de multiplier par d).

L'énoncé suivant utilise aussi la réduction modulo un nombre premier.

1.41. Un problème de poids

1. Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ avec $a_{ii} = 0$ pour tout i et $a_{ij} \in \{\pm 1\}$ pour $i \neq j$. Si n est pair, montrer que A est inversible.

2. On dispose de $2n+1$ cailloux, $n \geqslant 1$. On suppose que chaque sous ensemble de $2n$ cailloux peut se partager en deux paquets de n cailloux de même masse totale. Montrer que tous les cailloux ont la même masse.

(École normale supérieure)

▷ **Solution.**

1. On va montrer que le déterminant de A est un entier impair. Pour cela on passe dans $\mathbb{Z}/2\mathbb{Z}$. La classe de $\det A$ modulo 2 est égale au déterminant de

$$\begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \ldots & 1 & 0 \end{pmatrix} \in \mathcal{M}_n(\mathbb{Z}/2\mathbb{Z}).$$

En additionnant les $n-1$ premières colonnes à la dernière, on obtient une n-ième colonne ne contenant que des $n-1$ et par multilinéarité

$$\det A \equiv -\det \begin{pmatrix} 0 & 1 & \ldots & 1 \\ 1 & 0 & \ldots & 1 \\ \vdots & & \ddots & \vdots \\ 1 & \ldots & 1 & 1 \end{pmatrix} \quad [2]$$

car n est pair. En retranchant cette dernière colonne à toute les autres, on obtient

$$\det \mathrm{A} \equiv -\det \begin{pmatrix} -1 & 0 & \dots & 0 & 1 \\ 0 & -1 & \dots & 0 & 1 \\ \vdots & & \ddots & & \vdots \\ 0 & \dots & \dots & -1 & 1 \\ 0 & \dots & \dots & 0 & 1 \end{pmatrix} \equiv -(-1)^{n-1} \equiv 1 \ [2].$$

2. Notons $x_1, \dots, x_{2n+1}$ les masses des cailloux. Soit i fixé. On peut trouver deux sous-ensembles disjoints A_i, B_i de $[\![1, 2n+1]\!] \backslash \{i\}$ de cardinal n tels que $\sum_{k \in \mathrm{A}_i} x_k = \sum_{k \in \mathrm{B}_i} x_k$. Cela s'écrit aussi $\sum_{j=1}^{2n+1} a_{ij} x_j = 0$ où $a_{ii} = 0$, $a_{ij} = 1$ si $j \in \mathrm{A}_i$ et $a_{ij} = -1$ si $j \in \mathrm{B}_j$. On dispose donc d'une matrice $\mathrm{A} = (a_{ij}) \in \mathrm{M}_{2n+1}(\mathbb{R})$ de diagonale nulle et dont les autres coefficients valent ± 1 telle que $\mathrm{AX} = 0$, avec $\mathrm{X} = (x_1, \dots, x_{2n+1})$. Par ailleurs, chaque ligne de A contient exactement n coefficients égaux à 1 et n coefficients égaux à -1. Le vecteur $\mathrm{U} = (1, 1, \dots, 1)$ est donc aussi dans $\operatorname{Ker} \mathrm{A}$. Or, d'après la question 1, le mineur principal de taille $2n$ est non nul. Donc le rang de A est $2n$ et $\dim \operatorname{Ker} \mathrm{A} = 1$. Les vecteurs X et U sont donc proportionnels. ◁

Le dernier thème de ce chapitre est la décomposition LU. *On dit qu'une matrice* A *admet une décomposition* LU *si elle peut s'écrire comme produit d'une matrice* L *triangulaire inférieure (lower) et d'une matrice* U *triangulaire supérieure (upper). Dans l'exercice suivant, on cherche à quelle condition une telle décomposition existe pour une matrice inversible. La méthode exposée fournit un algorithme pour déterminer* L *et* U *basée sur la méthode du pivot de Gauss. La décomposition* LU *est utilisée dans la résolution numérique d'un système linéaire. Elle ramène la résolution de* $\mathrm{AX} = \mathrm{Y}$ *à la résolution successive de deux systèmes triangulaires :* $\mathrm{LZ} = \mathrm{Y}$ *et* $\mathrm{UX} = \mathrm{Z}$.

1.42. Décomposition LU avec pivot de Gauss

Soit $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$.

1. Donner la matrice qui, multipliée à gauche de A, va intervertir les lignes i et j. Quel est son déterminant ?

2. Donner la matrice qui, multipliée à gauche de A, va ajouter à la i-ième ligne la j-ième ligne multipliée par $\lambda \in \mathbb{R}$. Quel est son déterminant ? son inverse ?

3. On suppose dorénavant A inversible. Montrer qu'on peut trouver $M \in GL_n(\mathbb{R})$ telle que $MA = U$ où U est triangulaire supérieure.

4. Préciser à quelles conditions on peut écrire $A = LU$, où U (resp. L) est triangulaire supérieure (resp. inférieure).

(École polytechnique)

▷ **Solution.**

1. Soit $1 \leqslant i, j \leqslant n$, $i \neq j$. Considérons la matrice $M_{i,j}$ obtenue en échangeant les lignes i et j de la matrice I_n, c'est-à-dire

$$M_{ij} = \sum_{k \neq i,j} E_{kk} + E_{ij} + E_{ji}.$$

Pour tout couple d'indice (p, q), on obtient, $M_{ij}E_{pq} = E_{pq}$ si $p \neq i, j$, $M_{ij}E_{iq} = E_{jq}$ et $M_{ij}E_{jq} = E_{iq}$. La multiplication à gauche par M_{ij} échange les lignes i et j des matrices de base. Par linéarité de la multiplication par M_{ij}, il en est de même pour toute matrice de $\mathcal{M}_n(\mathbb{R})$.

Comme M_{ij} est obtenue en échangeant deux lignes de I_n, on a $\det M_{ij} = -\det I_n = -1$.

2. Soit $1 \leqslant i, j \leqslant n$, $i \neq j$. Posons $T_{ij}(\lambda) = I_n + \lambda E_{ij}$. Pour tout couple d'indice (p, q), on obtient

$$T_{ij}(\lambda)E_{pq} = E_{pq} \text{ si } p \neq j \quad \text{et} \quad T_{ij}(\lambda)E_{jq} = E_{jq} + \lambda E_{iq}.$$

Ainsi, pour les matrices de base, la multiplication à gauche par $T_{ij}(\lambda)$ ajoute à la i-ième ligne de la matrice sa j-ième ligne multipliée par λ. Par linéarité, il en est de même pour toute matrice de $\mathcal{M}_n(\mathbb{R})$.

La matrice $T_{ij}(\lambda)$ est triangulaire supérieure ou inférieure selon que $i < j$ ou $i > j$, avec des 1 sur la diagonale. On a donc $\det(T_{ij}(\lambda)) = 1$. La multiplication à gauche par $T_{ij}(-\lambda)$ ajoute à la i-ième ligne d'une matrice la j-ième multipliée par $-\lambda$. On a donc $T_{ij}(-\lambda)T_{ij}(\lambda)I_n = I_n$ et $(T_{ij}(\lambda))^{-1} = T_{ij}(-\lambda)$.

3. On applique la méthode du pivot de Gauss. La matrice $A = (a_{ij})$ étant inversible, sa première colonne n'est pas nulle. Si $a_{i_1 1} \neq 0$, la matrice $M_{1i_1}A = (b_{ij})$ vérifie $b_{11} \neq 0$. En remplaçant, pour $2 \leqslant i \leqslant n$, la ligne L_i de cette matrice par $L_i - \dfrac{b_{i1}}{b_{11}}L_1$, *i.e.* en multipliant à gauche par les matrices $T_{i1}\left(-\dfrac{b_{i1}}{b_{11}}\right)$, on obtient une matrice $A^{(1)} = (a_{ij}^{(1)})$, dont les termes de la première colonne sont nuls à part le premier.

Si on note B_1 la matrice obtenue en supprimant la première ligne et la première colonne de $A^{(1)}$, on a $\det(A^{(1)}) = a_{11}^{(1)} \det B_1 \neq 0$. La matrice B_1 est donc inversible et sa deuxième colonne possède un terme non nul $a_{i_2 2}^{(1)}$, avec $i_2 \geqslant 2$. En multipliant $A^{(1)}$ à gauche par M_{2i_2} on obtient une

matrice dont le terme d'indice $(2,2)$ est non nul. En multipliant à gauche la matrice obtenue par des matrices de la forme $T_{i2}(\lambda)$ pour $3 \leqslant i \leqslant n$, on obtient une matrice $A^{(2)}$ de la forme

$$\begin{pmatrix} a_{11}^{(2)} & a_{12}^{(2)} & \dots & a_{1n} \\ 0 & a_{22}^{(2)} & \dots & a_{2n}^{(2)} \\ 0 & 0 & & \\ \vdots & \vdots & B_2 & \\ 0 & 0 & & \end{pmatrix},$$

où $a_{11}^{(2)}$ et $a_{22}^{(2)}$ ne sont pas nuls et où B_2 est inversible.

On poursuit le procédé. On obtient une suite de matrice $A^{(1)}$, $A^{(2)}, \dots, A^{(n-1)}$ dans laquelle la matrice $A^{(p)}$ a tous ses termes d'indice (i,j) avec $j \leqslant p$ et $i > j$ qui sont nuls. La matrice $A^{(n-1)}$ est donc triangulaire supérieure. On pose $A^{(n-1)} = U$. La matrice U a été obtenue en multipliant à gauche A par de matrices de la forme M_{ij} ou $T_{ij}(\lambda)$. Le produit de ces matrices inversibles est une matrice inversible qu'on note M. On a donc $MA = U$.

4. Pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$ et $1 \leqslant p \leqslant n$, on note $A_p \in \mathcal{M}_p(\mathbb{R})$ la matrice obtenue en gardant les p premières lignes et les p premières colonnes de la matrice A.

• Supposons qu'il existe une matrice L triangulaire inférieure et une matrice U triangulaire supérieure telles que $A = LU$. On a alors pour tout $p \in [\![1,n]\!]$, $A_p = L_p U_p$. En effet, L est de la forme $\begin{pmatrix} L_p & 0 \\ \times & \times \end{pmatrix}$ et U de la forme $\begin{pmatrix} U_p & \times \\ 0 & \times \end{pmatrix}$. Par conséquent, A est de la forme $\begin{pmatrix} L_p U_p & \times \\ \times & \times \end{pmatrix}$ et $A_p = L_p U_p$.

La matrice A est inversible, donc les matrices L et U sont inversibles. Comme elles sont triangulaires, cela signifie que leurs termes diagonaux sont non nuls. Mais alors, pour $1 \leqslant p \leqslant n$, les matrices U_p et L_p, qui sont également triangulaires, ont leurs termes diagonaux non nuls. Elles sont donc inversibles et leur produit A_p est inversible. On a donc

$$\forall p \in [\![1,n]\!], \quad \det(A_p) \neq 0.$$

• Montrons que cette condition est suffisante. Soit donc $A \in \mathcal{M}_n(\mathbb{R})$ telle que, pour tout $p \in [\![1,n]\!]$, $\det A_p \neq 0$. Reprenons la construction effectuée dans la question 3.

On a $\det A_1 = a_{11} \neq 0$. D'emblée le coefficient d'indice $(1,1)$ est non nul. Il n'est donc pas besoin de multiplier A par une matrice du type M_{1i}. On obtient $A^{(1)}$ en multipliant A par un produit de matrices de la forme $T_{i1}(\lambda)$ avec $2 \leqslant i \leqslant n$. La matrice $A^{(1)}$ s'obtient à partir de A en effectuant des opérations élémentaires sur les lignes. Il en est de

même de $A_2^{(1)}$ à partir de A_2. On a donc $\det(A_2^{(1)}) = \det A_2 \neq 0$. Comme $A_2^{(1)} = \begin{pmatrix} a_{11} & * \\ 0 & a_{22}^{(1)} \end{pmatrix}$ est triangulaire supérieure, on a $a_{22}^{(1)} \neq 0$.

D'emblée le terme d'indice (2,2) de $A^{(1)}$ n'est pas nul. La matrice $A^{(2)}$ va s'obtenir à partir de $A^{(1)}$ en multipliant à gauche seulement par des matrices de la forme T_{i2}, avec $3 \leqslant i \leqslant n$. Le raisonnement peut se poursuivre. Si $A^{(p-1)}$ est construite, elle est déduite de A par des transformations élémentaires sur les lignes. Il en est de même de $A_p^{(p-1)}$ à partir de A_p. On obtient $\det A_p^{(p-1)} = \det A_p \neq 0$. Comme par construction $A_p^{(p-1)}$ est triangulaire supérieure, on a en particulier $a_{pp}^{(p-1)} \neq 0$ et on peut se servir de ce terme comme pivot.

Au total, la matrice M est le produit de matrices de la forme $T_{ij}(\lambda)$, avec $i > j$ (en effet, à la première étape, on a $j = 1$, $i \geqslant 2$, à la seconde $j = 2$, $i \geqslant 3, \ldots$). Toutes ces matrices sont triangulaires inférieures, avec des 1 sur la diagonale. Il en résulte que leur produit M est triangulaire inférieure, avec des 1 sur la diagonale. La matrice est inversible et $L = M^{-1}$ est triangulaire inférieure, avec des 1 sur la diagonale. De $MA = U$, on déduit $A = LU$.

Conclusion. La matrice $A \in GL_n(\mathbb{R})$ peut s'écrire sous la forme LU, avec L triangulaire inférieure et U triangulaire supérieure si et seulement si, pour tout $p \in [\![1, n]\!]$, on a $\det A_p \neq 0$. Quand la décomposition existe, on peut supposer de plus que la diagonale de L est constituée de 1. ◁

Lorsque A *est inversible, la décomposition* $A = LU$ *quand elle existe est unique, si on impose que* L *a une diagonale de* 1. *En effet, si on a deux telles décompositions* $A = LU = L'U'$, *on a alors* $L'^{-1}L = U'U^{-1}$. *Comme* $L'^{-1}L$ *est triangulaire inférieure avec une diagonale de* 1 *et* $U'U^{-1}$ *triangulaire supérieure, on a nécessairement* $L'^{-1}L = U'U^{-1} = I_n$, *et donc* $L = L'$, $U = U'$.

Pour démontrer l'existence de la décomposition LU, *d'autres méthodes que celle du pivot existent. On peut par exemple raisonner par récurrence sur la taille n de la matrice, la propriété étant évidente pour* $n = 1$. *Si elle est vérifiée au rang* $n-1$, *il existe* L_1 *et* U_1 *respectivement triangulaire inférieure et supérieure, avec des* 1 *sur la diagonale de* L_1 *telles que* $A_{n-1} = L_1U_1$. *On écrit* A *sous la forme* $A = \begin{pmatrix} A_{n-1} & Y \\ {}^tX & a \end{pmatrix}$, *où* X *et* Y *sont des matrices de* $\mathcal{M}_{n-1,1}(\mathbb{R})$ *et* $a \in \mathbb{R}$ *et on cherche* L *et* U *sous la forme* $L = \begin{pmatrix} L_1 & 0 \\ {}^tX_1 & 1 \end{pmatrix}$ *et* $U = \begin{pmatrix} U_1 & Y_1 \\ 0 & a_1 \end{pmatrix}$, *avec* $a_1 \in \mathbb{R}^*$, X_1 *et* Y_1 *dans* $\mathcal{M}_{n-1,1}(\mathbb{R})$. *On obtient*

$$LU = \begin{pmatrix} L_1U_1 & L_1Y_1 \\ {}^tX_1U_1 & {}^tX_1Y_1 + a_1 \end{pmatrix}.$$

On veut $Y = L_1Y_1$, ${}^tX = {}^tX_1U_1$ *et* ${}^tX_1Y_1 + a_1 = a$, *soit* $Y_1 = L_1^{-1}Y$,

$Y_1 = ({}^tU_1)^{-1}X$ *et* $a_1 = a - {}^tXU_1^{-1}L^{-1}Y = a - {}^tXA_{n-1}^{-1}Y$. *Comme* $\det A = a_1 \det U_1 \det L_1$, *on a* $a_1 \neq 0$. *Les matrices ainsi construites conviennent donc.*

Notons pour finir qu'on peut démontrer que pour toute matrice A *inversible, on peut trouver une matrice de permutation* P *telle que* PA *admette une décomposition LU. Le lecteur intéressé pourra faire le lien avec la décomposition de Bruhat qui a été présentée dans l'exercice 7.18 du tome algèbre 1.*

L'exercice suivant regroupe deux énoncés sur la décomposition LU *des matrices tridiagonales.*

1.43. Décomposition LU d'une matrice tridiagonale

Soit $(a_1, \dots, a_n) \subset \mathbb{R}^n$, $(b_1, \dots, b_{n-1}) \in \mathbb{R}^{n-1}$, $(c_1, \dots, c_{n-1}) \in \mathbb{R}^{n-1}$ et $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ définie par

$$\forall i \in [\![1, n]\!],\ a_{ii} = a_i,\ \forall i \in [\![1, n-1]\!],\ a_{i,i+1} = b_i,\ \forall i \in [\![2, n]\!],\ a_{i,i-1} = c_i$$

et $a_{ij} = 0$ sinon. Pour $k \in [\![1, n]\!]$, on pose $\delta_k = \det(a_{ij})_{1 \leqslant i,j \leqslant k}$.

1. Établir une relation de récurrence sur les δ_k.

2. On suppose que tous les δ_k sont non nuls. Établir l'existence d'une décomposition de A sous la forme LU :

$$A = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_1 & 1 & 0 & \dots & 0 \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ 0 & 0 & \dots & l_{n-1} & 1 \end{pmatrix} \begin{pmatrix} \delta_1 & b_1 & 0 & \dots & 0 \\ 0 & \dfrac{\delta_2}{\delta_1} & b_2 & \dots & 0 \\ & & \ddots & \ddots & \\ & & & \ddots & b_{n-1} \\ 0 & 0 & 0 & \dots & \dfrac{\delta_n}{\delta_{n-1}} \end{pmatrix}.$$

3. Déterminer les déterminants δ_k et la décomposition LU quand elle existe, dans le cas où

$$\forall i \in [\![1, n]\!],\ a_i = 2b\ ; \quad \forall i \in [\![1, n-1]\!],\ b_i = c_i = -1.$$

(École polytechnique)

▷ **Solution.**

1. Pour $k \in [\![1, n]\!]$, on note $A_k \in \mathcal{M}_k(\mathbb{R})$ la matrice constituée des k premières lignes et colonnes de A et $\delta_k = \det A_k$. On a

$$A_k = \begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ c_1 & a_2 & \ddots & & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & a_{k-1} & b_{k-1} \\ 0 & \dots & 0 & c_{k-1} & a_k \end{pmatrix}.$$

On obtient $\delta_1 = a_1$, $\delta_2 = a_1a_2 - b_1c_1$ et pour $k \geq 3$, on obtient en développant selon la dernière ligne

$$\delta_k = a_k\delta_{k-1} - c_{k-1}\det\begin{pmatrix} a_1 & b_1 & 0 & \dots & 0 \\ c_1 & a_2 & \ddots & & \vdots \\ 0 & \ddots & \ddots & b_{k-2} & \vdots \\ \vdots & & \ddots & a_{k-2} & 0 \\ 0 & \dots & 0 & c_{k-2} & b_{k-1} \end{pmatrix}$$

et en développant par rapport à la dernière colonne

$$\boxed{\delta_k = a_k\delta_{k-1} - b_{k-1}c_{k-1}\delta_{k-2}}.$$

2. Soit $(l_1, \dots, l_{n-1}) \in \mathbb{R}^{n-1}$. On pose $L = (l_{ij})$ et $U = (u_{ij})$, avec

$$l_{ii} = 1 \text{ si } i \in [\![1, n]\!], \qquad l_{i,i-1} = l_{i-1} \text{ si } i \in [\![2, n]\!];$$
$$u_{1,i} = \frac{\delta_i}{\delta_{i-1}} \text{ si } i \in [\![1, n]\!], \qquad u_{i,i+1} = b_i \text{ si } i \in [\![1, n-1]\!].$$

et $l_{ij} = u_{ij} = 0$ sinon, en posant $\delta_0 = 1$. On cherche à montrer l'existence de $(l_1, \dots, l_{n-1})$ tel que $A = LU$. On note que la matrice U est bien définie car les δ_k ne sont pas nuls. Il faut, pour tout $(i, j) \in [\![1, n]\!]$,

$$a_{ij} = \sum_{k=1}^{n} l_{ik}u_{kj} = u_{ij} + l_{i-1}u_{i-1,j}.$$

car $l_{ik} = 0$ si $k \neq i$ et $k \neq i-1$. Cela équivaut à $a_1 = u_{11} = \dfrac{\delta_1}{\delta_0}$ et

$$\forall i \in [\![2, n]\!], \qquad a_i = u_{ii} = \frac{\delta_i}{\delta_{i-1}} + l_{i-1}b_{i-1}$$
$$\forall i \in [\![1, n-1]\!], \quad b_i = u_{i,i+1} = u_{i,i+1} = b_i$$
$$\forall i \in [\![2, n]\!], \qquad c_{i-1} = a_{i,i-1} = l_{i-1}u_{i-1,i-1} = l_{i-1}\frac{\delta_{i-1}}{\delta_{i-2}},$$

car pour $j \notin \{i-1, i, i+1\}$, les deux membres de l'égalité sont nuls. L'égalité $a_1 = \dfrac{\delta_1}{\delta_0}$ est vérifiée. On doit avoir, pour $2 \leqslant i \leqslant n$,

$$\boxed{l_{i-1} = \frac{c_{i-1}\delta_{i-2}}{\delta_{i-1}}}.$$

ce qui définit $(l_1, \dots, l_{n-1})$. Il ne reste plus qu'à vérifier que, pour tout $i \in [\![2, n]\!]$,

$$a_i = \frac{\delta_i}{\delta_{i-1}} + l_{i-1}b_{i-1} = \frac{\delta_i}{\delta_{i-1}} + \frac{c_{i-1}b_{i-1}\delta_{i-2}}{\delta_{i-1}},$$

c'est-à-dire

$$\delta_i = a_i\delta_{i-1} - b_{i-1}c_{i-1}\delta_{i-2},$$

ce qui résulte de la première question. On remarque que cette décomposition est unique.

Il résulte de l'exercice 1.42 que la non-nullité de tous les δ_k est aussi une condition nécessaire pour l'existence d'une telle décomposition.

3. On a, toujours avec $\delta_0 = 1$, pour $2 \leqslant k \leqslant n$.

$$\delta_k = 2b\delta_{k-1} - \delta_{k-2}.$$

La suite $(\delta_k)_{k\geqslant 0}$ est récurrente linéaire d'ordre 2. Intéressons-nous aux racines de $r^2 - 2br + 1 = 0$ dont le discriminant est $\Delta = 4b^2 - 4$.

• Premier cas : $|b| > 1$.

Dans ces conditions, $\Delta > 0$. Les racines de $r^2 - 2br + 1 = 0$ sont réelles et distinctes $b + \sqrt{b^2-1}$ et $b - \sqrt{b^2-1}$. Il existe donc A et B réels tels que pour tout $k \in \mathbb{N}$,

$$\delta_k = \mathrm{A}(b + \sqrt{b^2-1})^k + \mathrm{B}(b - \sqrt{b^2-1})^k$$

En écrivant cette égalité pour $k = 0$ ou $k = 1$, on obtient le système

$$\begin{cases} \mathrm{A} + \mathrm{B} = 1 \\ \mathrm{A}(b + \sqrt{b^2-1}) + \mathrm{B}(b - \sqrt{b^2-1}) = 2b \end{cases}$$

qui donne $\mathrm{A} = \dfrac{b + \sqrt{b^2-1}}{2\sqrt{b^2-1}}$ et $\mathrm{B} = -\dfrac{b - \sqrt{b^2-1}}{2\sqrt{b^2-1}}$, ce qui permet d'écrire pour $k \geqslant 0$.

$$\boxed{\delta_k = \frac{1}{2\sqrt{b^2-1}}\left[(b + \sqrt{b^2-1})^{k+1} - (b - \sqrt{b^2-1})^{k+1}\right].}$$

La condition $\delta_k = 0$ équivaut à $(b + \sqrt{b^2-1})^{k+1} = (b - \sqrt{b^2-1})^{k+1}$. Dans ces conditions,

$$b + \sqrt{b^2-1} = \pm(b - \sqrt{b^2-1}) \text{ et en élevant au carré } b\sqrt{b^2-1} = 0,$$

ce qui est impossible puisque $|b| > 1$.

Par conséquent, si $|b| > 1$, on a, pour tout $k \in [\![1, n]\!]$, $\delta_k \neq 0$. La matrice A peut s'écrire sous la forme LU.

Si $b > 1$, il existe $t > 0$ tel que $b = \operatorname{ch} t$. Dans ces conditions, on obtient $b + \sqrt{b^2-1} = e^t$, $b - \sqrt{b^2-1} = e^{-t}$ et pour tout $k \geqslant 0$,

$$\boxed{\delta_k = \frac{\operatorname{sh}(k+1)t}{\operatorname{sh} t}}.$$

Si $b < -1$, il existe $t > 0$ tel que $b = -\operatorname{ch} t$. Dans ces conditions, on obtient, pour tout $k \geqslant 0$.

$$\boxed{\delta_k = (-1)^k \frac{\operatorname{sh}(k+1)t}{\operatorname{sh} t}}.$$

• Deuxième cas : $b = \pm 1$.

Dans ces conditions, $\Delta = 0$ et b est la seule racine de $r^2 - 2br + 1 = 0$. Il existe alors A et B réels tels que pour tout $n \geqslant 0$, $\delta_k = b^k(Ak + B)$. On a $D_0 = 1 = B$ et $D_1 = 2b = b(A + 1)$, ce qui donne $A = B = 1$ et

$$\boxed{\delta_k = (k+1) \text{ si } b = 1 \quad \text{et} \quad \delta_k = (-1)^k(k+1) \text{ si } b = -1}.$$

Là encore, aucun δ_k n'est nul et A possède une décomposition LU.

• Troisième cas : $|b| < 1$.

Dans ces conditions, $\Delta < 0$. Les racines de $r^2 - 2br + 1 = 0$ sont complexes conjuguées :

$$b + i\sqrt{1-b^2} \quad \text{et} \quad b - i\sqrt{1-b^2}.$$

Posons $\theta = \arccos b \in]0, \pi[$. On a alors $\sqrt{1-b^2} = \sin\theta$ et les racines sont $e^{i\theta}$ et $e^{-i\theta}$. Il existe A et B dans $\mathbb{C}$ tels que pour tout $k \geqslant 0$ on ait

$$\delta_k = Ae^{ik\theta} + Be^{-ik\theta}.$$

On a $A + B = 1$ et $Ae^{i\theta} + Be^{-i\theta} = 2\cos\theta$. Il vient $A = \dfrac{e^{i\theta}}{2i\sin\theta}$ et $B = -\dfrac{e^{i\theta}}{2i\sin\theta}$ et $\delta_k = \dfrac{1}{2i\sin\theta}\left[e^{i(k+1)\theta} - e^{-i(k+1)\theta}\right]$ ou plus simplement

$$\boxed{\delta_k = \frac{\sin(k+1)\theta}{\sin\theta}}.$$

$\delta_k = 0$ équivaut à $(k+1)\theta \in \pi\mathbb{Z}$. Comme $\theta \in]0, \pi[$, $\delta_k = 0$ si, et seulement si, $\theta = \dfrac{l\pi}{k+1}$ avec $1 \leqslant l \leqslant k$. On conclut que la décomposition LU de la matrice A existe si et seulement si b n'est pas de la forme $b = \cos\dfrac{l\pi}{k+1}$ avec $1 \leqslant l \leqslant k \leqslant n$. ◁

Chapitre 2
Réduction

On doit à Camille Jordan (1838-1922) de nombreux résultats sur la réduction des endomorphismes, qu'il découvre notamment à travers l'étude des groupes. Dépassant la notion des groupes de permutations pour en atteindre une plus abstraite, il s'intéresse à la classification des groupes finis à travers leurs représentations linéaires, autrement dit les morphismes d'un groupe fini G *dans le groupe linéaire* $\mathrm{GL}(\mathrm{E})$ *d'un espace vectoriel* E*. Il va même jusqu'à donner la description des classes de similitude à l'aide des formes dites de Jordan.*

Le problème fondamental de la réduction est bien celui de caractériser les classes de similitude de l'algèbre $\mathcal{L}(\mathrm{E})$ *où* E *est un* K*-espace vectoriel de dimension finie ou, ce qui revient au même, les classes de similitude de l'algèbre* $\mathcal{M}_n(\mathrm{K})$*. La recherche d'une matrice la plus simple possible pour représenter un endomorphisme donné vise de multiples buts : calculer les puissances successives de cet endomorphisme, son commutant, résoudre des systèmes différentiels linéaires... Une idée naturelle pour essayer de «réduire» l'étude d'un endomorphisme* u *donné à des choses plus simples consiste à essayer de décomposer l'espace vectoriel* E *en une somme directe de sous-espaces non triviaux stables par* u*. Cela n'est évidemment pas toujours possible (le lecteur se reportera aux exercices 2.50, 2.51 et 2.54 concernant la simplicité et la semi-simplicité). Les sous-espaces stables les plus simples sont ceux sur lesquels* u *coïncide avec une homothétie. On est ainsi naturellement amené à la notion de valeur propre. Si* λ *est un scalaire, on s'intéresse donc au sous-espace* $\mathrm{E}_\lambda = \mathrm{Ker}(u - \lambda\,\mathrm{Id}_\mathrm{E})$ *appelé sous-espace propre pour la valeur propre* λ *lorsqu'il n'est pas nul. Le théorème de décomposition des noyaux nous assure que les différents sous-espaces propres d'un endomorphisme sont en somme directe. Le cas où la somme remplit tout l'espace* E *mène à la notion d'endomorphisme diagonalisable : un tel endomorphisme peut être représenté par une matrice diagonale (il suffit de prendre une base formée de vecteurs propres). Pour les endomorphismes diagonalisables il est alors très facile de répondre à la question initiale de savoir quand ils sont semblables : il faut et suffit qu'ils aient les mêmes valeurs propres et que les espaces propres associés aient la même dimension. Il est aussi facile, en se ramenant à une matrice diagonale, de calculer les puissances d'un tel endomorphisme, son exponentielle (si on travaille sur un sous-corps de* $\mathbb{C}$*), son commutant... Cependant il ne s'agit là que d'une réponse partielle à notre problème. Un pas supplémentaire vers sa résolution, lorsque le corps de base est algébriquement clos (par exemple*

sur $\mathbb{C}$), consiste à établir le théorème de décomposition de Dunford : tout endomorphisme u d'un $\mathbb{C}$-espace vectoriel de dimension finie s'écrit de manière unique sous la forme $u = d + n$ où d est diagonalisable, n est nilpotent et commute avec d. Ce résultat, qui n'est pas dans les programmes actuels des classes préparatoires, fait l'objet de l'exercice 2.40. Cette décomposition amène l'attention sur les classes de similitude des endomorphismes nilpotents. Avec un peu de travail il est alors possible d'obtenir le théorème de Jordan qui règle complètement la question de la détermination des classes de similitude sur un corps algébriquement clos. Dans le cadre général, l'outil fondamental pour résoudre cette question est la notion d'endomorphisme cyclique : voir les exercices 2.52, 2.59 ainsi que le commentaire qui suit le second exercice. Après l'étude de ces deux exercices, le lecteur pourra sans mal se lancer dans la lecture d'un bon ouvrage d'algèbre linéaire, pour avoir enfin la solution complète au problème posé en introduction !

Voici maintenant la liste des thèmes successifs selon lesquels nous avons regroupé les exercices. Le plan suit la progression logique que nous venons d'évoquer plus haut.

• *Nos premiers exercices concernent les valeurs propres. Un scalaire λ est valeur propre d'un endomorphisme u si et seulement si $u - \lambda \mathrm{Id}_E$ n'est pas injectif donc, comme on est en dimension finie, non bijectif, soit encore de déterminant nul. Autrement dit, les valeurs propres de u sont exactement les racines du polynôme caractéristique χ_u de u. Cela permet de définir la multiplicité d'une valeur propre comme son ordre de multiplicité en tant que racine de χ_u. Il est facile de montrer que cette multiplicité majore la dimension de l'espace propre associé à λ. Les énoncés des exercices concernent des déterminations de spectre, des questions de multiplicité, ou de localisation des valeurs propres. Il y a notamment plusieurs exercices sur les matrices à coefficients réels positifs.*

• *Le thème suivant sera consacré aux polynômes d'endomorphismes (ou de matrices). On sait que le critère de diagonalisation le plus puissant est l'existence d'un polynôme annulateur qui soit scindé à racines simples. Les exercices porteront sur la notion de polynôme minimal, le théorème de décomposition des noyaux...*

• *Viendront ensuite les exercices concernant la diagonalisation, la trigonalisation, puis la réduction de Dunford avec quelques applications. Comme nous l'avons déjà dit, la décomposition $d+n$ ramène le problème de la réduction pour un corps algébriquement clos à la réduction des endomorphismes nilpotents. Plusieurs exercices sont alors consacrés à ces derniers.*

• *Nous regrouperons ensuite les exercices concernant les sous-espaces stables, la notion d'endomorphismes cycliques.*

• *Viendront ensuite diverses applications : recherche de commutant, équations à base de crochet de Lie, équation de Sylvester $AX - XB = C$.*

• *Nous terminerons en envisageant les effets d'un changement du corps de base.*

Par convention, sauf mention contraire, K *désigne un corps commutatif quelconque. Selon l'usage, on identifie souvent les éléments de* K^n *avec les matrices unicolonnes et les matrices avec l'application linéaire canoniquement associée.*

Dans les trois premiers exercices, on recherche les éléments propres d'un endomorphisme d'un espace vectoriel de dimension infinie.

2.1. Éléments propres d'un opérateur intégral

Soit $\mathrm{E} = \mathcal{C}([0,1],\mathbb{R})$. Pour $f \in \mathrm{E}$, on note $\mathrm{T}(f) : [0,1] \longrightarrow \mathbb{R}$ l'application définie par $\mathrm{T}(f)(x) = \displaystyle\int_0^1 \min(x,t) f(t)\mathrm{d}t$.

1. Montrer que $\mathrm{T} \in \mathcal{L}(\mathrm{E})$.

2. Déterminer les valeurs propres et les vecteurs propres de T.

(École polytechnique)

▷ **Solution.**

1. Soit $f \in \mathrm{E}$. Pour tout $x \in [0,1]$, on a

$$\int_0^1 \min(x,t) f(t)\mathrm{d}t = \int_0^x t f(t)\mathrm{d}t + x\int_x^1 f(t)\mathrm{d}t$$

Comme f est continue, $x \longmapsto \displaystyle\int_0^1 \min(x,t) f(t)\mathrm{d}t$ est de classe $\mathcal{C}^1$ et en particulier, elle est continue : l'application T est donc bien définie de E dans E. La linéarité de T est une conséquence de la linéarité de l'intégrale.

2. Soit λ une valeur propre de T et f un vecteur propre associé. Si $\lambda = 0$, comme $\mathrm{T}(f)$ est de classe $\mathcal{C}^1$, on a $\mathrm{T}(f)' = 0$ donc, pour tout $x \in [0,1]$, $\displaystyle\int_x^1 f(t)\mathrm{d}t = 0$. En dérivant cela une nouvelle fois on en déduit que f est nulle. Ainsi 0 n'est pas valeur propre de T.

Supposons $\lambda \neq 0$. La relation $\mathrm{T}(f) = \lambda f$ montre alors que f est nécessairement de classe $\mathcal{C}^1$. En la dérivant on obtient :

$$\forall x \in [0,1], \quad \lambda f'(x) = x f(x) + \int_x^1 f(t)\mathrm{d}t - x f(x) = \int_x^1 f(t)\mathrm{d}t.$$

La fonction f' est elle-même de classe $\mathcal{C}^1$ et en dérivant une deuxième fois, on obtient $\lambda f''(x) = -f(x)$ pour tout x. Ainsi $f'' + af = 0$, avec $a = \dfrac{1}{\lambda}$. Il existe donc $(\mathrm{A},\mathrm{B}) \in \mathbb{R}^2$ tel que

$$\forall x \in [0,1],\quad f(x) = \begin{cases} A\sin(\sqrt{a}x) + B\cos(\sqrt{a}x) & \text{si } a > 0, \\ A\,\mathrm{sh}(\sqrt{-a}x) + B\,\mathrm{ch}(\sqrt{-a}x) & \text{si } a < 0. \end{cases}$$

Comme $f(0) = 0$, B est nul. Comme $\lambda f'(1) = 0$ et $A \neq 0$, on a

$$\cos\sqrt{a} = 0 \text{ si } a > 0 \text{ et } \mathrm{ch}\sqrt{-a} = 0 \text{ si } a < 0.$$

La deuxième condition est impossible. On a donc $\lambda > 0$ et il existe $k \in \mathbb{N}$ tel que $a = \left(\frac{\pi}{2} + k\pi\right)^2$. Finalement, on obtient

$$\lambda = -\frac{1}{\left(\frac{\pi}{2} + k\pi\right)^2} \quad \text{et} \quad f : x \longmapsto A\sin\left(\left(\frac{\pi}{2} + k\pi\right)x\right).$$

Réciproquement, $f : x \longmapsto \sin\left(\left(\frac{\pi}{2} + k\pi\right)x\right)$ et $\lambda = -\frac{1}{\left(\frac{\pi}{2} + k\pi\right)^2}$ vérifient

$$\lambda f'' = -f, \quad f(0) = 0 \quad \text{et} \quad f'(1) = 0.$$

Par intégration entre x et 1, on a $\int_x^1 f = \lambda f'(x)$. La fonction $T(f) - \lambda f$ a donc une dérivée nulle sur $[0,1]$. Elle est donc constante et finalement nulle puisqu'en zéro, elle vaut 0.

Conclusion. Les valeurs propres de T sont les réels $\lambda_k = -\frac{1}{\left(\frac{\pi}{2} + k\pi\right)^2}$ avec $k \in \mathbb{N}$ et chaque valeur propre λ_k correspond à un sous-espace propre de dimension 1 engendré par $f_k : x \longmapsto \sin\left(\left(\frac{\pi}{2} + k\pi\right)x\right)$. ◁

2.2. Endomorphisme d'un espace de suites (1)

Soit $\mathcal{B} = \left\{u \in \mathbb{C}^{\mathbb{Z}},\ u \text{ bornée}\right\}$ et $T : \mathcal{B} \to \mathcal{B}$ qui à la suite $(u_n)_{n\in\mathbb{Z}}$ associe la suite $\left(\frac{u_{n+1} + u_{n-1}}{2}\right)_{n\in\mathbb{Z}}$. Déterminer les valeurs propres et les vecteurs propres de T.

(École normale supérieure)

▷ **Solution.**

Soit λ une valeur propre de T et u un vecteur propre associé. On a donc la relation de récurrence linéaire suivante :

$$(*) \quad \forall n \in \mathbb{Z},\quad u_{n+1} - 2\lambda u_n + u_{n-1} = 0.$$

Si on note $\mathcal{S}$ le sous-espace des suites de $\mathbb{C}^{\mathbb{Z}}$ vérifiant cette relation, il est clair que l'application qui à $v = (v_n)_{n\in\mathbb{Z}} \in \mathcal{S}$ associe le couple $(v_0, v_1) \in \mathbb{C}^2$ est un isomorphisme. Elle est bien linéaire. Si v est dans son

noyau, $v_0 = v_1 = 0$ et on montre par récurrence sur n que $v_n = v_{-n} = 0$ pour tout n. Enfin si on se donne $(a, b) \in \mathbb{C}^2$, on peut construire par récurrence sur $n \in \mathbb{N}$ à l'aide de $(*)$ les termes v_n et v_{-n} de telle manière que $v = (v_n)_{n\in\mathbb{Z}}$ vérifie $(*)$, $v_0 = a$ et $v_1 = b$.

On cherche donc l'intersection du plan $\mathcal{S}$ et de l'espace $\mathcal{B}$ et pour cela on va expliciter une base de $\mathcal{S}$. Considérons α et β les racines complexes (distinctes ou confondues) de $X^2 - 2\lambda X + 1$. Ces racines sont non nulles et les suites $(\alpha^n)_{n\in\mathbb{Z}}$ et $(\beta^n)_{n\in\mathbb{Z}}$ sont dans $\mathcal{S}$. Si $\alpha \neq \beta$, ces suites forment une famille libre et donc une base de $\mathcal{S}$. Si $\alpha = \beta$, on a $\alpha^2 = 1$ et $\alpha = \pm 1$ si bien que $\lambda = \alpha = \pm 1$. Dans ce dernier cas, comme pour les suites indexées par $\mathbb{N}$, on peut constater que $(n\alpha^n)_{n\in\mathbb{Z}}$ est dans $\mathcal{S}$ et forme avec $(\alpha^n)_{n\in\mathbb{Z}}$ une famille libre et donc une base de $\mathcal{S}$.

Considérons à nouveau le vecteur propre u et supposons dans un premier temps que α et β n'aient pas le même module, par exemple que $|\beta| < |\alpha|$. Comme $|\alpha\beta| = 1$, on a $|\beta| < 1 < |\alpha|$. D'après ce qui précède, il existe $(A, B) \in \mathbb{C}^2$ tel que pour tout $n \in \mathbb{Z}$,

$$u_n = A\alpha^n + B\beta^n.$$

En factorisant par α^n, on obtient $|u_n| = |\alpha|^n \left|A + B\left(\frac{\beta}{\alpha}\right)^n\right|$, ce qui est équivalent en $+\infty$ à $|A||\alpha|^n$ si $A \neq 0$. Comme u est bornée, ce cas est à rejeter et $A = 0$. Mais la suite $u_n = B\beta^n$ n'est bornée au voisinage de $-\infty$ que si $B = 0$ et on a donc $u = 0$. Les scalaires λ conduisant à des racines α, β de modules distincts ne sont donc pas des valeurs propres de T.

Supposons donc que $|\alpha| = |\beta| = 1$. Écrivons $\alpha = e^{i\theta}$ et $\beta = \frac{1}{\alpha} = e^{-i\theta}$. On a alors $\lambda = \frac{\alpha + \beta}{2} = \cos\theta$.

• Si $\theta \not\equiv 0 \pmod{\pi}$, alors $\alpha \neq \beta$ et il existe $A, B \in \mathbb{C}$ tels que pour tout $n \in \mathbb{Z}$,

$$u_n = Ae^{in\theta} + Be^{-in\theta}.$$

Cette suite est effectivement un élément non nul de $\mathcal{B}$ quand $(A, B) \neq 0$ et elle vérifie $(*)$ pour $\lambda = \cos\theta$: c'est un vecteur propre pour $\cos\theta$.

• Si $\theta \in \pi\mathbb{Z}$, alors $\lambda = \pm 1$ et il existe $A, B \in \mathbb{C}$ tels que pour tout $n \in \mathbb{Z}$, $u_n = (A + Bn)\lambda^n$. Comme u est bornée, nécessairement $B = 0$. *Effectivement, $u_n = A$ définit pour $A \neq 0$ un vecteur propre dans $\mathcal{B}$* pour $\lambda = 1$, et $u_n = A(-1)^n$ définit pour $A \neq 0$ un vecteur propre dans $\mathcal{B}$ pour $\lambda = -1$.

Conclusion. Le spectre de T est le segment $[-1, 1]$. Si $\lambda = \pm 1$, le sous-espace propre est de dimension 1 et il est engendré par $(\lambda^n)_{n\in\mathbb{Z}}$. Si $\lambda \in]-1, 1[$, le sous-espace propre est de dimension 2 et il est engendré par $(e^{in\theta})_{n\in\mathbb{Z}}$ et $(e^{-in\theta})_{n\in\mathbb{Z}}$, où $\theta = \arccos\lambda$. ◁

2.3. Endomorphisme d'un espace de suites (2)

Soit $\mathcal{B} = \left\{u \in \mathbb{C}^{\mathbb{Z}},\ u \text{ bornée}\right\}$ et $\mathrm{T} : \mathcal{B} \to \mathcal{B}$ qui à la suite $(u_n)_{n\in\mathbb{Z}}$ associe $(u_{n+1} + u_{n+2})_{n\in\mathbb{Z}}$. Déterminer les valeurs propres et les vecteurs propres de T.

(École normale supérieure)

▷ **Solution.**

Si $u = (u_n)_{n\in\mathbb{Z}}$ est dans $\operatorname{Ker}\mathrm{T}$, alors pour tout $n \in \mathbb{Z}$, $u_{n+2} = -u_{n+1}$ et finalement, $u_n = u_0(-1)^n$ pour tout $n \in \mathbb{Z}$. Ainsi, 0 est valeur propre de T et $\operatorname{Ker}\mathrm{T}$ est la droite engendrée par la suite $((-1)^n)_{n\in\mathbb{Z}}$.

Prenons λ une valeur propre non nulle de T et u un vecteur propre associé. La suite u vérifie la relation de récurrence double :

$$(*) \quad \forall n \in \mathbb{Z}, \quad u_{n+2} + u_{n+1} - \lambda u_n = 0.$$

Comme $\lambda \neq 0$, en procédant comme dans l'exercice précédent, on montre que le sous-espace $\mathcal{S}$ des suites $(u_n)_{n\in\mathbb{Z}}$ de $\mathbb{C}^{\mathbb{Z}}$ vérifiant la relation $(*)$ est de dimension 2. Soient α et β les racines (distinctes ou confondues) du trinôme $\mathrm{X}^2 + \mathrm{X} - \lambda$.

Supposons que $\alpha = \beta$. Alors la somme vaut $2\alpha = -1$ donc $\alpha = -\dfrac{1}{2}$ et $\lambda = -\alpha\beta = -\dfrac{1}{4}$. Dans ces conditions, il existe $\mathrm{A}, \mathrm{B} \in \mathbb{C}$ tels que pour tout $n \in \mathbb{Z}$,

$$u_n = (\mathrm{A} + \mathrm{B}n)\left(-\frac{1}{2}\right)^n.$$

En considérant la limite en $-\infty$ de $|u_n|$, le fait que u soit bornée impose $\mathrm{A} = \mathrm{B} = 0$. Mais alors u est nulle et ne peut être un vecteur propre. Ainsi, $-\dfrac{1}{4}$ n'est pas valeur propre de T.

Ce cas étant exclu on a $\alpha \neq \beta$ et il existe $(\mathrm{A}, \mathrm{B})^2 \in \mathbb{C}$ tel que $u_n = \mathrm{A}\alpha^n + \mathrm{B}\beta^n$ pour tout $n \in \mathbb{Z}$.

• Supposons $\mathrm{A} \neq 0$ et $\mathrm{B} \neq 0$. Supposons $|\beta| < |\alpha|$. On a en $+\infty$, $|u_n| \sim |\mathrm{A}||\alpha|^n$. Nécessairement, u étant bornée, on a $|\alpha| \leqslant 1$. En $-\infty$, l'équivalent est $|u_n| \sim |\mathrm{B}||\beta|^n$ et nécessairement $|\beta| \geqslant 1$ ce qui est contradictoire avec l'hypothèse $|\beta| < |\alpha|$. On a donc $|\alpha| = |\beta|$ et comme $\alpha + \beta = -1$, l'axe des abscisses est la médiatrice du segment $[\alpha, \beta]$ (puisque cette médiatrice contient l'origine et le milieu des deux points qui est $-\dfrac{1}{2}$). On en déduit que α et β sont conjugués et ont une partie réelle égale à $-\dfrac{1}{2}$. Quitte à les renommer, on peut supposer que β possède un argument θ dans $\left]\dfrac{\pi}{2}, \pi\right[$. Dans ces conditions,

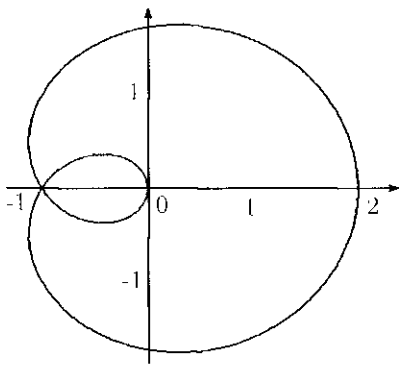
$$u_n = \alpha^n(\mathrm{A} + \mathrm{B}e^{2in\theta}).$$

Classiquement, $e^{2i\theta n}$ ne converge pas (en $+\infty$ et en $-\infty$) car $2\theta \notin 2\pi\mathbb{Z}$. Comme A et B sont toujours supposés différents de 0, $(A + Be^{2i\theta n})$ ne converge pas vers 0. Il existe $\varepsilon > 0$ et une infinité de n dans $\mathbb{N}$ et dans $\mathbb{Z}_-$ tels que $|A + Be^{2i\theta n}| \geqslant \varepsilon$. Ainsi, si $|\alpha| \neq 1$, une sous-suite de $|u_n| = |\alpha|^n|A + Be^{2i\theta n}|$ tend vers $+\infty$ (ou $+\infty$ ou $-\infty$). Comme u est bornée, on a donc nécessairement $|\alpha| = |\beta| = 1$ et finalement $\{\alpha, \beta\} = \{j, j^2\}$ et $\lambda = -1$.

Réciproquement, toute suite $(Aj^n + Bj^{2n})_{n\in\mathbb{Z}}$ avec $(A, B) \subset \mathbb{C}^2$ est élément de $\mathrm{Ker}(T+\mathrm{Id})$: -1 est valeur propre de T et $\dim \mathrm{Ker}(T+\mathrm{Id}) = 2$.

• Si $B = 0$, alors comme $u \neq 0$, on a $A \neq 0$ et on obtient $|\alpha| = 1$ en considérant les limites en $+\infty$ et $-\infty$. Ainsi, il existe $\varphi \in \mathbb{R}$ tel que $\alpha = e^{i\varphi}$ et $\beta = -1 - e^{i\varphi}$, $\lambda = -\alpha\beta = e^{i\varphi} + e^{2i\varphi}$.

Réciproquement, si λ s'écrit $\lambda = e^{i\varphi} + e^{2i\varphi}$ avec $\varphi \in \mathbb{R}$, $e^{i\varphi}$ et $-1-e^{i\varphi}$ sont les racines de $X^2 + X - \lambda$. La suite $u = (e^{in\varphi})_{n\in\mathbb{Z}}$ est vecteur propre pour la valeur propre λ. L'espace propre associé est de dimension 1 sauf si les racines sont j, j^2, ce qui correspond à $\varphi = \pm\dfrac{2\pi}{3} \pmod{2\pi}$ et $\lambda = -1$. Voici la représentation dans le plan complexe du spectre de T.



Remarquons que l'on retrouve la valeur propre $\lambda = 0$ en prenant $\varphi \equiv \pi \pmod{2\pi}$ et que -1 est le point double de la courbe.

Conclusion. Le spectre de T est l'ensemble des nombres complexes s'écrivant $e^{i\varphi} + e^{2i\varphi}$ avec $\varphi \in \mathbb{R}$, ce qui correspond à la courbe donnée en coordonnées cartésiennes par $\begin{cases} x(\varphi) = \cos(\varphi) + \cos(2\varphi) \\ y(\varphi) = \sin(\varphi) + \sin(2\varphi) \end{cases}$. Le sous-espace propre associé est la droite engendrée par $(e^{in\varphi})_{n\in\mathbb{Z}}$ sauf dans le cas $\lambda = -1$ ($\varphi \equiv \pi \pmod{2\pi}$) où le sous-espace propre est un plan engendré par $(j^n)_{n\in\mathbb{Z}}$ et $(j^{2n})_{n\in\mathbb{Z}}$. ◁

En dimension finie la dimension d'un espace propre pour une valeur propre λ est toujours comprise entre 1 et l'ordre de multiplicité de cette valeur propre dans le polynôme caractéristique, toutes les valeurs

intermédiaires étant possibles. Dans l'exercice ci-après on connaît la dimension de l'espace propre et on cherche l'ordre de multiplicité de la valeur propre.

2.4. Valeur propre simple

Soit K un corps commutatif, $\lambda \in K$, $A \in \mathcal{M}_n(K)$, X et Y dans $\mathcal{M}_{n,1}(K)$ tels que

$$AX = \lambda X, \quad {}^tYA = \lambda {}^tY, \quad {}^tYX \neq 0, \quad \mathrm{rg}(A - \lambda I_n) = n - 1.$$

Montrer que λ est valeur propre simple de A.

(École polytechnique)

▷ **Solution.**

Les vecteurs X et Y sont différents de 0 puisque ${}^tYX \neq 0$. Le scalaire λ est valeur propre de A (et aussi de ${}^tA$) et par hypothèse l'espace propre associé à λ pour A est une droite (car $\mathrm{rg}(A - \lambda I_n) = n - 1$). Cette droite est donc dirigée par X. Comme on souhaite calculer l'ordre de multiplicité de λ en tant que racine du polynôme caractéristique de A, on va essayer de calculer ce polynôme, en choisissant une matrice semblable à A qui soit plus simple. Il faut évidemment utiliser les informations que l'on a sur la transposée de A.

Le vecteur Y est un vecteur propre de ${}^tA$ pour la valeur propre λ. Il définit une forme linéaire non nulle sur K^n (par $Z \mapsto {}^tYZ$) dont le noyau est un hyperplan H. L'hypothèse ${}^tYX \neq 0$ signifie que X n'est pas dans cet hyperplan H. On remarque que, pour tout $Z \in H$, on a ${}^tYAZ = \lambda {}^tYZ = 0$ et $AZ \in H$. Ainsi l'hyperplan H est stable par A. Considérons une base de K^n constituée de X et d'une base de H. Dans cette base, l'endomorphisme canoniquement associé à A a une matrice de la forme

$$M = \begin{pmatrix} \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & N & \\ 0 & & & \end{pmatrix},$$

où $N \in \mathcal{M}_{n-1}(K)$. On a donc

$$\chi_A = \chi_M = \det \begin{pmatrix} X - \lambda & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & XI_{n-1} - N & \\ 0 & & & \end{pmatrix} = (X - \lambda)\chi_N.$$

La matrice $\lambda I_n - M$ est semblable à $\lambda I_n - A$, donc elle est également de

rang $n-1$. Comme $\lambda \mathrm{I}_n - \mathrm{M} = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \lambda \mathrm{I}_{n-1} - \mathrm{N} & \\ 0 & & & \end{pmatrix}$, $\lambda \mathrm{I}_{n-1} - \mathrm{N}$ est de rang $n-1$. Ainsi, λ n'est pas valeur propre de N et λ est bien une valeur propre simple de A. ◁

Il est important de retenir que si u est un endomorphisme d'un espace E et f une forme linéaire non nulle sur E, l'hyperplan $\operatorname{Ker} f$ est stable par u si et seulement si f est un vecteur propre de ${}^t u$.

L'énoncé suivant concerne encore des questions de multiplicité de valeur propre.

2.5. Existence d'une valeur propre double

Soient A, B et C trois matrices de $\mathcal{M}_2(\mathrm{K})$. Montrer qu'il existe un triplet $(\alpha, \beta, \gamma) \in \mathrm{K}^3 \backslash \{0\}$ tel que $\alpha \mathrm{A} + \beta \mathrm{B} + \gamma \mathrm{C}$ ait une valeur propre double.

(**École normale supérieure**)

▷ **Solution.**

Bien entendu si la famille $(\mathrm{A}, \mathrm{B}, \mathrm{C})$ est liée le résultat est clair, car on peut trouver une combinaison linéaire nulle des trois matrices avec des coefficients (α, β, γ) non tous nuls.

Supposons donc que la famille $(\mathrm{A}, \mathrm{B}, \mathrm{C})$ est libre. On connaît des matrices simples qui ont une valeur propre double : toutes les matrices triangulaires supérieures de la forme $\begin{pmatrix} \lambda & \mu \\ 0 & \lambda \end{pmatrix}$. Or, l'ensemble de ces matrices forme clairement un sous-espace vectoriel de $\mathcal{M}_2(\mathrm{K})$ de dimension 2. Ce sous-espace ne peut donc pas être en somme directe avec l'espace $\operatorname{Vect}(\mathrm{A}, \mathrm{B}, \mathrm{C})$ qui est de dimension 3 puisque $\dim \mathcal{M}_2(\mathrm{K}) = 4$. Il y a donc une matrice non nulle de la forme $\alpha \mathrm{A} + \beta \mathrm{B} + \gamma \mathrm{C}$ qui admet une valeur propre double. ◁

2.6. Détermination de spectre

On définit une suite de matrices par $\mathrm{A}_1 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ et la relation de récurrence $\mathrm{A}_{n+1} = \begin{pmatrix} \mathrm{A}_n & \mathrm{A}_n \\ \mathrm{A}_n & -\mathrm{A}_n \end{pmatrix}$. Déterminer les valeurs propres de A_n.

(**École polytechnique**)

▷ **Solution.**

On commence naturellement par regarder le cas $n = 1$. Le polynôme caractéristique de A_1 est

$$\chi_1(X) = \begin{vmatrix} X-1 & -1 \\ -1 & X+1 \end{vmatrix} = X^2 - 2.$$

On a donc $\operatorname{Sp} A_1 = \{\pm\sqrt{2}\}$.

Essayons maintenant de trouver une relation entre le polynôme caractéristique de A_n et celui de A_{n+1}. Notons que la matrice A_n est de taille 2^n. On a

$$\chi_{n+1}(X) = \begin{vmatrix} XI_{2^n} - A_n & -A_n \\ -A_n & XI_{2^n} + A_n \end{vmatrix}.$$

Les quatre matrices qui interviennent commutent deux à deux. On a donc (cf. exercice 1.31)

$$\chi_{n+1}(X) = \det(X^2 I_{2^n} - 2A_n^2) = \det(XI_{2^n} - \sqrt{2}A_n)\det(XI_{2^n} + \sqrt{2}A_n).$$

On obtient donc en factorisant $\sqrt{2}$, $\chi_{n+1}(X) = 2^{2^n}\chi_n\left(\frac{X}{\sqrt{2}}\right)\chi_n\left(\frac{-X}{\sqrt{2}}\right)$.

Les racines de χ_{n+1} s'obtiennent en multipliant celles de χ_n par $\pm\sqrt{2}$.

Conclusion. On a pour tout n, $\operatorname{Sp} A_n = \{\pm\sqrt{2}^n\}$. ◁

Pour une matrice donnée de taille importante le calcul des valeurs propres n'est pas chose facile. D'une part, le calcul du polynôme caractéristique n'est pas simple (rappelons que c'est un déterminant) et d'autre part, on sait bien que la recherche des racines d'un polynôme est un problème difficile. Ces questions ayant une grande importance pratique, on a été amené à chercher des algorithmes efficaces. On se préoccupe ici du calcul de χ_A où A est une matrice de $\mathcal{M}_n(\mathbb{C})$. Les coefficients de χ_A sont donnés par les fonctions symétriques élémentaires $\sigma_1, \ldots, \sigma_n$ des valeurs propres $\lambda_1, \ldots, \lambda_n$ (prises avec multiplicité). Une première idée est de calculer ces fonctions symétriques élémentaires à l'aide des formules de Newton qui relient les σ_k aux sommes de Newton $S_k = \sum_{i=1}^{n} \lambda_i^k$. En effet, ces sommes de Newton sont plus faciles à trouver puisque $S_k = \operatorname{Tr} A^k$ pour tout k (trigonaliser A pour le voir). Cela constitue la méthode de Leverrier. L'exercice suivant présente l'algorithme de Faddeev qui consiste à remplacer la suite A^k par une autre suite de matrices avec le même objectif : obtenir les coefficients successifs du polynôme caractéristique. On y utilisera les formules de Newton que le lecteur trouvera dans l'exercice 5.26 du premier tome d'algèbre.

2.7. Algorithme de Faddeev

Soit $A \in \mathcal{M}_n(\mathbb{C})$. On définit une suite $(A_k)_{k\geqslant 0}$ de matrices en posant $A_0 = A$ et pour $k \geqslant 1$,

$$A_k = A\left(A_{k-1} - \frac{1}{k}\operatorname{Tr}(A_{k-1})I_n\right).$$

Montrer que A_n est nulle.

(École polytechnique)

▷ **Solution.**

On commence par calculer les premiers termes de la suite $(A_k)_{k\geq 0}$:

$$A_0 = A\ ; \quad A_1 = A^2 - \operatorname{Tr}(A_0)A\ ; \quad A_2 = A^3 - \operatorname{Tr}(A_0)A^2 - \frac{1}{2}\operatorname{Tr}(A_1)A.$$

Par une récurrence immédiate, on montre alors que pour tout $k \in [\![0, n]\!]$,

$$A_k = A^{k+1} - \operatorname{Tr}(A_0)A^k - \frac{1}{2}\operatorname{Tr}(A_1)A^{k-1} - \cdots - \frac{1}{k}\operatorname{Tr}(A_{k-1})A.$$

Pour calculer A_n, il faut calculer $\operatorname{Tr}(A_k)$ pour tout $k \in [\![0, n-1]\!]$. On note $\lambda_1, \ldots, \lambda_n$ les valeurs propres de A comptées avec multiplicité. Les valeurs propres de A^k sont $\lambda_1^k, \ldots, \lambda_n^k$ et on a $\operatorname{Tr}(A^k) = \sum_{i=1}^{n} \lambda_i^k$. On obtient successivement $\operatorname{Tr}(A_0) = \operatorname{Tr}(A)$, puis

$$\operatorname{Tr}(A_1) = \operatorname{Tr}(A^2) - (\operatorname{Tr}(A))^2 = \sum_{i=1}^{n} \lambda_i^2 - \left(\sum_{i=1}^{n} \lambda_i\right)^2 = -2 \sum_{1\leqslant i<j\leqslant n} \lambda_i\lambda_j.$$

En notant $\sigma_1, \ldots, \sigma_n$ les fonctions symétriques élémentaires des valeurs propres, on a donc $\operatorname{Tr}(A_0) = \sigma_1$ et $\operatorname{Tr}(A_1) = -2\sigma_2$.

Montrons par récurrence sur k que $\operatorname{Tr}(A_k) = (-1)^k(k+1)\sigma_{k+1}$. C'est vrai pour $k = 1$ et $k = 2$ comme on vient de le voir. Si la proposition est vérifiée jusqu'au rang $k-1$, on obtient, en posant $S_p = \sum_{i=1}^{n} \lambda_i^p$,

$$\begin{aligned}\operatorname{Tr}(A_k) &= \operatorname{Tr}(A^{k+1}) - \operatorname{Tr}(A_0)\operatorname{Tr}(A^k) - \frac{1}{2}\operatorname{Tr}(A_1)\operatorname{Tr}(A^{k-1}) - \\ &\qquad - \cdots - \frac{1}{k}\operatorname{Tr}(A_{k-1})\operatorname{Tr}(A) \\ &= S_{k+1} - \sigma_1 S_k + \sigma_2 S_{k-1} - \sigma_3 S_{k-2} + \cdots + (-1)^k \sigma_k S_1.\end{aligned}$$

Les formules de Newton (cf exercice 5.26 du tome 1 d'algèbre) donnent

$$S_{k+1} - \sigma_1 S_k + \cdots + (-1)^k \sigma_k S_1 + (-1)^{k+1}(k+1)\sigma_{k+1} = 0$$

et on a donc $\mathrm{Tr}(\mathrm{A}_k) = -(-1)^{k+1}(k+1)\sigma_{k+1} = (-1)^k(k+1)\sigma_{k+1}$. La proposition est vérifiée au rang $k+1$ et donc pour tout k. On obtient en particulier

$$\mathrm{A}_n = \mathrm{A}^{n+1} - \sigma_1\mathrm{A}^n + \sigma_2\mathrm{A}^{n-1} + \cdots + \sigma_n\mathrm{A}.$$

Comme $\lambda_1, \ldots, \lambda_n$ sont les racines du polynôme caractéristique, on a

$$\chi_\mathrm{A} = \mathrm{X}^n + \sum_{k=0}^{n}(-1)^k\sigma_k\mathrm{X}^{n-k}$$

et donc $\mathrm{A}_n = \mathrm{A}\chi_\mathrm{A}(\mathrm{A})$. D'après le théorème de Cayley-Hamilton, on en déduit $\chi_\mathrm{A}(\mathrm{A}) = 0$ donc $\mathrm{A}_n = 0$. ◁

Les coefficients du polynôme caractéristique sont donc, au signe près, les $\dfrac{1}{k+1}\mathrm{Tr}(\mathrm{A}_k)$ *(pour* $0 \leqslant k \leqslant n-1$*). La suite* (A_k) *définie dans l'énoncé fournit ainsi un algorithme de calcul du polynôme caractéristique appelé algorithme de Fadeev, bien plus rapide que celui qui consiste à calculer* $\det(\mathrm{X}\mathrm{I}_n - \mathrm{A})$.

Notons à cette occasion que c'est le mathématicien allemand Frobenius (1849-1917) qui donna en 1878 la première démonstration du théorème de Cayley-Hamilton (voir l'exercice 4.10 pour une preuve analytique de ce théorème).

Le calcul des valeurs propres étant difficile, on peut être amené à simplement essayer de les localiser. Voici un résultat assez classique de ce type, basé sur le lemme d'Hadamard.

2.8. Lemme d'Hadamard, disques de Gershgörin

Soit $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{C})$. On pose $\mathrm{R}_i = \displaystyle\sum_{\substack{1\leqslant j\leqslant n\\ j\neq i}} |a_{ij}|$ pour tout $i \in [\![1, n]\!]$.

1. On suppose que pour tout $i \in [\![1, n]\!]$, $|a_{ii}| > \mathrm{R}_i$. Montrer que la matrice A est inversible.

2. Montrer que $\mathrm{Sp}\,\mathrm{A} \subset \displaystyle\bigcup_{i=1}^{n}\{z \in \mathbb{C}, |z - a_{ii}| \leqslant \mathrm{R}_i\}$.

3. On suppose à nouveau que pour tout $i \in [\![1, n]\!]$, $|a_{ii}| > \mathrm{R}_i$. Montrer que

$$|\det \mathrm{A}| \geqslant \prod_{i=1}^{n}(|a_{ii}| - \mathrm{R}_i).$$

4. On suppose que $A \in \mathcal{M}_n(\mathbb{R})$ et que pour tout $i \in [\![1, n]\!]$, $a_{ii} > R_i$. Montrer que $\det A \geqslant \prod_{i=1}^{n} (|a_{ii}| - R_i)$.

(École polytechnique)

▷ **Solution.**

1. Raisonnons par l'absurde et supposons A non inversible. Il existe alors un vecteur colonne non nul $X = {}^t(x_1, \ldots, x_n)$ tel que $AX = 0$. Pour tout $1 \leqslant i \leqslant n$, on a $\sum_{j=1}^{n} a_{ij}x_j = 0$ puis, en isolant le terme d'indice i,

$$|a_{ii}x_i| = \left|\sum_{j\neq i} a_{ij}x_j\right| \leqslant \sum_{j\neq i} |a_{ij}||x_j| \leqslant R_i \max_{1\leqslant k\leqslant n} |x_k|.$$

Prenons $i_0 \in [\![1, n]\!]$ tel que $|x_{i_0}| = \max_{1\leqslant k\leqslant n} |x_k|$. On a $|x_{i_0}| > 0$ car $X \neq 0$. On obtient $|a_{i_0i_0}||x_{i_0}| \leqslant R_{i_0}|x_{i_0}|$ et donc $|a_{i_0i_0}| \leqslant R_{i_0}$. Cela contredit l'hypothèse. On conclut que $A \in \mathrm{GL}_n(\mathbb{C})$.

Ce résultat constitue le lemme d'Hadamard.

2. Soit $\lambda \in \mathrm{Sp}(A)$. Comme $A - \lambda I_n$ n'est pas inversible, il existe $i_0 \in [\![1, n]\!]$ tel que $|a_{i_0i_0} - \lambda| \leqslant R_{i_0}$ *i.e.* $\lambda \in \{z \in \mathbb{C}, |z - a_{i_0i_0}| \leqslant R_{i_0}\}$. Finalement,

$$\lambda \in \bigcup_{i=1}^{n} \{z \in \mathbb{C}, |z - a_{ii}| \leqslant R_i\}.$$

Les disques de centre a_{ii} et de rayon R_i dont la réunion contient le spectre de A sont appelés disques de Gershgörin.

3. Notons A' la matrice obtenue à partir de A en multipliant, pour tout $1 \leqslant i \leqslant n$ la i-ième ligne par $\dfrac{1}{|a_{ii}| - R_i}$ (le dénominateur est bien non nul). On obtient

$$\det A = \det A' \prod_{i=1}^{n} (|a_{ii}| - R_i).$$

Il n'y a plus qu'à prouver que $\det A' \geqslant 1$. Posons $A' = (a'_{ij})_{1\leqslant i,j\leqslant n}$. On a pour tout $1 \leqslant i \leqslant n$,

$$|a'_{ii}| - \sum_{j\neq i} |a'_{ij}| = \frac{|a_{ii}| - R_i}{|a_{ii}| - R_i} = \frac{|a_{ii}| - R_i}{|a_{ii}| - R_i} = 1.$$

Pour toute valeur propre λ de A', il existe $1 \leqslant i \leqslant n$ tel que

$$|\lambda - a'_{ii}| \leqslant \sum_{j\neq i} |a'_{ij}|$$

et donc, par l'inégalité triangulaire,

$$|\lambda| - |a'_{ii}| \geqslant -\sum_{j\neq i} |a'_{ij}| \quad \text{et} \quad |\lambda| \geqslant |a'_{ii}| - \sum_{j\neq i} |a'_{ij}| = 1.$$

Toute valeur propre de A' a un module supérieur ou égal à 1. Or le déterminant de A' est le produit de ses valeurs propres (distinctes ou confondues). Donc $|\det A'| \geqslant 1$. On en déduit que

$$|\det A| = |\det A'| \prod_{i=1}^{n} (|a_{ii}| - R_i) \geqslant \prod_{i=1}^{n} (|a_{ii}| - R_i).$$

4. D'après la question 3, il suffit d'établir que $\det A > 0$. Pour cela nous allons utiliser un argument de convexité. Notons $\mathcal{C}$ l'ensemble des matrices $(a_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{R})$ vérifiant

$$\forall i \in [\![1, n]\!], \quad a_{ii} > \sum_{\substack{1\leqslant j\leqslant n \\ j\neq i}} |a_{ij}|$$

et montrons que $\mathcal{C}$ est convexe. Soient $A = (a_{ij})$ et $B = (b_{ij})$ dans $\mathcal{C}$ et $t \in]0, 1[$. Montrons que $(1-t)A + tB = ((1-t)a_{ij} + tb_{ij}) \in \mathcal{C}$.

Pour tout $i \in [\![1, n]\!]$, on a

$$\begin{aligned}
\sum_{j\neq i} |(1-t)a_{ij} + tb_{ij}| &\leqslant \sum_{j\neq i} ((1-t)|a_{ij}| + t|b_{ij}|) \\
&\leqslant (1-t)\sum_{j\neq i} |a_{ij}| + t\sum_{j\neq i} |b_{ij}| \\
&< (1-t)a_{ii} + tb_{ii}.
\end{aligned}$$

Donc $(1-t)A + tB \in \mathcal{C}$ et $\mathcal{C}$ est convexe et en particulier connexe par arcs. Or la fonction $M \in \mathcal{C} \longmapsto \det M$ est continue et, d'après la question 1, ne s'annule pas sur $\mathcal{C}$. Son image est donc un intervalle de $\mathbb{R}$ ne contenant pas 0 : cette image est contenue dans $\mathbb{R}_+^*$ ou dans $\mathbb{R}_-^*$. Comme $I_n \in \mathcal{C}$ et $\det I_n = 1 > 0$, pour tout $M \in \mathcal{C}$, $\det M > 0$. La question 3 donne le résultat voulu. ◁

On peut appliquer le résultat obtenu dans la question 2 à une matrice compagnon pour en déduire un résultat de localisation des racines d'un polynôme quelconque.

Nous commençons maintenant une série d'exercices sur les matrices à coefficients réels positifs et tout d'abord sur les matrices stochastiques. Celles-ci interviennent naturellement en Probabilités dans l'étude des chaînes de Markov (le lecteur pourra se reporter au commentaire précédent l'exercice 7.16 dans notre premier tome d'exercices d'algèbre pour un peu plus de détails). Le premier énoncé montre que le spectre d'une matrice stochastique est entièrement contenu dans le disque unité du plan complexe, résultat que précisera l'énoncé d'après.

2.9. Matrices stochastiques (1)

Soit S l'ensemble des matrices réelles stochastiques, c'est-à-dire l'ensemble des matrices $P = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$ telles que

$$\forall (i,j) \in [\![1,n]\!]^2,\ p_{ij} \geqslant 0 \quad \text{et} \quad \forall i \in [\![1,n]\!],\ \sum_{j=1}^{n} p_{ij} = 1.$$

1. Montrer que tous les éléments de S ont une valeur propre commune.

2. Si P, Q sont dans S, en est-il de même de PQ ?

3. Soit $P \in S$ et λ une valeur propre complexe de P. Montrer que $|\lambda| \leqslant 1$.

(École polytechnique)

▷ **Solution.**

1. Si U est le vecteur de $\mathbb{R}^n$ dont toutes les coordonnées valent 1, la seconde condition qui définit les éléments de S équivaut à PU = U. Donc tous les matrices stochastiques admettent 1 comme valeur propre et U comme vecteur propre associé.

2. Si P et Q sont dans S, les coefficients de PQ sont tous positifs. De plus (PQ)U = PU = U de sorte que la seconde condition est aussi remplie. Donc S est stable pour le produit.

3. Soit $\lambda \in \mathrm{Sp}\, P$ et $X = {}^t(x_1, \ldots, x_n)$ un vecteur propre associé. Choisissons un indice i tel que $|x_i| = \max_{1 \leqslant k \leqslant n} |x_k|$. Comme $PX = \lambda X$ on a, en regardant la i-ième coordonnée,

$$p_{i1}x_1 + \cdots + p_{in}x_n = \lambda x_i.$$

Et en passant au module, il vient

$$|\lambda x_i| = |\lambda||x_i| = |p_{i1}x_1 + \cdots + p_{in}x_n| \leqslant (p_{i1} + \cdots + p_{in})|x_i| = |x_i|.$$

On conclut que $|\lambda| \leqslant 1$. ◁

Dans l'énoncé suivant on prouve que les éventuelles valeurs propres de module 1 sont des racines de l'unité. C'est l'occasion de s'accoutumer au raisonnement suivant : pour montrer plusieurs égalités, on prouve une série d'inégalités dont le premier et le dernier membres sont identiques. Dans ces conditions, les inégalités sont donc toutes des égalités.

2.10. Matrices stochastiques (2)

Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{R})$ une matrice stochastique. Soit λ une valeur propre de A de module 1 et $X = (x_1, \dots, x_n) \in \mathbb{C}^n$ un vecteur propre associé.

1. Si x_i est une composante de X de module maximal, montrer que λx_i est encore une composante de X de module maximal.

2. En déduire que λ est une racine m-ième de l'unité avec $m \leqslant n$.

3. On suppose que pour tout $i \in [\![1, n]\!]$, $a_{ii} \neq 0$. Montrer que la seule valeur propre de A de module 1 est 1.

(**École normale supérieure**)

▷ **Solution.**

1. On suppose $\lambda \neq 1$ sinon la question est triviale. La i-ième ligne de l'égalité $AX = \lambda X$ conduit à :

$$\lambda - a_{ii} = \sum_{j \neq i} a_{ij} \frac{x_j}{x_i}. \tag{1}$$

On en déduit que

$$1 - a_{ii} = |\lambda| - a_{ii} \leqslant |\lambda - a_{ii}| \leqslant \sum_{j \neq i} a_{ij} \left| \frac{x_j}{x_i} \right| \leqslant \sum_{j \neq i} a_{ij} = 1 - a_{ii}.$$

Les inégalités sont donc toutes des égalités. On déduit de cela plusieurs choses.

• L'égalité $1 - a_{ii} = |\lambda - a_{ii}|$ montre que λ est sur le cercle de centre a_{ii} et de rayon $1 - a_{ii}$. Si $a_{ii} > 0$, ce cercle est tangent intérieurement au cercle unité de $\mathbb{C}$ en 1. On a alors $\lambda = 1$ ce qui est exclu. On a donc nécessairement $a_{ii} = 0$. Il en résulte notamment que l'ensemble $I = \{j \neq i,\ a_{ij} \neq 0\}$ n'est pas vide.

• D'après le cas d'égalité dans l'inégalité triangulaire, les nombres complexes $a_{ij} \dfrac{x_j}{x_i}$ pour $j \neq i$ sont tous sur une même demi-droite d'origine 0. Cela est intéressant uniquement si $a_{ij} \neq 0$ c'est-à-dire si $j \in I$.

• Pour $j \in I$ on a $a_{ij} \left| \dfrac{x_j}{x_i} \right| = a_{ij}$, c'est-à-dire $|x_j| = |x_i|$.

En combinant ces deux derniers points, on obtient l'existence d'un réel θ tel que, pour tout $j \in I$, $x_j = e^{i\theta} x_i$. En remplaçant dans la relation (1), on obtient

$$\lambda = e^{i\theta} \sum_{j \in I} a_{ij} = e^{i\theta} \sum_{j=1}^{n} a_{ij} = e^{i\theta}$$

Ainsi, $e^{i\theta} = \lambda$ et $x_j = \lambda x_i$ pour tout indice $j \in 1$. D'où le résultat demandé.

2. On suppose encore $\lambda \neq 1$. Soit $k \geqslant 1$ le nombre de coordonnées de X de module maximal. Parmi les $k+1$ complexes $x_i, \lambda x_i, \lambda^2 x_i, \ldots, \lambda^k x_i$, qui sont tous des composantes de X de module maximal, il y en a nécessairement deux qui sont égaux (principe des tiroirs). Il existe donc $r < s$ tels que $\lambda^r x_i = \lambda^s x_i$ et comme $x_i \neq 0$, on a $\lambda^m = 1$ avec $m \in [\![1, n]\!]$ en posant $m = s - r$.

3. La contraposée a été vue dans la question 1. ◁

L'énoncé suivant n'a rien à voir avec la réduction, mais comme il porte aussi sur le thème des matrices stochastiques nous avons choisi de le présenter ici.

2.11. Matrices stochastiques (3)

Soit S l'ensemble des matrices carrées d'ordre n réelles stochastiques.

1. Montrer que S est un convexe compact.

2. Déterminer le plus petit entier p tel qu'il existe un sous-espace affine de dimension p de $\mathcal{M}_n(\mathbb{R})$ contenant S.

(École polytechnique)

▷ **Solution.**

1. Rappelons qu'une matrice A est stochastique si ses coefficients sont tous positifs et si $AU = U$ où U désigne le vecteur dont toutes les composantes valent 1 (voir exercice 2.9). Si A et B sont dans S, il est aisé de voir que pour tout $t \in [0,1]$, $(1-t)A + tB \in S$. En effet, les coefficients sont positifs par convexité de l'intervalle $[0,1]$ et on a bien $((1-t)A + tB)U = (1-t)U + tU = U$. Donc S est convexe.

On choisit de munir $\mathcal{M}_n(\mathbb{R})$ de la norme définie par $\|A\|_1 = \max_{i,j} |a_{ij}|$ (toutes les normes sont équivalentes). Alors S est borné par 1 et fermé comme intersection d'hyperplans affines et de demi-espaces fermés. Donc S est compact.

2. Cette question demande de trouver la dimension du sous-espace affine engendré par S. Notons f_i la forme linéaire sur $\mathcal{M}_n(\mathbb{R})$ qui à une matrice associe la somme des coefficients de sa i-ième ligne. Par définition S est inclus dans le sous-espace affine $V = \bigcap_{i=1}^{n} f_i^{-1}(1)$. De manière plus précise S est l'intersection de V avec la partie de $\mathcal{M}_n(\mathbb{R})$ formée des matrices à coefficients positifs. Comme les formes linéaires $(f_1, \ldots, f_n)$ sont clairement libres, le sous-espace affine V est de dimension $n^2 - n$. On va prouver que S engendre affinement V.

Soit $A = (a_{ij})$ une matrice de V. Alors, A peut s'écrire comme bary-

centre de n matrices de V dont la dernière ligne est positive :

$$A = \sum_{j=1}^{n} a_{nj} B_j,$$

où B_j est obtenue en gardant les lignes $1, \dots, n-1$ de A et en remplaçant la dernière ligne par $(0, 0, \dots, 0, 1, 0, \dots, 0)$ avec le 1 en j-ième position (la somme des poids a_{nj} est égale à 1 puisque A est dans V).

De même, chacune des matrices B_j peut s'écrire comme barycentre de matrices de V dont les deux dernières lignes sont positives. Par associativité du barycentre, A est donc barycentre de matrices de V dont les deux dernières lignes sont positives. On itère cette idée pour obtenir finalement A comme barycentre de matrices stochastiques. ◁

En reprenant les calculs déjà menés dans l'exercice 2.10 on montre maintenant que pour une matrice strictement stochastique l'espace propre associé à la valeur propre 1 *est une droite.*

2.12. Matrices strictement stochastiques

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ une matrice strictement stochastique, c'est-à-dire stochastique avec des coefficients strictement positifs.

1. Montrer que $1 \in \operatorname{Sp} A$ et que $\dim \operatorname{Ker}(A - I_n) = 1$.

2. Montrer que toute valeur propre complexe de A est de module inférieur à 1 et que 1 est la seule valeur propre de module 1.

(École normale supérieure)

▷ **Solution.**

1. Le vecteur U dont toutes les coordonnées valent 1 est un vecteur propre pour la valeur propre 1. Soit $X = (x_1, \dots, x_n) \in \mathbb{C}^n$ un vecteur non nul tel que $AX = X$. On reprend la démonstration de la première question de l'exercice 2.10. On choisit un indice i tel que $|x_i|$ soit maximal. En regardant la i-ième coordonnée de l'égalité $AX = X$ il vient

$$1 - a_{ii} = \left| \sum_{j \neq i} a_{ij} \frac{x_j}{x_i} \right| \leqslant \sum_{j \neq i} a_{ij} \left| \frac{x_j}{x_i} \right| \leqslant \sum_{j \neq i} a_{ij} = 1 - a_{ii}.$$

Toutes les inégalités sont donc des égalités. Il en résulte d'une part que pour tout $j \in [\![1, n]\!]$, $|x_j| = |x_i|$ (car $a_{ij} \neq 0$) et d'autre part que tous les $a_{ij} \dfrac{x_j}{x_i}$ ont le même argument (cas d'égalité dans l'inégalité triangulaire). Comme a_{ij} est un réel strictement positif, les x_j sont donc tous égaux et X est colinéaire à U.

Le lecteur pourra montrer qu'en fait 1 *est une racine simple du polynôme caractéristique de* A *(ce qui est plus fort que le résultat prouvé ici).*

2. La première partie de la question a été résolue dans l'exercice 2.9 et la seconde est un cas particulier de la question 3 de l'exercice 2.10 puisque les coefficients diagonaux de A sont non nuls. ◁

Il y a en fait beaucoup de résultats concernant le spectre des matrices réelles dont les coefficients sont positifs. Le théorème de Perron-Frobenius, qui est prouvé dans l'exercice suivant, généralise les résultats vus en 2.9 et 2.10 pour les matrices stochastiques. Celui-ci affirme qu'une matrice réelle A *à coefficients positifs qui est irréductible (notion définie dans l'énoncé) admet une valeur propre réelle positive* ρ *telle que* $|\lambda| \leqslant \rho$ *pour toute valeur propre (complexe) de* A. *De plus, on peut trouver pour* ρ *un vecteur propre à coordonnées positives (vecteur de Perron). Ce résultat a été obtenu par Perron en 1907 pour les matrices à coefficients strictement positifs, puis généralisé par Frobenius par la suite. L'approche utilisée dans l'exercice est due à Wielandt. Le lecteur trouvera le cas particulier des matrices symétriques (plus facile) dans l'exercice 2.21 du tome algèbre 3.*

2.13. Théorème de Perron-Frobenius (1907)

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice à coefficients positifs. On note u l'endomorphisme de $\mathbb{R}^n$ canoniquement associé à A et $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$. On suppose que A est irréductible : cela signifie qu'il n'existe aucune permutation σ de $[\![1, n]\!]$ telle que la matrice de u dans la base $(e_{\sigma(1)}, \ldots, e_{\sigma(n)})$ soit triangulaire par blocs.

1. Montrer que $(I_n + A)^{n-1}$ est à coefficients strictement positifs. Si $X \in \mathbb{R}^n$ est un vecteur non nul à coordonnées positives, on pourra comparer le nombre de coordonnées strictement positives de X et de $Y = (I_n + A)X$.

2. Pour $X = (X_1, \ldots, X_n) \in \mathbb{R}_+^n$ non nul, on pose

$$r(X) = \min_{X_i \neq 0} \frac{(AX)_i}{X_i}.$$

où $(AX)_i$ désigne la i-ième coordonnée du vecteur AX. Montrer que r admet un maximum ρ sur $\mathbb{R}_+^n \setminus \{0\}$.

3. Montrer que $\rho \in \operatorname{Sp} A$ et que $|\lambda| \leqslant \rho$ pour tout $\lambda \in \operatorname{Sp} A$.

(École normale supérieure)

▷ **Solution.**

1. Soit $X = (X_1, \dots, X_n) \in \mathbb{R}_+^n$ et $Y = (I_n + A)X$ que l'on écrit $Y = (Y_1, \dots, Y_n)$. On a pour tout i, $Y_i = X_i + \sum_{j=1}^{n} a_{ij}X_j$ et il est donc clair que Y est aussi à coordonnées positives. Comme nous y invite l'énoncé, on va comparer le nombre de coordonnées strictement positives de X et de Y. Si $X = 0$ on a évidemment $Y = 0$. Si $X > 0$ (on note ainsi le fait que $X_i > 0$ pour tout i), alors $Y > 0$. Supposons donc que X ait exactement p coordonnées strictement positives avec $1 \leqslant p < n$. Il est clair que si $X_i > 0$, alors $Y_i > 0$. Donc Y a au moins p coordonnées strictement positives. On va prouver qu'il en a au moins une de plus. Notons $i_1, \dots, i_p$ les indices des p coordonnées de X strictement positives (les autres étant nulles). Supposons par l'absurde que l'on ait $Y_i = 0$ pour $i \notin \{i_1, \dots, i_p\}$. On a donc pour un tel indice i,

$$\sum_{j=1}^{n} a_{ij}X_j = \sum_{k=1}^{p} a_{ii_k}X_{i_k} = 0.$$

Comme les X_{i_k} sont tous non nuls, cela impose que $a_{ii_k} = 0$ pour tout $i \notin \{i_1, \dots, i_p\}$ et tout $k \in [\![1, p]\!]$. Cela contredit l'irréductibilité de la matrice A. En effet, si on effectue sur A une permutation consistant à mettre les colonnes et les lignes d'indice $i_1, \dots, i_p$ en premier, la matrice obtenue est triangulaire par blocs.

Il découle directement de ce résultat que si X est un vecteur non nul à coordonnées positives, alors $(A + I_n)^{n-1}X > 0$. C'est en particulier le cas en prenant pour X les vecteurs de la base canonique. Ainsi, toutes les colonnes de $(I_n + A)^{n-1}$ sont à coefficients strictement positifs.

2. Il est clair que pour tout $X = (X_1, \dots, X_n) \in \mathbb{R}_+^n \setminus \{0\}$ et tout $t > 0$ on a $r(tX) = r(X)$. Cette homogénéité permet de ramener la recherche du maximum de r sur l'ensemble

$$K = \left\{X = (X_1, \dots, X_n) \in \mathbb{R}_+^n,\ \sum_{i=1}^{n} X_i = 1\right\}$$

qui a le mérite d'être compact (il est clairement fermé et borné dans $\mathbb{R}^n$). Il reste à étudier la continuité de r. Les fonctions $X \mapsto \dfrac{(AX)_i}{X_i}$ sont toutes continues sur l'ensemble $P = (\mathbb{R}_+^*)^n$. Par ailleurs le minimum de deux fonctions continues (et donc, par récurrence, de n) est encore une fonction continue (sachant que $\min(a, b) = \dfrac{a+b}{2} - \dfrac{|b-a|}{2}$ pour tous réels a et b). Il en découle que r est continue sur P. Mais rien ne garantit que r reste continue en un point de la frontière de P, c'est-à-dire en un vecteur positif X qui a au moins une coordonnée nulle. Et d'ailleurs c'est faux en général ! Prenons par exemple la matrice $A = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$ et pour

tout $\varepsilon > 0$, $X_\varepsilon = (1, \varepsilon)$. On a

$$r(X_\varepsilon) = \min\left(\frac{2+\varepsilon}{1}, \frac{\varepsilon}{\varepsilon}\right) = 1$$

pour tout $\varepsilon > 0$. Ainsi $r(1,0) = 2$ n'est donc pas égal à la limite de $r(X_\varepsilon)$ lorsque $\varepsilon \to 0^+$. Certes A n'est pas irréductible, mais ce n'est pas la cause du problème : le lecteur fabriquera sans mal un exemple de matrice irréductible pour laquelle r n'est pas continue en tout point de K.

On va utiliser la première question pour contourner ce problème. Posons $K' = (I_n + A)^{n-1}(K)$. Comme image continue d'un compact, K' est encore compact, et il est inclus dans P d'après la première question. Donc r est continue sur K' et par suite atteint un maximum ρ. Il ne reste plus qu'à prouver que ρ est le maximum de r sur K. Soit $X \in K$ et $Y = (I_n + A)^{n-1}X \in K'$. On va montrer que $r(X) \leqslant r(Y)$. Par définition de $r(X)$ on a $(AX)_i \geqslant r(X)X_i$ pour tout $i \in [\![1, n]\!]$ (même si $X_i = 0$). Autrement dit, le vecteur $AX - r(X)X$ est positif. Son image par $(I_n + A)^{n-1}$ est donc à coordonnées positives. Or, cette image vaut $AY - r(X)Y$ car A et $(I_n + A)^{n-1}$ commutent. On a donc $(AY)_i \geqslant r(X)Y_i$ pour tout i et par suite $r(Y) = \min_{1 \leqslant i \leqslant n} \dfrac{(AY)_i}{Y_i} \geqslant r(X)$. Il en découle que r est majorée par ρ sur K. Mais cette valeur est atteinte dans K : en effet, il existe $Y \in K'$ tel que $r(Y) = \rho$ et il suffit de prendre l'unique vecteur de K colinéaire à Y.

Conclusion. La fonction r admet ρ pour maximum sur K et donc sur $\mathbb{R}_+^n \setminus \{0\}$.

3. Par définition on a $\rho \geqslant 0$. En fait ρ n'est pas nul, car il est clair que $r(1, 1, \ldots, 1) > 0$. Reprenons $Y \in K'$ tel que $r(Y) = \rho$. On va prouver qu'en fait $AY = \rho Y$. On a $AY - \rho Y \geqslant 0$. Supposons que ce vecteur ne soit pas nul. D'après la question 1 son image par $(I_n + A)^{n-1}$ est alors un vecteur strictement positif. Posons $Z = (I_n + A)^{n-1}Y$. On a alors $AZ - \rho Z > 0$. Pour $\varepsilon > 0$ assez petit le vecteur $AZ - \rho Z - \varepsilon Z$ reste strictement positif. On a alors $r(Z) \geqslant \rho + \varepsilon$, ce qui contredit la définition de ρ.

Montrons enfin que ρ majore le module de toute valeur propre complexe de A. Soit $\lambda \in \operatorname{Sp} A$ et $X \in \mathbb{C}^n$ un vecteur propre associé. On a pour tout i, $\lambda X_i = \sum_{j=1}^{n} a_{ij}X_j$. En prenant le module il vient, par inégalité triangulaire,

$$|\lambda||X_i| \leqslant \sum_{i=1}^{n} a_{ij}|X_j|$$

pour tout i. Autrement dit, si on note $\tilde{X}$ le vecteur $(|X_1|, \ldots, |X_n|)$, on a $(A\tilde{X})_i \geqslant |\lambda|\tilde{X}_i$ pour tout i et donc $|\lambda| \leqslant r(\tilde{X})$. Mais, par définition, $r(\tilde{X}) \leqslant \rho$. D'où le résultat.

Conclusion. ρ est une valeur propre de A et de plus $|\lambda| \leqslant \rho$ pour tout $\lambda \in \operatorname{Sp} A$. On notera par ailleurs que A admet un vecteur propre à coordonnées strictement positives pour ρ. ◁

Le lecteur pourra en fait prouver que l'espace propre associé à ρ est de dimension 1 et même, ce qui est un peu plus difficile, que ρ est une racine simple du polynôme caractéristique de A. *Indiquons que la seconde épreuve du concours de l'ENSAE en 1994 a porté sur ce sujet ainsi que le début de l'épreuve d'Agrégation Externe de 1995. Une très bonne référence pour toutes les questions relatives aux matrices positives est le livre de Henry Minc,* Nonnegative Matrices, *paru aux éditions Wiley.*

Avant d'aborder les exercices concernant la diagonalisation, nous regroupons ci-après des énoncés portant sur les polynômes d'endomorphismes. Rappelons que si u est un endomorphisme d'un K*-espace vectoriel* E *l'application ψ qui à $P \in K[X]$ associe $P(u)$ définit un morphisme d'algèbre de $\mathbb{K}[X]$ dans $\mathcal{L}(E)$. Son image $K[u]$ est donc une sous-algèbre commutative de $\mathcal{L}(E)$. Lorsque* E *est de dimension finie le morphisme ψ ne peut pas être injectif. Son noyau est un idéal non nul de $K[X]$ dont le générateur unitaire est appelé polynôme minimal de u. En revanche, lorsque* E *est de dimension infinie,* $\operatorname{Ker}\psi$ *peut très bien être réduit à $\{0\}$ et l'existence d'un polynôme annulateur non nul n'est plus assurée.*

2.14. Polynôme annulateur

Soit E un espace vectoriel de dimension infinie, u et v deux endomorphismes de E qui commutent. On suppose que u (resp. v) admet un polynôme annulateur non nul P (resp. Q). Montrer qu'il en est de même de $u + v$.

(École polytechnique)

▷ **Solution.**

Notons que l'existence d'un polynôme annulateur non nul de u équivaut au caractère lié de la famille $(u^k)_{k\in\mathbb{N}}$. Notons n le plus grand entier tel que $(\mathrm{Id}, u, \ldots, u^{n-1})$ soit libre. La famille $(\mathrm{Id}, u, \ldots, u^{n-1}, u^n)$ est alors liée et on a donc $u^n \in \operatorname{Vect}(\mathrm{Id}, u, \ldots, u^{n-1})$. Il est alors facile de prouver par récurrence sur k que $u^k \in \operatorname{Vect}(\mathrm{Id}, u, \ldots u^{n-1})$ pour tout $k \geqslant n$. Autrement dit, on a

$$\operatorname{Vect}(\mathrm{Id}, u, \ldots, u^{n-1}) = \operatorname{Vect}(u^k)_{k\geqslant 0}.$$

En fait, n n'est rien d'autre que le degré du polynôme minimal de u. Notons de même m le degré du polynôme minimal de v. Il est alors clair que

$$\mathrm{F} = \operatorname{Vect}(u^k v^\ell)_{(k,\ell)\in\mathbb{N}^2} = \operatorname{Vect}(u^i v^j)_{\substack{0\leqslant i\leqslant n-1\\ 0\leqslant j\leqslant m-1}}.$$

En particulier, il s'agit d'un espace de dimension finie. Comme u et v commutent, on a $(u+v)^k = \sum_{i=0}^{k} C_k^i u^i v^{k-i}$ pour tout k. Le sous-espace $\mathrm{Vect}((u+v)^k)_{k\geqslant 0}$ est donc contenu dans F et par suite il est de dimension finie. D'après la remarque initiale, cela prouve que $u+v$ admet un polynôme annulateur non nul. ◁

2.15. Noyaux et images de polynômes d'endomorphismes

Soient A et B dans K[X]. f un endomorphisme d'un K-espace vectoriel. On pose $D = \mathrm{pgcd}(A,B)$ et $M = \mathrm{ppcm}(A,B)$. Exprimer $\mathrm{Ker}\,D(f)$, $\mathrm{Im}\,D(f)$, $\mathrm{Ker}\,M(f)$, $\mathrm{Im}\,M(f)$ à l'aide de $\mathrm{Ker}\,A(f)$, $\mathrm{Im}\,A(f)$, $\mathrm{Ker}\,B(f)$ et $\mathrm{Im}\,B(f)$.

(École polytechnique)

▷ **Solution.**

On note A′ et B′ les quotients de A et B par D (sans risque de confusion avec la dérivée). On a donc $A = DA'$, $B = DB'$ et $M = DA'B'$. Les polynômes A′ et B′ sont premiers entre eux et nous allons utiliser une relation de Bezout $A'P + B'Q = 1$ (et donc $D = AP + BQ$). Nous utiliserons enfin le résultat évident suivant : si u, v sont deux endomorphismes, on a $\mathrm{Ker}\,u \subset \mathrm{Ker}(v\circ u)$ et $\mathrm{Im}(v\circ u) \subset \mathrm{Im}\,v$. Commençons par chercher le noyau et l'image de $D(f)$.

• Comme $A(f) = A'(f)\circ D(f)$, on a $\mathrm{Ker}\,D(f) \subset \mathrm{Ker}\,A(f)$. Il en est de même avec B et $\mathrm{Ker}\,D(f)$ est donc inclus dans $\mathrm{Ker}\,A(f) \cap \mathrm{Ker}\,B(f)$. On va montrer qu'il y a en fait égalité. Pour cela, on utilise la relation de Bezout : $D(f) = P(f)\circ A(f) + Q(f)\circ B(f)$. Il en découle directement que si $A(f)(x) = B(f)(x) = 0$, alors $D(f)(x) = 0$. On conclut que

$$\boxed{\mathrm{Ker}\,D(f) = \mathrm{Ker}\,A(f) \cap \mathrm{Ker}\,B(f)}.$$

• Considérons $\mathrm{Im}\,D(f)$. De l'égalité $A(f) = D(f)\circ A'(f)$, on déduit que $\mathrm{Im}\,A(f) \subset \mathrm{Im}\,D(f)$. De même, $\mathrm{Im}\,B(f) \subset \mathrm{Im}\,D(f)$ et on a donc l'inclusion $\mathrm{Im}\,A(f) + \mathrm{Im}\,B(f) \subset \mathrm{Im}\,D(f)$. Inversement, de la relation $D(f) = A(f)\circ P(f) + B(f)\circ Q(f)$, on déduit $\mathrm{Im}\,D(f) \subset \mathrm{Im}\,A(f) + \mathrm{Im}\,B(f)$. On conclut que

$$\boxed{\mathrm{Im}\,D(f) = \mathrm{Im}\,A(f) + \mathrm{Im}\,B(f)}.$$

• Regardons $\mathrm{Ker}\,M(f)$. On a $M = AB'$ et donc $M(f) = B'(f)\circ A(f)$. Ainsi $\mathrm{Ker}\,A(f) \subset \mathrm{Ker}\,M(f)$. De même, $\mathrm{Ker}\,B(f) \subset \mathrm{Ker}\,M(f)$ et par conséquent la somme $\mathrm{Ker}\,A(f) + \mathrm{Ker}\,B(f)$ est contenue dans $\mathrm{Ker}\,M(f)$.

Inversement, soit $x \in \mathrm{Ker}\,M(f)$. On a $\mathrm{Id} = A'(f)\circ P(f) + B'(f)\circ Q(f)$ et en appliquant en x,

$$x = (\mathrm{A}'(f) \circ \mathrm{P}(f))(x) + (\mathrm{B}'(f) \circ \mathrm{Q}(f))(x)$$

Or le premier vecteur est dans $\operatorname{Ker} \mathrm{B}(f)$, car $\mathrm{BA}'\mathrm{P} = \mathrm{MP}$ est un multiple de M, et pour des raisons analogues le second est dans $\operatorname{Ker} \mathrm{A}(f)$. Il s'ensuit que x est dans $\operatorname{Ker} \mathrm{A}(f) + \operatorname{Ker} \mathrm{B}(f)$. On conclut que

$$\boxed{\operatorname{Ker} \mathrm{M}(f) = \operatorname{Ker} \mathrm{A}(f) + \operatorname{Ker} \mathrm{B}(f)}.$$

• Considérons enfin $\operatorname{Im} \mathrm{M}(f)$. De $\mathrm{M}(f) = \mathrm{A}(f) \circ \mathrm{B}'(f)$, on déduit que $\operatorname{Im} \mathrm{M}(f) \subset \operatorname{Im} \mathrm{A}(f)$. Par symétrie, $\operatorname{Im} \mathrm{M}(f) \subset \operatorname{Im} \mathrm{B}(f)$ et finalement $\operatorname{Im} \mathrm{M}(f) \subset \operatorname{Im} \mathrm{A}(f) \cap \operatorname{Im} \mathrm{B}(f)$. Ici encore il y a égalité. En effet, on écrit toujours

$$x = (\mathrm{A}'(f) \circ \mathrm{P}(f))(x) + (\mathrm{B}'(f) \circ \mathrm{Q}(f))(x).$$

Si $x \in \operatorname{Im} \mathrm{B}(f)$ on peut l'écrire $x = \mathrm{B}(f)(y)$ et dans ce cas

$$(\mathrm{A}'(f) \circ \mathrm{P}(f))(x) = (\mathrm{A}'(f) \circ \mathrm{P}(f) \circ \mathrm{B}(f))(y) = (\mathrm{M}(f) \circ \mathrm{P}(f))(y) \in \operatorname{Im} \mathrm{M}(f).$$

De même, si $x \in \operatorname{Im} \mathrm{A}(f)$ le vecteur $(\mathrm{B}'(f) \circ \mathrm{Q}(f))(x)$ est dans $\operatorname{Im} \mathrm{M}(f)$. Donc, si $x \in \operatorname{Im} \mathrm{A}(f) \cap \operatorname{Im} \mathrm{B}(f)$ il est dans $\operatorname{Im} \mathrm{M}(f)$. On conclut donc que

$$\boxed{\operatorname{Im} \mathrm{M}(f) = \operatorname{Im} \mathrm{A}(f) \cap \operatorname{Im} \mathrm{B}(f)}. \quad \lhd$$

Nous en venons maintenant aux exercices concernant la diagonalisation.

2.16. Diagonalisabilité d'une matrice

Soit K un sous-corps de $\mathbb{C}$ et $(a_1, \ldots, a_n) \in \mathrm{K}^n$. On considère la matrice

$$\mathrm{A} = \begin{pmatrix} 0 & 0 & \ldots & 0 & a_1 \\ 0 & \ddots & & 0 & a_2 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 0 & a_{n-1} \\ a_1 & a_2 & \ldots & a_{n-1} & a_n \end{pmatrix}.$$

À quelle condition A est-elle diagonalisable ?

(École polytechnique)

▷ **Solution.**

Si $a_1 = \cdots = a_{n-1} = 0$, la matrice A est diagonale. Nous supposerons désormais l'un des a_i, pour $i \in [\![1, n-1]\!]$, non nul. Alors A est de rang 2 et donc $\dim \operatorname{Ker} \mathrm{A} = n - 2$. Regardons les valeurs propres non nulles :

soit $\lambda \in \mathrm{K}^*$, $\mathrm{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ et le système (S) $\mathrm{AX} = \lambda \mathrm{X}$. Explicitons ce système :

$$(\mathrm{S}) \iff (\forall i \in [\![1, n-1]\!],\ a_i x_n = \lambda x_i) \text{ et } (a_1 x_1 + \cdots + a_n x_n = \lambda x_n\ (*)).$$

Comme λ est non nul, on peut multiplier $(*)$ par λ qui devient

$$a_1^2 x_n + \cdots + a_{n-1}^2 x_n + \lambda a_n x_n = \lambda^2 x_n$$

ou encore

$$(\lambda^2 - a_n \lambda - (a_1^2 + \cdots + a_{n-1}^2)) x_n = 0.$$

Le système n'est pas de Cramer si, et seulement si

$$\lambda^2 - a_n \lambda - (a_1^2 + \cdots + a_{n-1}^2) = 0$$

et dans ces conditions, le sous-espace propre associé à λ est de dimension 1 (le rang du système étant $n-1$). Pour que A soit diagonalisable, il faut et il suffit que la somme directe des sous-espaces propres soit de dimension n, autrement dit dans notre cas, il faut et il suffit que le polynôme $\mathrm{X}^2 - a_n \mathrm{X} - (a_1^2 + \cdots + a_{n-1}^2)$ possède deux racines distinctes non nulles, ce qui se traduit par les conditions

$$a_1^2 + \cdots + a_{n-1}^2 \neq 0 \quad \text{et} \quad a_n^2 + 4(a_1^2 + \cdots + a_{n-1}^2) \neq 0. \lhd$$

On remarquera que si $\mathrm{K} = \mathbb{R}$, *ces conditions sont toujours réalisées, ce à quoi on pouvait s'attendre, puisqu'alors* A *est symétrique réelle, donc diagonalisable.*

On s'intéresse maintenant au caractère diagonalisable des matrices tridiagonales qui jouent un rôle dans la résolution numérique par discrétisation des équations différentielles d'ordre 2.

2.17. Diagonalisabilité d'une matrice tridiagonale

Soit $(a, b) \in \mathbb{C}^2$. Déterminer les vecteurs propres et les valeurs propres de $\mathrm{M} = \begin{pmatrix} a & b & 0 & \dots & 0 \\ b & a & b & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & b & a & b \\ 0 & \dots & 0 & b & a \end{pmatrix}$.

(École polytechnique)

▷ **Solution.**

Comme $M = aI_n + b\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & & 0 \\ & \ddots & \ddots & \ddots & \\ 0 & & 1 & 0 & 1 \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}$, il suffit de traiter le cas $a = 0$ et $b = 1$. Supposons donc $a = 0$ et $b = 1$. Soit $\lambda \in \mathbb{C}$ une valeur propre de M et $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ un vecteur propre associé à λ. On a alors

$$(*) \quad \begin{cases} x_2 = \lambda x_1 \\ x_{k-1} + x_{k+1} = \lambda x_k \text{ pour } k \in [\![2, n-1]\!] \\ x_{n-1} = \lambda x_n. \end{cases}$$

En posant $x_0 = x_{n+1} = 0$, on écrit de manière plus élégante que pour tout $k \in [\![1, n]\!]$

$$x_{k+1} - \lambda x_k + x_{k-1} = 0.$$

On reconnaît là les termes d'une suite à récurrence linéaire d'ordre 2. Le polynôme caractéristique associé est $X^2 - \lambda X + 1$. Appelons α et β les racines distinctes ou confondues de ce polynôme. Supposons $\alpha \neq \beta$. Il existe alors $A, B \in \mathbb{C}$ tels que $x_k = A\alpha^k + B\beta^k$ pour tout $k \in [\![0, n+1]\!]$. Comme $x_0 = 0$ on a $B = -A$. Comme $X \neq 0$ on a $A \neq 0$ et l'égalité $x_{n+1} = 0$ implique $\alpha^{n+1} - \beta^{n+1} = 0$. Or le produit $\alpha\beta$ vaut 1 si bien que $\alpha^{2n+2} = \beta^{2n+2} = 1$. On en déduit qu'il existe $p \in [\![0, 2n+1]\!]$ tel que $\alpha = e^{\frac{2ip\pi}{2n+2}} = e^{\frac{ip\pi}{n+1}}$. On obtient donc

$$x_k = A\left(e^{\frac{ipk\pi}{n+1}} - e^{-\frac{ipk\pi}{n+1}}\right) = 2iA\sin\frac{pk\pi}{n+1}.$$

D'autre part $\lambda = \alpha + \beta = 2\cos\dfrac{p\pi}{n+1}$. On remarque que $2\cos\dfrac{p\pi}{n+1}$ prend exactement $n+1$ valeurs distinctes (celles qui correspondent à p et $2n+2-p$ sont égales). Le cas $p = 0$ est à exclure car alors $\lambda = 0$, $\alpha = \beta = 1$ et $x_k = 0$ pour tout k ce qui est impossible puisque X est un vecteur propre.

Passons à la synthèse : soit $p \in [\![1, n]\!]$. On pose $\lambda_p = 2\cos\dfrac{p\pi}{n+1}$. Alors la suite définie par $x_k = \sin\dfrac{pk\pi}{n+1}$ pour $k \in [\![0, n+1]\!]$ vérifie la relation de récurrence linéaire

$$x_{k+1} - 2\lambda_p x_k + x_{k-1} = 0,$$

car $e^{\frac{ip\pi}{n+1}}$ et $e^{-\frac{ip\pi}{n+1}}$ sont les racines de $X^2 - \lambda_p X + 1$. Comme $x_0 = x_{n+1} =$

0, le système $(*)$ est vérifié et le vecteur $e_p = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ est donc un vecteur propre de M pour la valeur propre λ_p. Comme nous disposons de n valeurs propres distinctes, M est diagonalisable dans la base $(e_1, \ldots, e_n)$.

Dans le cas où a et b sont quelconques, les vecteurs $(e_1, \ldots, e_n)$ constituent encore une base de diagonalisation de M et la valeur propre associée à e_k est $\lambda_p = a + 2b\cos\dfrac{pk\pi}{n+1}$. $\lhd$

Si a et b sont réels, la matrice M *est redevable du théorème spectral et on peut dire directement que* M *est diagonalisable. En particulier, les vecteurs e_k sont orthogonaux deux à deux pour la structure euclidienne canonique de $\mathbb{R}^n$.*

2.18. Racine d'une matrice diagonalisable inversible

Soit $A \in GL_n(\mathbb{C})$. On suppose qu'il existe $p \in \mathbb{N}^*$ tel que A^p soit diagonalisable. Montrer que A est diagonalisable.

La propriété subsiste-t-elle si A n'est pas inversible ?

(École polytechnique)

$\rhd$ **Solution.**

Il existe $\lambda_1, \ldots, \lambda_r$ dans $\mathbb{C}$ deux à deux distincts tels que le polynôme $(X-\lambda_1)\ldots(X-\lambda_r)$ annule A^p. Par conséquent, $(X^p-\lambda_1)\ldots(X^p-\lambda_r)$ annule A. Ce polynôme est scindé à racines simples, car pour tout $\lambda \in \mathbb{C}^*$, $X^p - \lambda$ est scindé à racines simples (dans $\mathbb{C}^n$, tout nombre complexe non nul admet exactement p racines p-ièmes) et si $\lambda \neq \mu$, $X^p - \lambda$ et $X^p - \mu$ n'ont aucune racine en commun. Il s'ensuit que A est diagonalisable.

La propriété ne subsiste pas si A n'est pas inversible : par exemple, si A est nilpotente non nulle, $A^n = 0$ est diagonalisable sans que A le soit. $\lhd$

Voici une étude plus détaillée du cas $p = 2$, sur $\mathbb{C}$ et sur $\mathbb{R}$.

2.19. Diagonalisabilité de f dans le cas f^2 diagonalisable

1. Soit $f \in \mathcal{L}(\mathbb{C}^n)$. Montrer que f est diagonalisable si, et seulement si, f^2 est diagonalisable et $\operatorname{Ker} f = \operatorname{Ker} f^2$.

2. Soit $f \in \mathcal{L}(\mathbb{R}^n)$. On suppose f^2 diagonalisable. À quelle condition nécessaire et suffisante f est-elle diagonalisable ?

3. À quelle condition nécessaire et suffisante la matrice complexe (resp. réelle)

$$A = \begin{pmatrix} 0 & 0 & \dots & a_n \\ 0 & 0 & \dots & 0 \\ 0 & a_2 & & 0 \\ a_1 & 0 & \dots & 0 \end{pmatrix}$$

est-elle diagonalisable ?

(École polytechnique)

▷ **Solution.**

1. Si f est diagonalisable, il existe une base $\mathcal{B}$ de $\mathbb{C}^n$ telle que la matrice de f dans cette base soit $\mathrm{Diag}(\lambda_1, \dots, \lambda_r, 0, \dots, 0)$ avec λ_i valeur propre non nulle de f. La matrice de f^2 dans cette même base est $\mathrm{Diag}(\lambda_1^2, \dots, \lambda_r^2, 0 \dots, 0)$: l'endomorphisme f^2 est donc diagonalisable et $\operatorname{rg} f^2 = \operatorname{rg} f = r$. Comme on a $\operatorname{Ker} f \subset \operatorname{Ker} f^2$ il en découle que $\operatorname{Ker} f^2 = \operatorname{Ker} f$.

Réciproquement, supposons f^2 diagonalisable et $\operatorname{Ker} f = \operatorname{Ker} f^2$. Notons $\mu_1, \dots, \mu_r$ les valeurs propres non nulles de f^2. Pour $1 \leqslant i \leqslant p$, on prend $\lambda_i \in \mathbb{C}^*$ telle que $\lambda_i^2 = \mu_i$. Comme $X - \lambda_i$ et $X + \lambda_i$ sont premiers entre eux, le théorème de décomposition des noyaux donne $\operatorname{Ker}(f^2 - \mu_i \operatorname{Id}) = \operatorname{Ker}(f - \lambda_i \operatorname{Id}) \oplus \operatorname{Ker}(f + \lambda_i \operatorname{Id})$. Dans ces conditions, on peut écrire

$$\begin{aligned} \mathbb{C}^n &= \operatorname{Ker} f^2 \oplus \left(\bigoplus_{i=1}^{r} \operatorname{Ker}(f^2 - \mu_i \operatorname{Id}) \right) \\ &= \operatorname{Ker} f \oplus \left(\bigoplus_{i=1}^{r} (\operatorname{Ker}(f - \lambda_i \operatorname{Id}) \oplus \operatorname{Ker}(f + \lambda_i \operatorname{Id})) \right), \end{aligned}$$

si bien que $\mathbb{C}^n$ est somme de sous-espaces propres de f, *i.e.* f est diagonalisable.

2. Si f est diagonalisable, en écrivant la matrice de f dans une base de vecteurs propres, il apparaît que $\operatorname{Ker} f = \operatorname{Ker} f^2$ et que les valeurs propres de f^2 sont positives ou nulles.

Réciproquement, si $\operatorname{Ker} f = \operatorname{Ker} f^2$ et $\operatorname{Sp} f^2 \subset \mathbb{R}_+$, ce qui a été fait dans la première question peut être repris (chaque valeur propre μ non nulle de f^2 possède bien une racine carrée λ dans $\mathbb{R}$), et on démontre de manière analogue que f est diagonalisable.

La condition nécessaire et suffisante cherchée est donc $\operatorname{Ker} f = \operatorname{Ker} f^2$ et $\operatorname{Sp} f^2 \subset \mathbb{R}_+$.

3. Soit $(e_1, \dots, e_n)$ la base canonique de $\mathbb{C}^n$ (resp. $\mathbb{R}^n$). On a pour tout $1 \leqslant i \leqslant n$, $Ae_i = a_i e_{n-i+1}$ et donc $A^2 e_i = a_i a_{n-i+1} e_i$. Il s'ensuit que A^2 est diagonalisable, puisque diagonale. Pour que A le soit, il faut

et il suffit que $\operatorname{Ker} A = \operatorname{Ker} A^2$ si $K = \mathbb{C}$ (resp. $\operatorname{Ker} A = \operatorname{Ker} A^2$ et les produits $a_i a_{n-i+1}$ positifs ou nuls si $K = \mathbb{R}$). Traduisons la condition $\operatorname{Ker} A = \operatorname{Ker} A^2$. Le sous-espace $\operatorname{Ker} A$ est engendré par les e_i tels que $a_i = 0$ et le sous-espace $\operatorname{Ker} A^2$ par les e_i tels que $a_i a_{n-i+1} = 0$. Pour que $\operatorname{Ker} A = \operatorname{Ker} A^2$, il faut et il suffit que pour tout i,

$$a_i a_{n-i+1} = 0 \implies a_i = 0,$$

ce qui est encore équivalent à, pour tout i, $a_i = 0 \iff a_{n-i+1} = 0$.

Conclusion. La matrice complexe (resp. réelle) A est diagonalisable si et seulement si pour tout $1 \leqslant i \leqslant n$, $a_i = 0 \iff a_{n-i+1} = 0$ (resp. a_i et a_{n-i+1} sont du même signe strict pour tout i). ◁

On peut par exemple utiliser l'exercice précédent pour résoudre le suivant, mais ce n'est pas obligatoire.

2.20. Étude de diagonalisabilité

Soit X et Y dans $\mathcal{M}_{n,1}(\mathbb{R})$ et A la matrice par blocs $\begin{pmatrix} 0 & X \\ {}^tY & 0 \end{pmatrix}$ de $\mathcal{M}_n(\mathbb{R})$. À quelle condition A est-elle diagonalisable ?

(**École polytechnique**)

▷ **Solution.**

Si $X = Y = 0$, $A = 0$ est diagonalisable. Si $X = 0$ et $Y \neq 0$ (ou l'inverse), A est triangulaire à diagonale nulle. Si A était diagonalisable, elle serait semblable à 0 et serait en fait nulle ce qui n'est pas le cas. Donc A n'est pas diagonalisable.

Plaçons-nous dans le cas où X et Y sont non nuls. La matrice A est de rang 2. En élevant au carré, on trouve $A^2 = \begin{pmatrix} X\,{}^tY & 0 \\ 0 & {}^tYX \end{pmatrix}$. On va utiliser le résultat de l'exercice précédent : la matrice A est diagonalisable si, et seulement si, A^2 est diagonalisable, $\operatorname{Ker} A = \operatorname{Ker} A^2$ et $\operatorname{Sp} A^2 \subset \mathbb{R}_+$.

Supposons A diagonalisable. Alors le scalaire ${}^tYX$ est valeur propre de A^2 et la matrice $X\,{}^tY$ de taille $n-1$ est de rang 1. Comme $\operatorname{rg} A = \operatorname{rg} A^2$, on a ${}^tYX \neq 0$ et même ${}^tYX > 0$.

Réciproquement, si ${}^tYX > 0$, alors $\operatorname{rg} A^2 = 2 = \operatorname{rg} A$ car la matrice carrée de taille $n-1$, $X\,{}^tY$, est de rang 1. Rappelons ce résultat classique :

Lemme. *Une matrice de $\mathcal{M}_m(K)$ de rang 1 est diagonalisable si et seulement si sa trace est non nulle.*

Démonstration. Soit M une matrice carré de taille m et de rang 1. Comme $\dim \operatorname{Ker} M = m-1$, 0 est racine de χ_M avec une multiplicité égale à $m-1$ ou m. Si $\operatorname{Tr} M = 0$, la dernière racine de χ_M est 0 et $\chi_M = X^m$ si bien

que M est nilpotente non nulle donc non diagonalisable. Par contre, si $\operatorname{Tr} M \neq 0$, $\lambda = \operatorname{Tr} M$ est racine de χ_M et la somme des sous-espaces propres remplit l'espace : M est diagonalisable. $\diamondsuit$

En écrivant explicitement $X\,{}^tY$, il apparaît que $\operatorname{Tr} X\,{}^tY = {}^tYX > 0$. Donc le bloc $X\,{}^tY$ est diagonalisable et possède une unique valeur propre non nulle ${}^tYX$ qui est strictement positive. On en déduit que A^2 est diagonalisable et que son spectre est dans $\mathbb{R}_+$. Comme $\operatorname{rg} A^2 = \operatorname{rg} A$, A est elle-même diagonalisable.

Conclusion. A est diagonalisable si, et seulement si, $X = Y = 0$ ou ${}^tYX > 0$. $\lrcorner$

On aurait pu également procéder directement comme à l'exercice 2.16.

L'algèbre des matrices diagonales est commutative. Pour qu'on puisse diagonaliser simultanément plusieurs endomorphismes il est donc nécessaire que ceux-ci commutent. En fait, cela est suffisant et c'est le résultat que propose l'exercice suivant.

2.21. Diagonalisation simultanée

Soit E un espace vectoriel de dimension finie.

1. Montrer que l'endomorphisme induit par un endomorphisme diagonalisable dans un sous-espace stable est diagonalisable.

2. Soit $A \subset \mathcal{L}(E)$ formée d'endomorphismes diagonalisables commutant deux à deux. Montrer que les éléments de A admettent une base commune de diagonalisation.

(École polytechnique)

▷ **Solution.**

1. Soit u un endomorphisme diagonalisable du K-espace vectoriel E. Il existe P polynôme de K[X], scindé à racines simples annulant u. Supposons que u laisse stable un sous-espace F. Alors P annule aussi $u_{|F}$ et $u_{|F}$ est diagonalisable.

2. La partie A engendre un sous-espace de $\mathcal{L}(E)$ qui est de dimension finie. Il existe alors une famille $(u_1, \ldots, u_p)$ de A constituant une base de ce sous-espace. Si on trouve une base diagonalisant tous les éléments de cette famille, tout endomorphisme de A s'exprimant comme combinaison linéaire des u_i, il sera diagonalisable dans cette base.

Montrons l'existence d'une base commune de diagonalisation des u_i par une récurrence sur $p \geqslant 1$. C'est trivial si $p = 1$. Supposons $p \geqslant 2$ et le résultat vrai au rang $p-1$. Comme u_p est diagonalisable, on peut écrire, en notant $\lambda_1, \ldots, \lambda_k$ les valeurs propres deux à deux distinctes de u_p,

$$\mathrm{E} = \bigoplus_{i=1}^{k} \mathrm{Ker}(u_p - \lambda_i \mathrm{Id}).$$

Comme chaque u_j, pour $1 \leqslant j \leqslant p-1$, commute avec u_p, u_j laisse stable chaque sous-espace propre $\mathrm{Ker}(u_p - \lambda_i \mathrm{Id})$. Notons u_{ji} l'endomorphisme induit par u_j sur $\mathrm{Ker}(u_p - \lambda_i \mathrm{Id})$. À i fixé dans $\{1, 2, \ldots, k\}$, les u_{ji} en tant que restrictions d'endomorphismes diagonalisables, sont eux-mêmes diagonalisables et commutent deux à deux. Il y en a $p-1$ et ils sont donc justifiables de l'hypothèse de récurrence : il existe $\mathcal{B}_i$ base de $\mathrm{Ker}(u_p - \lambda_i \mathrm{Id})$ qui diagonalise chaque u_{ji}. La base $\mathcal{B}_i$ est donc constituée de vecteurs propres pour u_j, $1 \leqslant j \leqslant p-1$, mais aussi pour u_p puisque les vecteurs de $\mathcal{B}_i$ sont dans $\mathrm{Ker}(u_p - \lambda_i \mathrm{Id})$. Comme E est somme directe des sous-espaces propres de u_p, $(\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_k)$ est une base de E constituée de vecteurs propres pour $u_1, \ldots, u_p$: c'est une base commune de diagonalisation. ◁

Il est aussi possible de faire une récurrence (forte) sur la dimension de l'espace E. *Si tous les éléments de* A *sont des homothéties, le résultat est trivial. Sinon, on choisit un élément* u *de* A *qui n'est pas une homothétie. Les sous-espaces propres de* u *sont stables par tous les éléments de* A *et de dimension strictement inférieure à celle de* E. *Les endomorphismes induits sur ces sous-espaces commutent deux à deux ce qui permet d'appliquer l'hypothèse de récurrence. Le lecteur pourra aussi montrer par une récurrence du même type que, sur un* $\mathbb{C}$*-espace vectoriel de dimension finie* E, *toute famille commutative d'endomorphismes de* E *admet un vecteur propre commun.*

Voici un exemple où l'on codiagonalise, mais élémentairement, une sous-algèbre commutative de $\mathcal{M}_n(\mathbb{C})$.

2.22. Matrices circulantes (1)

Soit $\mathcal{A}$ l'ensemble des matrices de la forme

$$\begin{pmatrix} a_0 & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & a_2 \\ \vdots & \vdots & \cdots & \vdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{pmatrix}$$

de $\mathcal{M}_n(\mathbb{C})$ (on les appelle des *matrices circulantes*).

1. Montrer que $\mathcal{A}$ est une sous-algèbre commutative de $\mathcal{M}_n(\mathbb{C})$.

2. Montrer que les éléments de $\mathcal{A}$ sont simultanément diagonalisables.

3. On remplace $\mathbb{C}$ par un corps K dans lequel $X^n - 1$ est scindé. Le résultat de la question 2 subsiste-t-il ? Discuter selon la caractéristique de K.

(École normale supérieure)

▷ **Solution.**

1. Chaque ligne d'une matrice $A \in \mathcal{A}$ s'obtient par permutation circulaire à partir de la précédente (ce qui explique le nom de matrices circulantes qu'on leur donne).

Soit $J = \begin{pmatrix} 0 & \cdots & \cdots & 0 & 1 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & \ddots & \ddots & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$, c'est-à-dire $J = (m_{i,j})$ où $m_{i,j} = 1$ si $i \equiv j+1 \mod n$ et $m_{i,j} = 0$ sinon. C'est la matrice de permutation associée au n-cycle $(1, 2, \ldots, n)$. Cette interprétation de J rend facile le calcul de ses puissances : J agit sur les vecteurs de la base canonique $(e_1, \ldots, e_n)$ de $\mathbb{C}^n$ en augmentant l'indice d'une unité (modulo n). On obtient alors facilement que, pour tout $k \in \mathbb{N}$, $J^k = (m_{i,j}^k)$, où $m_{i,j}^k = 1$ si $i \equiv j+k \mod n$ et $m_{i,j}^k = 0$. On obtient en particulier $J^n = I_n$.

Ainsi une matrice $A = \begin{pmatrix} a_0 & a_{n-1} & \cdots & a_1 \\ a_1 & a_0 & \cdots & a_2 \\ \vdots & \vdots & & \vdots \\ a_{n-1} & a_{n-2} & \cdots & a_0 \end{pmatrix}$ de $\mathcal{A}$ s'écrit tout simplement comme un polynôme en J : $A = \sum_{k=0}^{n-1} a_k J^k$. On en déduit que $\mathcal{A} = \mathrm{Vect}(I, J, \ldots, J^{n-1})$ est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{C})$. La famille $(I, J, \ldots, J^{n-1})$ est libre donc $\mathcal{A}$ est de dimension n. Comme le produit de deux vecteurs de cette base appartient encore à la base et que ce produit est commutatif, par linéarité $\mathcal{A}$ est stable pour la multiplication et la multiplication est commutative dans $\mathcal{A}$. Enfin $\mathcal{A}$ contient la matrice I_n, et est donc une sous-algèbre commutative de $\mathcal{M}_n(\mathbb{C})$.

2. Puisque $J^n = I_n$, la matrice J possède un polynôme annulateur scindé à racines simples $X^n - 1$. Il existe donc D diagonale et $P \in \mathrm{GL}_n(\mathbb{C})$ telles que $J = PDP^{-1}$. On obtient, pour tout $k \in \mathbb{N}$, $J^k = PD^kP^{-1}$ et avec les notations précédentes, pour tout $A \in \mathcal{A}$,

$$A = \sum_{k=0}^{n-1} a_k J^k = P \left(\sum_{k=0}^{n-1} a_k D^k \right) P^{-1}.$$

Comme $\sum_{k=0}^{n-1} a_k D^k$ est diagonale, toutes les matrices de $\mathcal{A}$ sont diagonalisables dans la même base que J. Les valeurs propres de J étant les racines n-ièmes de l'unité, celles de A sont les $\sum_{k=0}^{n-1} a_k \omega^k$, où $\omega \in U_n$.

3. Si K est un corps quelconque, on exprime de la même façon les matrices de $\mathcal{A}$ en fonction de J. On suppose $n \geqslant 2$ sinon toutes les matrices sont diagonales.

• Si K est de caractéristique nulle, le polynôme $P = X^n - 1$ est encore scindé à racines simples puisque $P' = nX^{n-1}$ a comme seule racine 0. Le résultat de la question 2 subsiste.

• Supposons que K soit de caractéristique p, nombre premier.

Si p ne divise pas n, la seule racine de P' est encore 0 qui n'est pas racine de P. Le résultat subsiste. Si p divise n, on pose $n = pq$. On a alors $X^n - 1 = X^{pq} - 1 = (X^q - 1)^p$, car pour $1 \leqslant k \leqslant p-1$, p divise C_p^k. Les valeurs propres de J sont des racines de $X^q - 1$. Si J est diagonalisable, on a $J^q = I_n$. C'est manifestement faux : le plus petit entier pour lequel $J^k = I_n$ est n. Donc J n'est pas diagonalisable et le résultat ne subsiste pas. ◁

On peut vérifier de suite que, pour toute racine n-ième de l'unité ω, le vecteur de coordonnées $(1, \omega, \ldots, \omega^{n-1})$ est une base de la droite propre de J *associée à la valeur propre ω.*

Si on se place dans l'espace quotient $\mathbb{C}[X]/(X^n - 1)$ muni de la base $(1, X, \ldots, X^{n-1})$, la multiplication par X *(plus précisément par la classe de* X*) induit un n-cycle sur les vecteurs de la base. Cela permet de donner une réalisation abstraite des matrices circulantes étudiées dans l'exercice précédent. C'est ce que propose l'énoncé suivant qui se place toutefois dans $\mathbb{C}_{n-1}[X]$ pour éviter de parler de quotient...*

2.23. Matrices circulantes (2)

Soit $F \in \mathbb{C}_{n-1}[X]$, $F = \sum_{k=0}^{n-1} a_k X^k$ et $\Phi : \mathbb{C}_{n-1}[X] \longrightarrow \mathbb{C}_{n-1}[X]$ l'application qui à P associe le reste de la division euclidienne de PF par $X^n - 1$.

1. Vérifier que Φ est linéaire et écrire sa matrice dans la base canonique.

2. Trouver les valeurs et les vecteurs propres de Φ. L'application Φ est-elle diagonalisable ?

(École polytechnique)

▷ **Solution.**

1. Le degré de $\Phi(P)$ est inférieur à celui de $X^n - 1$, donc $\Phi(P)$ appartient à $\mathbb{C}_{n-1}[X]$. Montrons que Φ est linéaire. Si P_1 et P_2 sont deux éléments de $\mathbb{C}_{n-1}[X]$, R_1 et R_2 leurs images par Φ, il existe Q_1 et Q_2 dans $\mathbb{C}[X]$ tels que

$$FP_1 = (X^n - 1)Q_1 + R_1 \quad \text{et} \quad FP_2 = (X^n - 1)Q_2 + R_2.$$

Pour tout $(\lambda_1, \lambda_2) \in \mathbb{C}^2$, on a

$$F(\lambda_1 P_1 + \lambda_2 P_2) = (X^n - 1)(\lambda_1 Q_1 + \lambda_2 Q_2) + (\lambda_1 R_1 + \lambda_2 R_2).$$

Puisque $\lambda_1 R_1 + \lambda_2 R_2 \in \mathbb{C}_{n-1}[X]$, c'est le reste dans la division de $F(\lambda_1 P_1 + \lambda_2 P_2)$ par $X^n - 1$ et

$$\Phi(\lambda_1 P_1 + \lambda_2 P_2) = \lambda_1 R_1 + \lambda_2 R_2 = \lambda_1 \Phi(P_1) + \lambda_2 \Phi(P_2).$$

Déterminons $\Phi(X^j)$ pour $j \in [\![0, n-1]\!]$. On a $FX^j = \sum_{k=0}^{n-1} a_k X^{k+j}$. Si $k \geqslant n - j$, alors $n \leqslant k + j \leqslant 2n - 2$. On écrit $X^{k+j} = X^{k+j-n} X^n$, d'où l'on déduit $X^{k+j} \equiv X^{k+j-n}[X^n - 1]$ avec $0 \leqslant k + j - n \leqslant n - 2$. On obtient modulo $X^n - 1$

$$FX^j = \sum_{k=0}^{n-j-1} a_k X^{k+j} + \sum_{k=n-j}^{n-1} a_k X^{k+j} \equiv \sum_{k=0}^{n-j-1} a_k X^{k+j} + \sum_{k=n-j}^{n-1} a_k X^{k+j-n}.$$

Puisque ce dernier polynôme est dans $\mathbb{C}_{n-1}[X]$, on en déduit

$$\Phi(X^j) = \sum_{k=0}^{n-j-1} a_k X^{k+j} + \sum_{k=n-j}^{n-1} a_k X^{k+j-n} = \sum_{\ell=j}^{n-1} a_{\ell+n-j} X^\ell + \sum_{\ell=0}^{j-1} a_{\ell-j} X^\ell.$$

Nous en déduisons que la matrice M de Φ dans la base canonique de $\mathbb{C}_{n-1}[X]$ est

$$M = \begin{pmatrix} a_0 & a_{n-1} & \dots & a_1 \\ a_1 & a_0 & \dots & a_2 \\ \vdots & \vdots & \dots & \vdots \\ a_{n-1} & a_{n-2} & \dots & a_0 \end{pmatrix}.$$

2. La matrice de Φ dans la base canonique de $\mathbb{C}_{n-1}[X]$ est donc une matrice circulante. Nous n'allons pas utiliser les résultats de l'exercice 2.22 mais chercher directement les valeurs propres et vecteurs propres de F.

• Soit λ une valeur propre de Φ et P un vecteur propre associé. Par définition de Φ, il existe $Q \in \mathbb{C}[X]$ tel que $FP = (X^n - 1)Q + \lambda P$, c'est-à-dire

$$(F(X) - \lambda)P(X) = (X^n - 1)Q(X).$$

Si $\omega \in \mathrm{U}_n$ (l'ensemble des racines n-ièmes de l'unité), on a l'égalité $\mathrm{P}(\omega)(\mathrm{F}(\omega) - \lambda) = 0$. On en déduit que si $\lambda \neq \mathrm{F}(\omega)$, alors $\mathrm{P}(\omega) = 0$, et donc que si, pour tout $\omega \in \mathrm{U}_n$, $\lambda \neq \mathrm{F}(\omega)$, le polynôme P est nul, puisqu'il appartient à $\mathbb{C}_{n-1}[\mathrm{X}]$ et a n racines distinctes. Cela montre que $\mathrm{Sp}(\Phi) \subset \{\mathrm{F}(\omega), \omega \in \mathrm{U}_n\}$.

• Réciproquement, soit $\omega \in \mathrm{U}_n$. Alors ω est racine de $\mathrm{F}(\mathrm{X}) - \mathrm{F}(\omega)$ ainsi que de $\mathrm{X}^n - 1$. Il existe donc Q et P_ω dans $\mathbb{C}_{n-1}[\mathrm{X}]$ tels que

$$\mathrm{Q}(\mathrm{X}) = \frac{\mathrm{F}(\mathrm{X}) - \mathrm{F}(\omega)}{\mathrm{X} - \omega} \quad \text{et} \quad \mathrm{P}_\omega = \frac{\mathrm{X}^n - 1}{\mathrm{X} - \omega}.$$

On a alors $(\mathrm{F}(\mathrm{X}) - \mathrm{F}(\omega))\mathrm{P}_\omega(\mathrm{X}) = (\mathrm{X}^n - 1)\mathrm{Q}(\mathrm{X})$, c'est-à-dire

$$\mathrm{F}(\mathrm{X})\mathrm{P}_\omega(\mathrm{X}) - (\mathrm{X}^n - 1)\mathrm{Q}(\mathrm{X}) = \mathrm{F}(\omega)\mathrm{P}_\omega(\mathrm{X}).$$

On obtient donc $\Phi(\mathrm{P}_\omega) = \mathrm{F}(\omega)\mathrm{P}_\omega$, et P_ω est vecteur propre de Φ pour la valeur propre $\mathrm{F}(\omega)$. On en déduit que $\mathrm{Sp}(\Phi) = \{\mathrm{F}(\omega), \omega \in \mathrm{U}_n\}$.

• Montrons que la famille $(\mathrm{P}_\omega)_{\omega \in \mathrm{U}_n}$ est une famille libre. Supposons que $\sum_{\omega \in \mathrm{U}_n} \lambda_\omega \mathrm{P}_\omega = 0$, où les λ_ω sont des complexes. En évaluant cela en tout $\omega_0 \in \mathrm{U}_n$, il vient $\lambda_{\omega_0}\mathrm{P}_{\omega_0}(\omega_0) = 0$. Comme $\mathrm{P}_{\omega_0}(\omega_0) \neq 0$, car $\mathrm{X}^n - 1$ n'a que des racines simples, on en déduit que $\lambda_{\omega_0} = 0$, et ceci pour tout $\omega_0 \in \mathrm{U}_n$. La famille est libre, et est donc une base de $\mathbb{C}_{n-1}[\mathrm{X}]$ puisqu'elle comporte n vecteurs.

L'endomorphisme Φ est donc diagonalisable puisque $(\mathrm{P}_\omega)_{\omega \in \mathrm{U}_n}$ est une base de vecteurs propres. On notera que les $\mathrm{F}(\omega)$ ne sont pas nécessairement distincts. Pour $\lambda \in \mathrm{Sp}(\Phi)$, la dimension du sous-espace propre correspondant est égale à $\mathrm{Card}\{\omega \in \mathrm{U}_n, \lambda = \mathrm{F}(\omega)\}$. ◁

Voici maintenant quelques exercices concernant l'étude de la diagonalisabilité de matrices définies par blocs. Les variantes sont nombreuses mais les techniques sont toujours un peu les mêmes.

2.24. Diagonalisabilité d'une matrice par blocs (1)

Soit $\mathrm{U} \in \mathcal{M}_n(\mathbb{C})$ et $\mathrm{V} = \begin{pmatrix} 0 & \mathrm{I}_n \\ \mathrm{U} & 0 \end{pmatrix} \in \mathcal{M}_{2n}(\mathbb{C})$.

1. Relier les sous-espaces propres de V à ceux de U.

2. Trouver une condition nécessaire et suffisante sur U pour que V soit diagonalisable.

(École polytechnique)

▷ **Solution.**

1. Soit X et Y dans $\mathbb{C}^n$, $\mathrm{Z} = \begin{pmatrix} \mathrm{X} \\ \mathrm{Y} \end{pmatrix} \in \mathbb{C}^{2n}$, $\lambda \in \mathbb{C}$. Le vecteur VZ est égal à $\begin{pmatrix} \mathrm{Y} \\ \mathrm{UX} \end{pmatrix}$ et on en déduit que

$$Z \in \operatorname{Ker}(V - \lambda I_{2n}) \iff Y = \lambda X \text{ et } UX = \lambda Y$$
$$\iff Z = \begin{pmatrix} X \\ \lambda X \end{pmatrix} \text{ et } X \in \operatorname{Ker}(U - \lambda^2 I_n).$$

On en déduit donc que les valeurs propres de V sont les racines carrées des valeurs propres de U. Comme de plus l'application $X \mapsto \begin{pmatrix} X \\ \lambda X \end{pmatrix}$ de $\mathbb{C}^n$ dans $\mathbb{C}^{2n}$ est injective, si $\mu^2 = \lambda \in \operatorname{Sp} U$, la dimension du sous-espace propre relatif à la valeur propre μ pour V est exactement la dimension du sous-espace propre relatif à λ pour U.

2. La solution la plus naturelle consiste à utiliser la question précédente. La matrice V est diagonalisable si et seulement si la somme des dimensions de ses espaces propres vaut $2n$. Tout complexe non nul admet deux racines carrées distinctes mais pas 0. On est obligé de distinguer le noyau de V. La question 1 montre que

$$\dim \operatorname{Ker} V + \sum_{\mu \in \operatorname{Sp} V \setminus \{0\}} \dim \operatorname{Ker}(V - \mu I_{2n})$$

vaut aussi :

$$\dim \operatorname{Ker} U + 2 \sum_{\lambda \in \operatorname{Sp} U \setminus \{0\}} \dim \operatorname{Ker}(U - \mu I_n).$$

Cette somme vaut $2n$ si et seulement si $\operatorname{Ker} U = \{0\}$ et U est diagonalisable. Donc V est diagonalisable si et seulement si U est inversible et diagonalisable. ◁

Voici une autre approche possible qui ne fait pas appel à la question 1. On observe que $V^2 = \begin{pmatrix} U & 0 \\ 0 & U \end{pmatrix}$. *Il en découle que* V^2 *est diagonalisable si et seulement si* U *l'est. Si* V *est diagonalisable il en est de même de* V^2 *mais la réciproque est fausse en général (cf. exercice 2.18). Il faut en plus que* $\operatorname{Ker} V = \operatorname{Ker} V^2$ *ce qui ici n'a lieu que lorsque* U *est inversible.*

2.25. Diagonalisabilité d'une matrice par blocs (2)

Soit $A \in \mathcal{M}_n(\mathbb{C})$.

1. On pose $B = \begin{pmatrix} A & 0 \\ A & A \end{pmatrix} \in \mathcal{M}_{2n}(\mathbb{C})$. Donner une condition nécessaire et suffisante sur A pour B soit diagonalisable.

2. Même question avec $B = \begin{pmatrix} A & 0 \\ C & A \end{pmatrix} \in \mathcal{M}_{2n}(\mathbb{C})$ où $C \in \mathcal{M}_n(\mathbb{C})$ commute avec A. Que peut-on dire dans le cas où A et C ne commutent pas ?

(École polytechnique)

▷ **Solution.**

1. Regardons le cas $n = 1$ pour avoir des idées. Pour $a \in \mathbb{C}$, si la matrice $B = \begin{pmatrix} a & 0 \\ a & a \end{pmatrix}$ est diagonalisable, B est semblable à la matrice aI_2 car a est sa seule valeur propre. Mais alors $B = aI_2$ et nécessairement $a = 0$. Cela donne déjà une idée du résultat.

Supposons $B = \begin{pmatrix} A & 0 \\ A & A \end{pmatrix}$ diagonalisable. On va regarder les polynômes en B. On a $B^2 = \begin{pmatrix} A^2 & 0 \\ 2A^2 & A^2 \end{pmatrix}$ et $B^3 = \begin{pmatrix} A^3 & 0 \\ 3A^3 & A^3 \end{pmatrix}$ ce qui nous invite à penser que $B^k = \begin{pmatrix} A^k & 0 \\ kA^k & A^k \end{pmatrix}$ pour tout k. Cela se vérifie immédiatement par une récurrence sur k. Comme la matrice B est diagonalisable, elle est annulée par un polynôme scindé à racines simples P. On a alors $P(A) = 0$ et aussi, en posant $P = a_k X^k + \cdots + a_0$,

$$ka_k A^k + (k-1)a_{k-1}A^{k-1} + a_1 A = 0 \ \ i.e. \ \ (XP')(A) = 0.$$

On en déduit déjà que A doit être diagonalisable (ce qu'on pouvait voir directement car si $(e_1, \ldots, e_{2n})$ est la base canonique de $\mathbb{C}^{2n}$, le sous-espace $F = \mathrm{Vect}(e_{n+1}, \ldots, e_{2n})$ est stable par B et A est la matrice de cette restriction dans la base $(e_{n+1}, \ldots, e_{2n})$). Mais plus précisément, le polynôme minimal μ_A de A divise P et XP'. Comme les racines de P sont simples, P et P' sont premiers entre eux, et μ_A est donc égal à X. Ce qui signifie que $A = 0$.

Réciproquement, si $A = 0$, $B = 0$ est évidemment diagonalisable.

Conclusion. La matrice B est diagonalisable si, et seulement si $A = 0$.

2. • On utilise les mêmes idées. Si A et C commutent, on vérifie par récurrence que pour tout $k \geqslant 0$,

$$B^k = \begin{pmatrix} A^k & 0 \\ kCA^{k-1} & A^k \end{pmatrix}.$$

Si B est diagonalisable,elle est annulée par un polynôme scindé à racines simples P. Comme pour tout polynôme P on a $P(B) = \begin{pmatrix} P(A) & 0 \\ CP'(A) & P(A) \end{pmatrix}$, il vient d'une part $P(A) = 0$, et A est donc diagonalisable, et d'autre part $CP'(A) = 0$. Or la matrice $P'(A)$ est inversible : en effet, comme P est scindé à racines simples, P et P' sont premiers entre eux, si bien qu'il existe $(U, V) \in (\mathbb{C}[X])^2$ tel que $UP + VP' = 1$ d'après le théorème de Bezout. En évaluant en A, il vient $V(A)P'(A) = I_n$ et $P'(A)$ est bien inversible. Par conséquent $C = 0$.

Réciproquement, si A est diagonalisable et si $C = 0$, il existe P scindé à racines simples annulant A. Dans ces conditions, on a $P(B) = 0$ et B est diagonalisable.

• Dans le cas où C et A ne commutent pas, B peut être diagonalisable sans que C soit nulle (par contre A doit être nécessairement diagonalisable). Il suffit de prendre $A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}$ et $C = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ pour avoir un contre-exemple. On vérifie que $B^2 = B$ donc B est diagonalisable (c'est la matrice d'un projecteur) mais C est nilpotente non nulle. ◁

Si $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(K)$ *et* $B \in \mathcal{M}_m(K)$, *la matrice* $A \otimes B$ *définie comme la matrice par blocs* $\begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & & \vdots \\ a_{n1}B & \dots & a_{nn}B \end{pmatrix}$ *est appelée produit tensoriel de* A *par* B *(ou encore produit de Kronecker). Elle sert à représenter le produit tensoriel de deux applications linéaires. Voici un exercice traitant de ce produit.*

2.26. Diagonalisabilité d'un produit tensoriel de matrices

Soit $A \in \mathcal{M}_n(\mathbb{C})$ et $B = \begin{pmatrix} A & 2A \\ -A & 2A \end{pmatrix}$. Montrer que A est diagonalisable si, et seulement si, B est diagonalisable.

(École polytechnique)

▷ **Solution.**

La matrice $\begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix}$ est diagonalisable dans $\mathcal{M}_2(\mathbb{C})$ car son polynôme caractéristique $X^2 - 3X + 4$ est scindé à racines simples puisque son discriminant est $9 - 16 = -7 \neq 0$. Il existe donc $\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C})$ et $(\lambda, \mu) \in (\mathbb{C}^*)^2$ tels que

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{-1} \begin{pmatrix} 1 & 2 \\ -1 & 2 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} = \begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}.$$

Notons $\begin{pmatrix} \alpha' & \beta' \\ \gamma' & \delta' \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{-1}$, $P = \begin{pmatrix} \alpha I_n & \beta I_n \\ \gamma I_n & \delta I_n \end{pmatrix}$, $Q = \begin{pmatrix} \alpha' I_n & \beta' I_n \\ \gamma' I_n & \delta' I_n \end{pmatrix}$. On a $PQ = QP = \begin{pmatrix} I_n & 0 \\ 0 & I_n \end{pmatrix}$. La matrice P est inversible d'inverse Q. Dans ces conditions, on trouve

$$P^{-1} \begin{pmatrix} A & 2A \\ -A & 2A \end{pmatrix} P = \begin{pmatrix} \lambda A & 0 \\ 0 & \mu A \end{pmatrix}.$$

Si A est diagonalisable, il existe des matrices $Q \in \mathrm{GL}_n(\mathbb{C})$ et $D \in \mathcal{M}_n(\mathbb{C})$ diagonale, telles que $Q^{-1}AQ = D$. On a alors

$$\begin{pmatrix} Q^{-1} & 0 \\ 0 & Q^{-1} \end{pmatrix}\begin{pmatrix} \lambda A & 0 \\ 0 & \mu A \end{pmatrix}\begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix} = \begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}^{-1}\begin{pmatrix} \lambda A & 0 \\ 0 & \mu A \end{pmatrix}\begin{pmatrix} Q & 0 \\ 0 & Q \end{pmatrix}$$
$$= \begin{pmatrix} \lambda D & 0 \\ 0 & \mu D \end{pmatrix}.$$

On en déduit que B est diagonalisable.

Réciproquement, si B est diagonalisable, la matrice $\begin{pmatrix} \lambda A & 0 \\ 0 & \mu A \end{pmatrix}$ l'est aussi. Si on considère un polynôme annulateur P scindé à racines simples de cette matrice, on a en particulier $P(\lambda A) = 0$. Donc λA et finalement A sont diagonalisables. $\lhd$

En reprenant l'idée des matrices de passage par blocs, on démontre facilement que si M *et* N *sont des matrices carrées diagonalisables, alors la matrice* $M \otimes N$ *est diagonalisable.*

Le résultat de l'exercice qui suit concerne la théorie algébrique des graphes (le lecteur trouvera l'interprétation du résultat en termes de graphes après la solution). Il a aussi fait l'objet d'une partie de l'épreuve d'algorithmique du concours 1996 de l'ENS Lyon. Concernant la réduction on y utilise juste le fait qu'une matrice symétrique réelle est diagonalisable.

2.27. Théorème de Hoffman et Singleton (1960)

Soit A une matrice carrée réelle d'ordre n à coefficients dans $\{0, 1\}$, de trace nulle, symétrique et telle qu'il existe un entier $d \geqslant 1$ vérifiant $A^2 + A - (d-1)I_n = J$ où J est la matrice carrée d'ordre n dont tous les coefficients valent 1. On note U le vecteur colonne dont tous les coefficients sont égaux à 1.

1. Montrer que $AU = dU$. En déduire que $n = d^2 + 1$.

2. Soit a, b les racines du polynôme $X^2 + X - (d-1)$. Montrer que le spectre de A est inclus dans $\{a, b, d\}$.

3. Montrer que $d \in \{1, 2, 3, 7, 57\}$. Déterminer A pour $d = 1$ ou $d = 2$.

(École polytechnique)

$\rhd$ **Solution.**

1. Notons $A=(a_{ij})_{1\leqslant i,j\leqslant n}$ et $A^2 = (b_{ij})_{1\leqslant i,j\leqslant n}$. La nullité de la trace de A implique la nullité de tous les coefficients diagonaux a_{ii}. Par ailleurs, on a

$$b_{ij} = \sum_{k=1}^{n} a_{ik}a_{kj} = \sum_{k=1}^{n} a_{ik}a_{jk},$$

car A est symétrique. Il s'agit du nombre de 1 communs à la ligne i et à la ligne j de A. En particulier, b_{ii} est égal au nombre de 1 qu'il y a dans la ligne i. La relation vérifiée par A conduit à $b_{ii} + a_{ii} - (d-1) = 1$ c'est-à-dire $b_{ii} = d$. Or, AU est exactement le vecteur dont la i-ième coordonnée est la somme des termes de la i-ième ligne de A, somme qui vaut donc d. Ainsi, $AU = dU$.

Il vient alors $A^2U + AU - (d-1)U = JU = nU$, donc

$$d^2 + d - (d-1) = d^2 + 1 = n.$$

2. Soit λ une valeur propre de A, X un vecteur propre associé. On a alors $JX = (\lambda^2 + \lambda - (d-1))X$. Donc X est aussi un vecteur propre de J pour la valeur propre $\lambda^2 + \lambda - (d-1)$. La matrice J est de rang 1 et diagonalisable : ses valeurs propres sont 0 et n, l'espace propre associé à n étant la droite vectorielle dirigée par U. Il en résulte que, soit X est lié à U et alors d'après la question 1 $\lambda = d$, soit $X \in \operatorname{Ker} J$ et alors $\lambda \in \{a, b\}$. C'est le résultat voulu.

3. Observons que a, b, d sont deux à deux distincts. La matrice A est symétrique réelle donc diagonalisable. L'espace propre associé à d est la droite dirigée par U. Notons α et β les dimensions des espaces propres associés à a et b. On a

$$n = 1 + \alpha + \beta = d^2 + 1 \tag{1}$$

car A est diagonalisable. Le fait que la trace de A est nulle conduit à la seconde relation

$$\alpha a + \beta b + d = 0 \tag{2}.$$

Les contraintes sur d vont provenir de ce que α et β doivent être *entiers*. Explicitons pour commencer a et b :

$$a = \frac{-1 + \sqrt{4d-3}}{2} \quad \text{et} \quad b = \frac{-1 - \sqrt{4d-3}}{2}.$$

En substituant dans (2) on obtient

$$-\frac{\alpha+\beta}{2} + \sqrt{4d-3}\frac{\alpha-\beta}{2} = -d$$

et en remplaçant $\alpha + \beta$ par d^2 : $\sqrt{4d-3}(\alpha - \beta) = d(d-2)$ (3).

• Si $\alpha = \beta$, on a nécessairement $d = 2$.

• Si $\alpha \neq \beta$, $\sqrt{4d-3}$ est un nombre rationnel. Donc $4d-3$ est un carré d'entier : $4d - 3 = p^2$. La relation (3) impose alors que p divise $d(d-2) = \dfrac{p^2+3}{4} \cdot \dfrac{p^2-5}{4}$ et donc que $16p$ divise $(p^2+3)(p^2-5)$. Cela impose $p | 15$ c'est-à-dire $p \in \{1, 3, 5, 15\}$.

Les seules valeurs possibles de d sont donc $1, 2, 3, 7, 57$.

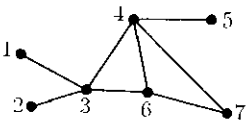
Si $d = 1$, alors $n = 2$. Les termes diagonaux de A sont nuls et chaque ligne de A contient un terme égal à 1. On trouve une seule solution $\mathrm{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, dont on vérifie qu'elle convient.

Si $d = 2$, alors $n = 5$. Chaque ligne de A contient exactement deux termes égaux à 1 et les termes diagonaux sont nuls. Si on échange deux lignes de A ainsi que les colonnes correspondantes, la matrice obtenue reste symétrique et il est clair, au vu des relations que doivent vérifier les coefficients, qu'elle est encore solution. On peut donc supposer qu'on a $a_{12} = a_{13} = 1$ et $a_{14} = a_{15} = 0$. Comme $b_{12} = 1 - a_{12} = 0$, les lignes 1 et 2 n'ont pas de 1 en commun et $a_{23} = 0$. On a donc $a_{32} = 0$. De $b_{23} = 1 - a_{23} = 1$, on déduit que les lignes 2 et 3 n'ont qu'un 1 en commun, le premier terme. On a donc $a_{24} = 0$ ou $a_{34} = 0$. Quitte à échanger les deux dernières lignes et colonnes, on peut supposer $a_{24} = 1$ et $a_{34} = 0$. Alors A est entièrement déterminée :

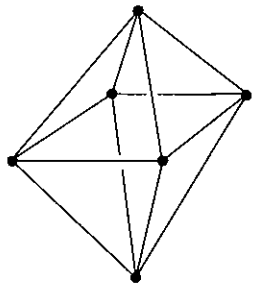
$$\mathrm{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

et cette matrice convient. Le raisonnement qui précède montre qu'une fois fixés les coefficients non nuls de la première ligne a_{1i} et a_{1j}, il y a deux possibilités pour les lignes i et j et qu'ensuite A est entièrement déterminé. On obtient donc $\mathrm{C}_4^2 \times 2 = 12$ matrices différentes qui s'obtiennent à partir de la précédente en permutant les lignes et les colonnes. ◁

Exposons un peu de théorie des graphes pour bien comprendre le résultat que l'on vient de démontrer. Un graphe G *est un couple* (S, A)*, où* S *est un ensemble fini dont les éléments sont appelés les sommets et* A *un ensemble de paires d'éléments distincts de* S*, appelés les arêtes (par paire, on entend un ensemble à deux éléments* $\{x, y\}$*). On appelle arête une paire de sommets parce que sur un graphe concret (voir figure), il suffit de donner les deux extrémités d'une arête pour la définir. La terminologie sommets, arêtes provient du graphe naturellement associé à un polyèdre. Le cardinal de* S *est appelé l'ordre du graphe. Deux sommets* x *et* y *sont adjacents si* $\{x, y\} \in \mathrm{A}$*. Étant donné un sommet* x*, le nombre de sommets adjacents à* x *est appelé le degré de* x*. Un chemin de longueur* p *entre les sommets* x *et* y *est une suite* $(x = s_0, s_1, \ldots, s_p = y)$ *de sommets adjacents. La distance entre deux sommets* x *et* y *est la longueur minimale d'un chemin de* x *à* y *et le diamètre du graphe est la distance maximale entre deux sommets distincts.*



Un graphe



Graphe associé à un polyèdre de l'espace

Un graphe d'ordre n peut être représenté par sa matrice d'adjacence $A \in \mathcal{M}_n(\mathbb{R})$. *On numérote les sommets* $S = \{x_i,\ i \in [\![1, n]\!]\}$ *et on définit* $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ *par* $a_{ij} = 1$ *si* $\{x_i, x_j\} \in A$ *et* $a_{ij} = 0$ *sinon. La matrice d'adjacence est donc une matrice symétrique.*

On considère un graphe d'ordre n, de diamètre 2 (deux sommets quelconques sont donc à une distance 1 ou 2) et de degré maximal d. On a alors $n \leqslant d^2 + 1$. *En effet, si on fixe un sommet s, il possède au plus d voisins et chacun de ceux-ci possède au plus $d - 1$ voisins autres que s. Donc le graphe possède au plus* $1 + d + d(d-1) = d^2 + 1$ *sommets. S'il possède le nombre maximal de sommets, c'est-à-dire* $d^2 + 1$, *on dit que c'est un graphe de Moore.*

Le raisonnement précédent montre qu'alors chaque sommet est de degré d (on dit que le graphe est d-régulier) et est relié à tout autre sommet par exactement un chemin (d'ordre 1 ou 2).

Montrons que la matrice d'adjacence A d'un graphe de Moore vérifie l'équation $A^2 + A - (d-1)I_n = J$. *En effet, avec les notations de l'énoncé, cela équivaut à* $b_{ij} + a_{ij} = 1 + (d-1)\delta_{ij}$. *On remarque que* $b_{ij} = \sum_{k=1}^{n} a_{ik}a_{ki}$ *est le nombre de chemins de longueur 2 du sommet i au sommet j.*

• *Pour $j = i$, on a* $a_{ii} = 0$ *et* $b_{ii} = \sum_{k=1}^{n} a_{ik}^2 = d$, *nombre de sommets voisins du i-ième et donc* $a_{ii} + b_{ii} = d$.

• *Si $i \neq j$, soit* $a_{ij} = 0$ *et alors les sommets i et j sont reliés par un chemin d'ordre 2 et* $b_{ij} = 1$, *soit* $a_{ij} = 1$ *et alors i et j ne sont pas reliés par un chemin d'ordre 2 donc* $b_{ij} = 0$. *On a dans tous les cas* $a_{ij} + b_{ij} = 1$.

L'exercice 2.27 montre que $d \in \{2, 3, 7, 57\}$ *(le cas $d = 1$ doit être écarté car alors $n = 2$ et le diamètre est 1). Pour $d = 2$, $d = 3$ et $d = 7$, on a montré l'existence de graphes de Moore. Pour $d = 57$, on ignore encore s'il en existe un.*

Voici maintenant une série d'exercices concernant la trigonalisation. Rappelons le résultat fondamental : une matrice est trigonalisable si et seulement si son polynôme caractéristique est scindé. Le premier exercice

est un lemme classique permettant de caractériser les endomorphismes nilpotents, il sera maintes fois utilisé dans la suite.

2.28. Caractérisation des matrices nilpotentes avec la trace

Soit K un sous-corps de $\mathbb{C}$ et $A \in \mathcal{M}_n(K)$. On suppose que pour tout $k \geqslant 1$, la trace de A^k est nulle. Montrer que A est nilpotente.
(École polytechnique)

▷ **Solution.**

Nous allons proposer trois solutions de cette question classique :

• Le polynôme caractéristique de A est scindé sur $\mathbb{C}$. Raisonnons par l'absurde et supposons A non nilpotente. Alors A possède des valeurs propres (complexes) non nulles. On va noter $\lambda_1, \ldots, \lambda_r$ ces valeurs propres non nulles de A ($r \geqslant 1$) et $n_1, \ldots, n_r$ leurs multiplicités respectives. Nous savons que A est semblable à une matrice triangulaire avec sur la diagonale les valeurs propres apparaissant autant de fois que leur multiplicité. En élevant à la puissance k-ième cette matrice triangulaire, on obtient une matrice triangulaire semblable à A^k si bien que pour tout $k \geqslant 1$, on a

$$\mathrm{Tr}(A^k) = n_1\lambda_1^k + \cdots + n_r\lambda_r^k = 0.$$

Si on écrit ses relations pour k variant de 1 à r, on obtient que $(n_1, \ldots, n_r)$ est solution du système linéaire

$$\begin{pmatrix} \lambda_1 & \lambda_2 & \ldots & \lambda_r \\ \lambda_1^2 & \lambda_2^2 & \ldots & \lambda_r^2 \\ \vdots & & & \vdots \\ \lambda_1^r & \lambda_2^r & \ldots & \lambda_r^r \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{pmatrix} = 0.$$

Or, ce système est de Cramer puisque le déterminant de la matrice du système vaut

$$\lambda_1 \ldots \lambda_r \prod_{1 \leqslant i < j \leqslant r} (\lambda_j - \lambda_i) \neq 0 \quad \text{(cf. exercice 1.10)}$$

Nécessairement, on a $n_1 = \cdots = n_r = 0$, ce qui est exclu.

• Voici une autre solution à l'aide des formules de Newton. On note désormais $\lambda_1, \ldots, \lambda_n$ les racines complexes de χ_A prises avec multiplicité. D'après ce qui a été vu plus haut, dire que $\mathrm{Tr}(A^k) = 0$ pour tout $k \geqslant 1$, c'est exactement dire que les sommes de Newton relatives à $\lambda_1, \ldots, \lambda_n$ sont nulles :

$$\mathrm{Tr}(A^k) = \lambda_1^k + \cdots + \lambda_n^k = 0.$$

Les formules de Newton (voir exercice 5.31 tome 1 d'algèbre) donnent alors que les fonctions symétriques élémentaires des λ_i sont nulles : $\sigma_1 = \cdots = \sigma_n = 0$. On en déduit que

$$\chi_A = X^n - \sigma_1 X^{n-1} + \cdots + (-1)^{n-1} X + (-1)^n \sigma_n = X^n,$$

et $A^n = 0$ d'après le théorème de Cayley-Hamilton.

• Voici enfin une troisième solution par récurrence sur la taille de la matrice. Si $n = 1$, comme $\mathrm{Tr}(A) = 0$, on a $A = 0$. Soit $n \geqslant 2$. Supposons le résultat vrai jusqu'au rang $n-1$. Écrivons le polynôme caractéristique de A :

$$\chi_A = X^n - \sigma_1 X^{n-1} + \cdots + (-1)^{n-1} X + (-1)^n \sigma_n.$$

D'après le théorème de Cayley-Hamilton, on a

$$\chi_A(A) = 0 = A^n - \sigma_1 A^{n-1} + \cdots + (-1)^{n-1} A + (-1)^n \sigma_n I_n,$$

et en exprimant la trace de cette somme, par hypothèse, il vient

$$(-1)^n \sigma_n n = 0,$$

ou encore $n \det A = 0$, soit $\det A = 0$. Donc 0 est valeur propre de A. Écrivons $\chi_A = X^p Q$ avec $Q(0) \neq 0$, $p \geqslant 1$.

D'après le théorème de décomposition de noyaux, on a

$$K^n = \mathrm{Ker}(\chi_A(A)) = \mathrm{Ker}\, A^p \oplus \mathrm{Ker}\, Q(A).$$

Supposons $\mathrm{Ker}\, Q(A) \neq \{0\}$. Prenons une base obtenue par réunion d'une base de $\mathrm{Ker}\, A^p$ et d'une base de $\mathrm{Ker}\, Q(A)$. L'endomorphisme A dans cette base admet une matrice de la forme

$$\begin{pmatrix} \boxed{A'} & 0 \\ 0 & \boxed{B'} \end{pmatrix}.$$

On a $A'^p = 0$ donc A' est nilpotente et, pour tout $k \geqslant 1$, $\mathrm{Tr}(A'^k) = 0$. Comme A^k est semblable à

$$\begin{pmatrix} \boxed{A'^k} & 0 \\ 0 & \boxed{B'^k} \end{pmatrix}.$$

on a $\mathrm{Tr}(B'^k) = 0$. D'autre part, la restriction de A à $\mathrm{Ker}\, Q(A)$ est injective, si bien que B' est inversible. Or par hypothèse de récurrence, B' est aussi nilpotente ce qui fournit la contradiction.

On peut donc affirmer que $\mathrm{Ker}\, Q(A) = \{0\}$ et $K^n = \mathrm{Ker}\, A^p$: la matrice A est bien nilpotente. ◁

Les solutions reste valables sur un corps commutatif de caractéristique nulle quelconque à condition d'admettre l'existence d'un surcorps dans lequel le polynôme caractéristique est scindé.

Lorsque u est un endomorphisme trigonalisable de valeurs propres $(\lambda_1, \ldots, \lambda_n)$ complétés avec multiplicité, pour tout $P \in K[X]$, le n-uplet $(P(\lambda_1), \ldots, P(\lambda_n))$ correspond aux valeurs propres de $P(u)$ comptées avec multiplicité. Cela est encore utilisé dans l'exercice suivant qui montre un résultat déjà établi, mais différemment, dans notre tome 1 d'algèbre (exercice 5.33). Comme dans l'exercice 1.17, on utilise le fait que tout polynôme est le polynôme caractéristique d'une certaine matrice.

2.29. Théorème de Kronecker

1. Soit $P \in \mathbb{Z}[X]$ unitaire de degré n. Existe-t-il $A \in \mathcal{M}_n(\mathbb{Z})$ telle que $\chi_A = P$?

2. On note $\lambda_1, \ldots, \lambda_n \in \mathbb{C}$ les racines distinctes ou confondues d'un polynôme P unitaire de $\mathbb{Z}[X]$. Montrer que si $q \in \mathbb{N}^*$,

$$P_q = \prod_{i=1}^{n} (X - \lambda_i^q) \in \mathbb{Z}[X].$$

3. Soit $P \in \mathbb{Z}[X]$ unitaire dont toutes les racines complexes sont de module inférieur ou égal à 1. Montrer que les racines non nulles de P sont des racines de l'unité.

(**École polytechnique**)

▷ **Solution.**

1. Il suffit de penser à la matrice compagnon du polynôme P : si on écrit $P = X^n + a_{n-1}X^{n-1} + \cdots + a_1X + a_0$, il s'agit de la matrice

$$A = \begin{pmatrix} 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & & & -a_1 \\ 0 & 1 & \ddots & & -a_2 \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \ldots & \ldots & 1 & -a_{n-1} \end{pmatrix}.$$

Cette matrice admet le polynôme P comme polynôme caractéristique : le lecteur trouvera ce calcul classique dans l'exercice 1.17.

2. On reprend la matrice A introduite à la question précédente. Dans $\mathcal{M}_n(\mathbb{C})$, la matrice A est semblable à une matrice triangulaire supérieure T avec $\lambda_1, \ldots, \lambda_n$ sur la diagonale. Dans ces conditions, si $q \in \mathbb{N}^*$, A^q est semblable à T^q. Comme A^q est à coefficients entiers, son polynôme caractéristique est à coefficients entiers, autrement dit,

$$\chi_{A^q} = \chi_{T^q} - \prod_{i=1}^{n}(X - \lambda_i^q) = P_q \in \mathbb{Z}[X].$$

3. On reprend les notations de la question précédente. L'argument crucial est que, du fait des hypothèses sur les racines de P, les polynômes P_q sont en nombre fini. En effet, le coefficient de X^{n-k} dans P_q est $\sum_{i_1<\cdots<i_k} \lambda_{i_1} \ldots \lambda_{i_k}$ qu'on peut majorer en module :

$$\left| \sum_{i_1<\cdots<i_k} \lambda_{i_1} \ldots \lambda_{i_k} \right| \leqslant \sum_{i_1<\cdots<i_k} 1 = C_n^k \leqslant 2^n.$$

Comme les coefficients de P_q sont entiers, ils ne peuvent prendre qu'un nombre fini de valeurs et les P_q sont en nombre fini. Les n-uplets $(\lambda_1^q, \ldots, \lambda_n^q)$ sont donc eux aussi en nombre fini. D'après le lemme des tiroirs, il existe donc $q \neq q'$ dans $\mathbb{N}^*$ tels que

$$(\lambda_1^q, \ldots, \lambda_n^q) = (\lambda_1^{q'}, \ldots, \lambda_n^{q'}).$$

Si λ_i est non nulle, $\lambda_i^q = \lambda_i^{q'}$ et $\lambda_i^{q-q'} = 1$ ce qui fait de λ_i une racine de l'unité. ◁

Les exercices proposés maintenant concernent presque tous des questions de trigonalisation simultanée.

2.30. Trigonalisation simultanée (1)

Soit A, B deux matrices de $\mathcal{M}_n(\mathbb{C})$ qui commutent. Montrer que A et B possèdent un vecteur propre commun, puis établir qu'elles sont trigonalisables simultanément.

(École polytechnique)

▷ **Solution.**

• Comme $\mathbb{C}$ est algébriquement clos, le polynôme caractéristique de A qui est de degré $n \geqslant 1$ admet au moins une racine. Donc A admet au moins une valeur propre λ. L'espace propre $\mathrm{Ker}(A - \lambda I_n)$ est stable par B car A et B commutent. Le même argument montre que la restriction de B à cet espace propre admet une valeur propre μ. Si X est un vecteur propre associé on a alors $BX = \mu X$ et $AX = \lambda X$. Donc X répond au problème.

• On va montrer par récurrence sur n qu'il existe $S \in \mathrm{GL}_n(\mathbb{C})$ tel que $S^{-1}AS$ et $S^{-1}BS$ soient triangulaires supérieures. C'est trivial si $n = 1$. Supposons $n \geqslant 2$ et le résultat vrai au rang $n - 1$. D'après la première

question, il existe $e \in \mathbb{C}^n$ non nul et λ et μ complexes vérifiant $Ae = \lambda e$ et $Be = \mu e$. Complétons e en une base $(e, \varepsilon_2, \ldots, \varepsilon_n)$ de $\mathbb{C}^n$. Si Q désigne la matrice de passage de la base canonique de $\mathbb{C}^n$ à $(e, \varepsilon_2, \ldots, \varepsilon_n)$, on a :

$$Q^{-1}AQ = \left(\begin{array}{c|c} \lambda & \times \cdots \times \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right) \quad \text{et} \quad Q^{-1}BQ = \left(\begin{array}{c|c} \mu & \times \cdots \times \\ \hline 0 & \\ \vdots & B' \\ 0 & \end{array}\right).$$

Comme A et B commutent, il en va de même pour A' et B'. D'après l'hypothèse de récurrence, il existe $R \in GL_{n-1}(\mathbb{C})$ tel que les matrices $R^{-1}A'R = T$ et $R^{-1}B'R = U$ soient triangulaires supérieures. Posons

$$P = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & R \\ 0 & \end{array}\right). \text{ Alors } P \in GL_n(\mathbb{C}) \text{ et } P^{-1} = \left(\begin{array}{c|c} 1 & 0 \cdots 0 \\ \hline 0 & \\ \vdots & R^{-1} \\ 0 & \end{array}\right).$$

On en déduit

$$P^{-1}Q^{-1}AQP = \left(\begin{array}{c|c} \lambda & \times \cdots \times \\ \hline 0 & \\ \vdots & R^{-1}A'R \\ 0 & \end{array}\right) = \left(\begin{array}{c|c} \lambda & \times \cdots \times \\ \hline 0 & \\ \vdots & T \\ 0 & \end{array}\right) \in T_n(\mathbb{C})$$

et de même,

$$P^{-1}Q^{-1}BQP = \left(\begin{array}{c|c} \mu & \times \cdots \times \\ \hline 0 & \\ \vdots & R^{-1}B'R \\ 0 & \end{array}\right) = \left(\begin{array}{c|c} \mu & \times \cdots \times \\ \hline 0 & \\ \vdots & U \\ 0 & \end{array}\right) \in T_n(\mathbb{C}).$$

Comme $(QP)^{-1} = P^{-1}Q^{-1}$, la matrice inversible $S = QP$ convient. ◁

On a déjà signalé plus haut (cf. page 99) que si $(A_i)_{i\in I}$ est une famille commutative quelconque de matrices de $\mathcal{M}_n(\mathbb{C})$ on peut trouver un vecteur propre commun à toutes les matrices de la famille. La même récurrence montre alors que la famille $(A_i)_{i\in I}$ est cotrigonalisable.

Bien entendu, comme deux matrices triangulaires ne commutent pas forcément, le fait que A *et* B *commutent est suffisant mais n'est pas nécessaire pour que* A *et* B *soient cotrigonalisables. On va voir dans la suite plusieurs autres conditions suffisantes. La variante suivante est une première généralisation puisqu'on retrouve l'exercice précédent si* $C = 0$. *Les mêmes idées interviennent dans la solution.*

2.31. Trigonalisation simultanée (2)

Soient A, B et C trois matrices de $\mathcal{M}_n(\mathbb{C})$ telles que

$$AB - BA = C, \quad AC = CA, \quad BC = CB.$$

1. Montrer que ces trois matrices ont un vecteur propre commun.

2. Montrer que A, B et C sont trigonalisables simultanément.

(École polytechnique)

▷ **Solution.**

1. On remarque que si X est un vecteur propre pour A et B, alors X est forcément un vecteur du noyau de C. On va donc regarder ce qui se passe sur le noyau de C.

• Supposons que $\operatorname{Ker} C$ n'est pas réduit à $\{0\}$. Alors les endomorphismes définis par A et B laissent stable ce sous-espace puisque A et B commutent avec C. Leurs restrictions A′ et B′ à $\operatorname{Ker} C$ commutent car si $X \in \operatorname{Ker} C$,

$$ABX - BAX = CX = 0.$$

Comme $\mathbb{C}$ est algébriquement clos, A′ et B′ ont un vecteur propre commun (cf. exercice précédent).

• Il reste donc à montrer que $\operatorname{Ker} C \neq \{0\}$. Supposons par l'absurde C inversible. On a alors

$$C^{-1}AB - C^{-1}BA = I_n$$

Or, $\operatorname{Tr}(C^{-1}AB) = \operatorname{Tr}(BC^{-1}A) = \operatorname{Tr}(C^{-1}BA)$ car B et C^{-1} commutent. C'est absurde car I_n serait de trace nulle.

2. En prenant une base $\mathcal{B}$ qui admet comme premier vecteur un vecteur propre commun à A, B et C, on peut trouver $P \in \mathrm{GL}_n(\mathbb{C})$ telle que

$$P^{-1}AP = \left(\begin{array}{c|c} \lambda & \times \cdots \times \\ \hline 0 & \\ \vdots & A' \\ 0 & \end{array}\right), \quad P^{-1}BP = \left(\begin{array}{c|c} \mu & \times \cdots \times \\ \hline 0 & \\ \vdots & B' \\ 0 & \end{array}\right),$$

$$P^{-1}CP = \left(\begin{array}{c|c} 0 & \times \cdots \times \\ \hline 0 & \\ \vdots & C' \\ 0 & \end{array}\right).$$

On a alors $A'B' - B'A' = C'$, $A'C' = C'A'$ et $B'C' = C'B'$. Une récurrence sur la taille de la matrice permet de conclure. ◁

2.32. Trigonalisation simultanée (3)

1. Soient $A, B \in \mathcal{M}_n(\mathbb{C})$ telles que $AB - BA = B$. Montrer que B est nilpotente.

2. Soient $A, B \in \mathcal{M}_n(\mathbb{C})$ telles qu'il existe $\lambda, \mu \in \mathbb{C}$ vérifiant $AB - BA = \lambda A + \mu B$. Montrer que A et B ont un vecteur propre commun, puis qu'elles sont simultanément trigonalisables.

(École polytechnique)

▷ **Solution.**

1. On calcule les puissances de B à partir de la relation $AB - BA = B$. En multipliant par B à droite on a $AB^2 - BAB = B^2$ et en multipliant par B à gauche $BAB - B^2A = B^2$. Ainsi, $2B^2 = AB^2 - B^2A$. Montrons alors par récurrence sur $k \geqslant 1$ que $AB^k - B^kA = kB^k$. C'est vu pour $k = 1$ ou $k = 2$. Supposons $k \geqslant 3$ et le résultat vrai au rang $k - 1$. On a par hypothèse de récurrence

$$\begin{aligned} AB^k - B^kA &= (AB^{k-1} - B^{k-1}A)B + B^{k-1}(AB - BA) \\ &= (k-1)B^k + B^k = kB^k. \end{aligned}$$

D'où le résultat. Pour en déduire que B est nilpotente, nous vous proposons pas moins de trois arguments.

• On peut utiliser le résultat classique de l'exercice 2.28 puisque pour tout $k \geqslant 1$, on a $\operatorname{Tr}(kB^k) = \operatorname{Tr}(AB^k) - \operatorname{Tr}(B^kA) = 0$ et donc $\operatorname{Tr}(B^k) = 0$.

• Il est aussi possible d'utiliser la notion de polynôme minimal. En effet, les relations précédentes montrent que pour tout polynôme P on a

$$AP(B) - P(B)A = BP'(B).$$

Si on prend pour P le polynôme minimal μ de B, on voit que $X\mu'(X)$ annule aussi B. Donc $\mu(X)$ divise $X\mu'(X)$. Ces deux polynômes ayant le même degré, ils sont colinéaires. La seule racine de μ est alors 0 (l'ordre de multiplicité d'une autre racine serait le même dans μ et dans μ' ce qui est impossible). Donc $\mu = X^p$ pour un certain entier p et $B^p = 0$: B est nilpotente.

• Voici enfin un dernier argument très court : considérons l'endomorphisme de $\mathcal{M}_n(\mathbb{C})$ $f : M \mapsto AM - MA$. Les égalités obtenues plus haut signifient que si B^k est non nulle, alors k est dans le spectre de f. Comme $\mathcal{M}_n(\mathbb{C})$ est de dimension finie, le spectre de f est fini et les puissances de B sont donc forcément nulles à partir d'un certain rang.

2. Si $\lambda = \mu = 0$, A et B commutent et on se retrouve avec la situation de l'exercice 2.30. On supposera donc $(\lambda, \mu) \neq 0$ et pour des raisons de symétrie, on peut supposer $\mu \neq 0$. Comme dans les deux exercices précédent il suffit de prouver que A et B admettent un vecteur propre

commun, la cotrigonalisation s'obtenant alors facilement par récurrence sur la taille n (la relation se transmet bien au rang inférieur).

Pour cela, on se ramène à une relation de la forme étudiée dans la question 1. On pose $B' = \lambda A + \mu B$. On a alors $AB' - B'A = \mu B'$. La question précédente, appliquée avec $\frac{1}{\mu}A$ et B' montre que B' est nilpotente. En particulier son noyau est non nul. Si $X \in \operatorname{Ker} B'$ on a $B'AX = 0$ ce qui montre que $\operatorname{Ker} B'$ est stable par A. Comme on travaille sur $\mathbb{C}$, la restriction de A à $\operatorname{Ker} B'$ admet une valeur propre α. Soit X un vecteur propre associé. On a alors $B'X = 0$ ce qui donne $\lambda AX + \mu BX = 0$ soit encore $BX = -\dfrac{\lambda\alpha}{\mu}X$. Donc X est un vecteur propre commun à A et à B. ◁

Voici encore une question du même type, mais plus difficile.

2.33. Trigonalisation simultanée (4)

Soient A et B dans $\mathcal{M}_n(\mathbb{C})$ telles que $\operatorname{rg}(AB - BA) \leqslant 1$. Montrer que A et B sont cotrigonalisables.

(École normale supérieure)

▷ **Solution.**

Posons $C = AB - BA$. Si C est nulle, alors A et B commutent et on retrouve la situation de l'exercice 2.30. On va donc supposer C de rang 1. Une première idée est de chercher un vecteur propre commun à A et B pour faire ensuite une récurrence sur la taille n. Il est clair qu'un tel vecteur propre commun X de A et B doit être dans le noyau de C. Par hypothèse, ce noyau est un hyperplan de $\mathbb{C}^n$. Soit λ une valeur propre de A. Quitte à remplacer A par $A' = A - \lambda I_n$, on peut supposer que $\lambda = 0$ (on a toujours $A'B - BA' = C$). Deux cas se présentent :

• Dans le cas où $\operatorname{Ker} A \subset \operatorname{Ker} C$, tout se passe bien. En effet, pour tout X de $\operatorname{Ker} A$ on a $ABX = CX + BAX = 0$. Donc $\operatorname{Ker} A$ est stable par B. Un vecteur propre de la restriction de B à $\operatorname{Ker} A$ est alors vecteur propre commun à A et B.

• Regardons maintenant le cas où $\operatorname{Ker} A$ n'est pas inclus dans $\operatorname{Ker} C$. On peut donc trouver $X \in \mathbb{C}^n$ tel que $AX = 0$ et $CX \neq 0$. On a alors $CX = ABX \in \operatorname{Im} A$. Comme CX est un vecteur non nul de l'image de C et que $\operatorname{Im} C$ est une droite, on a $\operatorname{Im} C \subset \operatorname{Im} A$. Dans ce cas on voit que $\operatorname{Im} A$ est stable par B : en effet, soit $Y \in \operatorname{Im} A$ et Z tel que $Y = AZ$. Alors $ABZ - BAZ = CZ$ ce qui donne $BY = ABZ - CZ \in \operatorname{Im} A$. Problème : on ne voit pas pourquoi A et B auraient un vecteur propre commun dans $\operatorname{Im} A$.

Mais en fait, il n'est pas essentiel de trouver un vecteur propre com-

mun à A et B, c'est-à-dire une droite stable par A et B : n'importe quel sous-espace strict de $\mathbb{C}^n$ stable par A et B permet de faire une récurrence sur la dimension! Et on vient justement de montrer que si $\lambda \in \operatorname{Sp} A$, alors soit $\operatorname{Ker}(A - \lambda I_n)$ soit $\operatorname{Im}(A - \lambda I_n)$ est stable par B (et naturellement aussi par A). Comme A n'est pas scalaire (sinon $C = 0$), ces deux sous-espaces sont non nuls et strictement inclus dans $\mathbb{C}^n$.

Rédigeons alors la récurrence. Le résultat est trivial pour $n = 1$. Supposons le résultat vrai pour tout couple de matrices de taille $< n$ et prenons A, B dans $\mathcal{M}_n(\mathbb{C})$ telles que $\operatorname{rg}(AB - BA) \leqslant 1$. Si $AB = BA$ on sait que A et B sont cotrigonalisables. Si $C = AB - BA$ est de rang 1 on vient de prouver qu'il existe un sous-espace non trivial F de $\mathbb{C}^n$ stable par A et par B. En prenant une base de F que l'on prolonge en base de $\mathbb{C}^n$, les endomorphismes canoniquement associés à A et B admettent dans cette base de $\mathbb{C}^n$ des matrices de la forme

$$A' = \begin{pmatrix} \boxed{A_1} & \times \\ 0 & \boxed{A_2} \end{pmatrix} \text{ et } B' = \begin{pmatrix} \boxed{B_1} & \times \\ 0 & \boxed{B_2} \end{pmatrix} \text{ respectivement.}$$

On a alors

$$A'B' - B'A' = \begin{pmatrix} \boxed{A_1B_1 - B_1A_1} & \times \\ 0 & \boxed{A_2B_2 - B_2A_2} \end{pmatrix}.$$

Puisque $\operatorname{rg}(A'B' - B'A') = 1$, on a également $\operatorname{rg}(A_1B_1 - B_1A_1) \leqslant 1$ et $\operatorname{rg}(A_2B_2 - B_2A_2) \leqslant 1$, et par hypothèse de récurrence il existe P et Q inversibles telles que $P^{-1}A_1P$ et $P^{-1}B_1P$ d'une part, $Q^{-1}A_1Q$ et $Q^{-1}B_1Q$ d'autre part, soient triangulaires supérieures. Si R est la matrice diagonale par blocs $\operatorname{Diag}(P, Q)$, $R^{-1}A'R$ et $R^{-1}B'R$ sont triangulaires supérieures et donc A et B sont simultanément trigonalisables. ◁

Après tous ces exercices le lecteur se demandera légitimement s'il y a une condition nécessaire et suffisante pour que deux matrices A *et* B *de* $\mathcal{M}_n(\mathbb{C})$ *soient cotrigonalisables. On voit de suite que si* A, B *sont cotrigonalisables, la matrice* AB − BA *se trigonalise en une matrice dont la diagonale est nulle. Elle est donc nilpotente. En fait, il en est de même de toute matrice de la forme* $P(A, B)(AB - BA)$ *où* $P(X, Y)$ *désigne un polynôme en deux variables* X, Y *non commutatives : par exemple si* $P(X, Y) = XY^2X^3$ *on a* $P(A, B) = AB^2A^3$. *Et bien un théorème de Mc Coy affirme que la réciproque est vraie : si* $P(A, B)(AB - BA)$ *est nilpotente pour tout polynôme en deux variables non commutatives* P, *alors les matrices* A *et* B *sont cotrigonalisables. On en déduit directement le résultat de l'exercice 2.30 (cas où* AB − BA = 0*) mais aussi celui de l'exercice 2.31. En effet, si* C = AB−BA *commute avec* A *et* B *on a, pour tout* $k \geqslant 1$, $C^k = (AC^{k-1})B - B(AC^{k-1})$ *et on en déduit que* $\operatorname{Tr}(C^k) = 0$. *Cela implique classiquement que* C *est nilpotente (voir exercice 2.28). Pour tout polynôme* P *en deux variables non commutatives on a alors* $[P(A, B)C]^n = P(A, B)^nC^n = 0$ *car* C *commute avec* $P(A, B)$.

Voici un exercice délicat où on l'on cherche un vecteur propre commun à une famille commutative de matrices.

2.34. Vecteur propre commun à une famille de matrices

Soit K un corps commutatif et S une partie de $\mathcal{M}_n(\mathrm{K})$ telle que les éléments de S commutent deux à deux. On suppose de plus que les éléments de S possèdent un vecteur propre commun. Démontrer l'existence d'un vecteur propre commun aux éléments de ${}^t\mathrm{S}$.

(École normale supérieure)

▷ **Solution.**

Le résultat a déjà été évoqué lorsque K est $\mathbb{C}$ ou plus généralement un corps algébriquement clos (voir page 99).

Quitte à rajouter les matrices $\lambda \mathrm{I}_n$ pour $\lambda \in \mathrm{K}$, on peut supposer sans perte de généralité qu'elles appartiennent à S. On peut remarquer que les éléments du sous-espace engendré par S possèdent la même propriété que ceux de S : ils commutent deux à deux et possèdent un vecteur propre en commun. On peut donc supposer que S est un sous-espace vectoriel de $\mathcal{M}_n(\mathrm{K})$ et en prendre une base $\mathcal{B}$ dont le premier élément est I_n. Écrivons $\mathcal{B} = (\mathrm{I}_n, \mathrm{A}_1, \dots, \mathrm{A}_p)$. Nous savons qu'il existe $\mathrm{X}_0 \in \mathrm{K}^n$ non nul tel que pour tout $i \in [\![1, p]\!]$, il existe $\lambda_i \in \mathrm{K}$ vérifiant $\mathrm{A}_i\mathrm{X}_0 = \lambda_i \mathrm{X}_0$. En changeant A_i en $\mathrm{A}_i - \lambda_i \mathrm{I}_n$, la famille $\mathcal{B}$ reste une base de S et pour tout $i \in [\![1, n]\!]$, $\mathrm{A}_i\mathrm{X}_0 = 0$. Nous allons démontrer l'existence d'un vecteur non nul appartenant aux noyaux de chaque ${}^t\mathrm{A}_i$. Ce vecteur sera alors un vecteur propre commun à tous les éléments de ${}^t\mathrm{S}$.

Pour cela, opérons une récurrence sur l'entier n. Le résultat est trivial pour $n = 1$. Supposons $n \geqslant 2$ et le résultat vrai en taille $< n$.

Le polynôme caractéristique des matrices A_i peut se mettre sous la forme $\mathrm{X}^p\mathrm{Q}$, où $p \geqslant 1$ et $\mathrm{Q}(0) \neq 0$. Supposons que pour l'une des matrices A_i, disons A_1 pour fixer les idées, on ait $p < n$ et $\deg \mathrm{Q} \geqslant 1$. Le théorème de décomposition des noyaux permet d'écrire $\mathrm{K}^n = \operatorname{Ker} \mathrm{A}_1^p \oplus \operatorname{Ker} \mathrm{Q}(\mathrm{A}_1)$, ces deux sous-espaces n'étant pas réduits à $\{0\}$. Considérons $\mathcal{B}_1$ une base de $\operatorname{Ker} \mathrm{A}_1^p$, $\mathcal{B}_2$ une base de $\operatorname{Ker} \mathrm{Q}(\mathrm{A}_1)$ et P la matrice de passage de la base canonique à la base $(\mathcal{B}_1, \mathcal{B}_2)$. Comme les A_i commutent deux à deux, ils laissent stables les sous-espaces $\operatorname{Ker} \mathrm{A}_1^p$ et $\operatorname{Ker} \mathrm{Q}(\mathrm{A}_1)$. Pour tout $1 \leqslant i \leqslant p$, la matrice $\mathrm{P}^{-1}\mathrm{A}_i\mathrm{P}$ est de la forme $\begin{pmatrix} \mathrm{A}'_i & 0 \\ 0 & \mathrm{B}_i \end{pmatrix}$ avec A'_i (resp. B_i) la matrice de l'endomorphisme induit par A_i sur $\operatorname{Ker} \mathrm{A}_1^p$ (resp. sur $\operatorname{Ker} \mathrm{Q}(\mathrm{A}_1)$). Les matrices A'_i commutent deux à deux et possèdent un vecteur non nul appartenant à tous les noyaux $\operatorname{Ker} \mathrm{A}'_i$. Comme la taille de ces matrices est strictement inférieure à n, par hypothèse de récurrence, il existe Y_0 non nul tel que ${}^t\mathrm{A}'_i\mathrm{Y}_0 = 0$ (la seule valeur propre de A'_i est 0

car A'_i est nilpotente). Dans ces conditions, le vecteur $Y'_0 = \begin{pmatrix} Y_0 \\ 0 \end{pmatrix} \in K^n$ est non nul et ${}^t(P^{-1}A_iP)Y'_0 = 0$ pour tout i. Comme ${}^tP\,{}^tA_i\,{}^tP^{-1}Y'_0 = 0$, le vecteur non nul ${}^tP^{-1}Y'_0$ est dans le noyau de chaque ${}^tA_i$ et répond au problème.

Il nous reste à traiter le cas où pour chaque matrice A_i, le polynôme caractéristique est égal à X^n (ce qui correspond avec les notations utilisées plus haut à $p = n$ et $Q = 1$). Les matrices A_i sont donc nilpotentes. Ainsi, la famille $({}^tA_1, \ldots, {}^tA_p)$ est composée de matrices nilpotentes commutant deux à deux. L'existence d'un vecteur propre commun à toutes les matrices ${}^tA_i$ est assurée par le lemme suivant :

Lemme. *Soit* E *un* K*-espace vectoriel de dimension finie non nulle et* $(u_1, \ldots, u_p)$ *une famille d'endomorphismes nilpotents commutant deux à deux. Il existe alors* $x \in E$*, vecteur propre commun à tous les endomorphismes* u_i*.*

Démonstration. Procédons par récurrence sur l'entier p. Si $p = 1$, comme $E \neq \{0\}$, $\operatorname{Ker} u_1$ n'est pas réduit à $\{0\}$ et un vecteur non nul de $\operatorname{Ker} u_1$ convient.

Supposons $p \geqslant 2$ et le résultat vrai au rang $p-1$. Si $u_p = 0$, on applique l'hypothèse de récurrence à la famille $u_1, \ldots, u_{p-1}$. Le vecteur x trouvé convient aussi pour u_p. Si $u_p \neq 0$, $F = \operatorname{Ker} u_p$ est un sous-espace non réduit à $\{0\}$ laissé stable par les u_i. Pour $1 \leqslant i \leqslant p-1$, notons u'_i l'endomorphisme induit par u_i sur F. Les endomorphismes $u'_1, \ldots, u'_{p-1}$ commutent deux à deux et sont nilpotents, et donc redevables de l'hypothèse de récurrence : il existe $x \in F$ un vecteur propre commun à tous les u'_i. C'est donc un vecteur propre pour u_i avec $i \leqslant p-1$ et même pour u_p car $x \in F = \operatorname{Ker} u_p$. La preuve du lemme est achevée. ◇

Et cela termine l'exercice. ◁

Le théorème de Cayley-Hamilton nous assure que χ_u *est un polynôme annulateur de* u *lorsque* u *est un endomorphisme d'un espace de dimension finie. Si* χ_u *est scindé (ce qui est toujours le cas si le corps de base est* $\mathbb{C}$*), on peut écrire* $\chi_u = (X-\lambda_1)^{n_1} \ldots (X-\lambda_r)^{n_r}$ *où les* λ_i *sont les valeurs propres deux à deux distinctes de* u*. Le théorème de décomposition des noyaux nous donne*

$$E = \bigoplus_{i=1}^{r} \operatorname{Ker}(u - \lambda_i \operatorname{Id}_E)^{n_i}.$$

Le sous-espace $F_i = \operatorname{Ker}(u - \lambda_i \operatorname{Id}_F)^{n_i}$ *est le sous-espace caractéristique de* u *relatif à la valeur propre* λ_i*. Il contient le sous-espace propre relatif à* λ_i*. Notons que l'endomorphisme* u *est diagonalisable si, et seulement*

si, pour tout i,

$$\mathrm{Ker}(u-\lambda_i \mathrm{Id}_E)^{n_i} = \mathrm{Ker}(u-\lambda_i \mathrm{Id}_E).$$

On rappelle le résultat classique sur les noyaux itérés prouvés à l'exercice 6.14 du tome 1 d'algèbre : en dimension finie, la suite des noyaux $\mathrm{Ker}(u-\lambda_i \mathrm{Id}_E)^k$ *est strictement croissante jusqu'à un rang p, puis stationnaire. Ce sont ces idées que l'on retrouve dans les exercices qui suivent.*

2.35. Étude d'un opérateur d'un espace de suites (1)

Soit $\mathcal{B} = \left\{u \in \mathbb{C}^{\mathbb{Z}},\ u \text{ bornée}\right\}$ et $T : \mathcal{B} \to \mathcal{B}$ qui à $(u_n)_{n\in\mathbb{Z}}$ associe $(u_{n+1})_{n\in\mathbb{Z}}$. Déterminer les valeurs propres et les vecteurs propres de T. Que dire d'un sous-espace de dimension finie stable par T ?

(**École normale supérieure**)

▷ **Solution.**

• Soit λ une valeur propre de T et u un vecteur propre associé. Si $\lambda = 0$, u est nulle ce qui est exclu. Donc $\lambda \neq 0$. On vérifie par récurrence sur $n \in \mathbb{N}$ que $u_n = u_0\lambda^n$ et $u_{-n} = \dfrac{u_0}{\lambda^n}$. Finalement, $u_n = u_0\lambda^n$ pour tout $n \in \mathbb{Z}$. Comme u est bornée, nécessairement $|\lambda| = 1$.

Réciproquement, si $\lambda \in \mathbb{C}$ est de module 1, $(\lambda^n)_{n\in\mathbb{Z}}$ est dans $\mathcal{B}$ et vérifie $T(u) = \lambda u$: le spectre de T est le cercle unité S^1 et les espaces propres sont tous de dimension 1.

• Soit F un sous-espace de $\mathcal{B}$ stable par T, de dimension finie. On note $\widetilde{T}$ l'endomorphisme de F induit par T. Comme $\mathbb{C}$ est algébriquement clos, on peut écrire $\chi_{\widetilde{T}} = \prod_{i=1}^{r} (X-\lambda_i)^{n_i}$ avec les λ_i deux à deux distincts dans $\mathbb{C}$ et $n_i \in \mathbb{N}^*$. Les λ_i sont des valeurs propres de $\widetilde{T}$, donc de T : ce sont des complexes de module 1. D'après le théorème de Cayley-Hamilton, $\chi_{\widetilde{T}}(\widetilde{T}) = 0$ et en appliquant le théorème de décomposition des noyaux :

$$F = \bigoplus_{i=1}^{r} \mathrm{Ker}(\widetilde{T}-\lambda_i \mathrm{Id})^{n_i}.$$

Soit λ une des valeurs propres λ_i. On a $\mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})^2 = \mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})$. En effet, si $u = (u_n)_{n\in\mathbb{Z}} \in \mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})^2$, on a $u_{n+2} - 2\lambda u_{n+1} + \lambda^2 u_n = 0$ et on sait qu'il existe $(A,B) \in \mathbb{C}^2$ tel que $u_n = (An+B)\lambda^n$. Comme $|\lambda| = 1$ et u est bornée, nécessairement $A = 0$ et $u \in \mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})$, d'où $\mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})^2 \subset \mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})$. L'autre inclusion est triviale.

Il s'ensuit que la suite des noyaux itérés $\mathrm{Ker}(\widetilde{T}-\lambda \mathrm{Id})^n$ est stationnaire à partir de $n = 1$. On a en particulier $\mathrm{Ker}(\widetilde{T}-\lambda_i \mathrm{Id})^{n_i} = \mathrm{Ker}(\widetilde{T}-\lambda_i \mathrm{Id})$ pour tout i, et donc

$$\mathrm{F} = \bigoplus_{i=1}^{r} \mathrm{Ker}(\widetilde{\mathrm{T}} - \lambda_i \,\mathrm{Id}).$$

Ainsi $\widetilde{\mathrm{T}}$ est diagonalisable et les $(\lambda_i^n)_{n\in\mathbb{Z}}$ constituent une base de vecteurs propres de F pour $\widetilde{\mathrm{T}}$.

Réciproquement, si $\lambda_1, \ldots, \lambda_r$ sont r complexes distincts de module 1, le sous-espace engendré par les $(\lambda_i^n)_{n\in\mathbb{Z}}$ est stable par T et de dimension finie.

Conclusion. Les sous-espaces stables par T sont les sous-espaces engendrés par les $(\lambda_i^n)_{n\in\mathbb{Z}}$ avec $\lambda_1, \ldots, \lambda_r$ complexes distincts de module 1. ◁

2.36. Étude d'un opérateur d'un espace de suites (2)

On pose $\mathrm{E} = \mathbb{C}^{\mathbb{N}}$ et on définit l'opérateur $\mathrm{T} : \mathrm{E} \to \mathrm{E}$ qui à u associe $((n+1)u_{n+1})_{n\in\mathbb{N}}$.

1. Déterminer les éléments propres de T.

2. Pour $\lambda \in \mathbb{C}$, déterminer $\mathrm{Ker}(\mathrm{T} - \lambda \,\mathrm{Id_E})^2$.

3. Soit V un sous-espace de E de dimension finie et stable par T et $u \in \mathrm{V}$. Montrer que la série de terme général u_n est absolument convergente.

(École normale supérieure)

▷ **Solution.**

1. Si u est un vecteur propre pour la valeur propre 0, on a pour tout $n \geqslant 0$, $(n+1)u_{n+1} = 0$ et finalement $u_n = 0$ pour tout $n \geqslant 1$. Il est alors clair que $\mathrm{Ker}\,\mathrm{T} = \mathbb{C}(1, 0, 0, \ldots, 0\ldots)$: 0 est valeur propre de T et l'espace propre associé est une droite.

Supposons que maintenant $\lambda \in \mathbb{C}^*$ est une valeur propre de T et que u est un vecteur propre associé. Alors pour tout $n \in \mathbb{N}$, $u_{n+1} = \dfrac{\lambda}{n+1} u_n$ et, par une récurrence immédiate, $u_n = \dfrac{\lambda^n}{n!} u_0$. Réciproquement, si la suite u est proportionnelle à $\left(\dfrac{\lambda^n}{n!}\right)_{n\in\mathbb{N}}$, alors $\mathrm{T}(u) = \lambda u$. On en déduit que $\mathrm{Ker}(\mathrm{T} - \lambda\,\mathrm{Id}) = \mathbb{C}\left(\dfrac{\lambda^n}{n!}\right)_{n\in\mathbb{N}}$ et $\mathrm{Sp}\,\mathrm{T} = \mathbb{C}$.

2. Soit $u = (u_n)_{n\in\mathbb{N}}$, $v = (v_n)_{n\in\mathbb{N}} = \mathrm{T}(u)$ et $w = (w_n)_{n\in\mathbb{N}} = \mathrm{T}(v)$ dans E. Alors on a, pour $n \in \mathbb{N}$,

$$v_n = (n+1)u_{n+1} \quad \text{et} \quad w_n = (n+1)v_{n+1} = (n+1)(n+2)u_{n+2}.$$

Soit $\lambda \in \mathbb{C}^*$. Dire que $u \in \mathrm{Ker}(\mathrm{T} - \lambda\,\mathrm{Id})^2 = \mathrm{Ker}(\mathrm{T}^2 - 2\lambda\mathrm{T} + \lambda^2\,\mathrm{Id})$, c'est dire que, pour tout $n \in \mathbb{N}$,

$$w_n - 2\lambda v_n + \lambda^2 u_n = (n+2)(n+1)u_{n+2} - 2\lambda(n+1)u_{n+1} + \lambda^2 u_n = 0,$$

ou encore

$$\alpha_{n+2} - 2\lambda\alpha_{n+1} + \lambda^2\alpha_n = 0$$

avec $\alpha_n = n!u_n$. Comme $X^2 - 2\lambda X + \lambda^2 = (X-\lambda)^2$, $u \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^2$ signifie qu'il existe $(A,B) \in \mathbb{C}^2$ tel que, pour tout n, $\alpha_n = (An+B)\lambda^n$ ou encore $u_n = \dfrac{(An+B)\lambda^n}{n!}$: $\mathrm{Ker}(T - \lambda\,\mathrm{Id})^2$ est un plan.

Dire que $u \in \mathrm{Ker}\,T^2$, c'est dire que $w_n = (n+2)(n+1)u_{n+2} = 0$ pour tout $n \geqslant 0$, autrement dit que $u_n = 0$ pour $n \geqslant 2$. Le noyau $\mathrm{Ker}\,T^2$ est un plan de base $(1,0,0,\dots,0,\dots)$ et $(0,1,0,\dots,0,\dots)$.

3. Notons $\widetilde{T}$ la restriction de T à V. $\chi_{\widetilde{T}} = \prod_{i=1}^{r}(X-\lambda_i)^{n_i}$ le polynôme caractéristique de $\widetilde{T}$ avec les $\lambda_i \in \mathbb{C}$ deux à deux distincts et les $n_i \geqslant 1$. Le théorème de Cayley-Hamilton assure que $\chi_{\widetilde{T}}(\widetilde{T}) = 0$. Le théorème de décomposition des noyaux permet ensuite d'écrire

$$V = \bigoplus_{i=1}^{r} \mathrm{Ker}(\widetilde{T} - \lambda_i\,\mathrm{Id})^{n_i}.$$

Il suffit donc de démontrer que chaque élément de $\mathrm{Ker}(\widetilde{T} - \lambda_i\,\mathrm{Id})^{n_i}$ correspond au terme général d'une série absolument convergente. Pour cela, démontrons que si $u = (u_n)_{n\in\mathbb{N}} \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^p$ avec $\lambda \in \mathbb{C}$ et $p \geqslant 1$, alors $\sum u_n$ est absolument convergente. Pour $\lambda = 0$, $\mathrm{Ker}\,T^p$ correspond aux suites nulles à partir de $n = p$, c'est donc vérifié.

Supposons $\lambda \neq 0$. Nous allons prouver que pour $p \geqslant 1$, $\mathrm{Ker}(T - \lambda\,\mathrm{Id})^p$ est constitué des suites $u_n = \dfrac{P(n)\lambda^n}{n!}$ $(n \in \mathbb{N})$, où P est un polynôme quelconque de $\mathbb{C}_{p-1}[X]$.

Pour commencer montrons par récurrence sur $p \geqslant 1$ que si, pour tout $n \in \mathbb{N}$, $u_n = \dfrac{P(n)\lambda^n}{n!}$, avec $P \in \mathbb{C}_{p-1}[X]$, alors $u = (u_n)_{n\in\mathbb{N}}$ appartient à $\mathrm{Ker}(T - \lambda\,\mathrm{Id})^p$. Cela a été vu pour $p = 1$ et $p = 2$. Supposons la propriété vraie au rang $p-1$, avec $p \geqslant 3$. On note $v = (v_n)_{n\in\mathbb{N}} = T(u) - \lambda u$. On a $v_n = \dfrac{(P(n+1) - P(n))\lambda^{n+1}}{n!}$ pour tout $n \in \mathbb{N}$. Comme $\lambda(P(X+1) - P(X))$ a un degré inférieur ou égal à $p-2$, on a $v \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^{p-1}$ par hypothèse de récurrence et donc $u \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^p$.

Montrons maintenant, toujours par récurrence sur $p \geqslant 1$, que $\dim \mathrm{Ker}(T - \lambda\,\mathrm{Id})^p \leqslant p$. Cela a été vu pour $p = 1, 2$. On suppose que c'est vrai pour $p-1$, pour $p \geqslant 3$. On considère l'application linéaire, $f : u \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^p \longmapsto T(u) - \lambda u \in \mathrm{Ker}(T - \lambda\,\mathrm{Id})^{p-1}$. Son image est un sous-espace de $\mathrm{Ker}(T - \lambda\,\mathrm{Id})^{p-1}$ qui est par hypothèse de récurrence de dimension finie inférieure ou égale à $p-1$. En prenant une base de $\mathrm{Im}\,f$, $(\varepsilon_1,\dots,\varepsilon_q)$, et pour tout i, un antécédent e_i de ε_i par f, il est aisé de vérifier que $H = \mathrm{Vect}(e_1,\dots,e_q)$ est un supplémentaire du noyau isomorphe à $\mathrm{Im}\,f$. Comme $\mathrm{Ker}\,f = \mathrm{Ker}(T - \lambda\,\mathrm{Id})$ est une

droite, le sous-espace $\mathrm{Ker}(\mathrm{T}-\lambda\,\mathrm{Id})^p = \mathrm{Ker}\, f \supset \Pi$ est de dimension finie $q+1 \leqslant p-1+1 = p$.

Enfin, les suites $\left(\dfrac{n^k\lambda^n}{n!}\right)_{n\in\mathbb{N}}$ pour $k \in [\![0, p-1]\!]$ forment une famille libre de $\mathrm{Ker}(\mathrm{T}-\lambda\,\mathrm{Id})^p$: on en déduit que ces p suites forment une base de $\mathrm{Ker}(\mathrm{T}-\lambda\,\mathrm{Id})^p$. Comme $\dfrac{n^k\lambda^n}{n!} = O\left(\dfrac{1}{n^2}\right)$ quand n tend vers l'infini, le théorème de comparaison des séries à termes positifs assure que $\sum \dfrac{n^k\lambda^n}{n!}$ est absolument convergente. Il en va de même pour $\sum u$ lorsque $u \in \mathrm{Ker}(\mathrm{T}-\lambda\,\mathrm{Id})^p$. ◁

2.37. Étude d'un opérateur d'un espace de suites (3)

Soit $\mathrm{E} = \mathbb{C}^{\mathbb{N}^*}$ et $\mathrm{T} : \mathrm{E} \to \mathrm{E}$ qui à la suite $(u_n)_{n\geqslant 1}$ associe la suite $\left(\dfrac{u_1+\cdots+u_n}{n}\right)_{n\geqslant 1}$.

1. Déterminer les éléments propres de T.

2. Soit $\lambda \in \mathbb{C}$. Déterminer $\mathrm{Ker}(\mathrm{T}-\lambda\,\mathrm{Id})^2$.

3. Déterminer les sous-espaces de dimension finie stables par T.

(**École normale supérieure**)

▷ **Solution.**

1. Soit λ une valeur propre de T et $u = (u_n)_{n\geqslant 1}$ un vecteur propre associé à λ. Notons p le plus entier k tel que $u_k \neq 0$. Alors, on a

$$\lambda u_p = \frac{u_1+\cdots+u_p}{p} = \frac{u_p}{p}.$$

Il s'ensuit que $\lambda = \dfrac{1}{p}$. Dans ces conditions, pour $n > p$, on peut écrire

$$\frac{u_n}{p} = \frac{u_p+\cdots+u_n}{n} \quad \text{et} \quad u_n = \frac{p}{n-p}\left(u_p+\cdots+u_{n-1}\right).$$

Si on se donne une autre suite non nulle $v = (v_n)_{n\geqslant 1}$ de $\mathrm{Ker}\left(\mathrm{T}-\dfrac{1}{p}\,\mathrm{Id}\right)$, en notant q le plus petit entier k tel que $v_k \neq 0$, on a $\dfrac{1}{p}v_q = \dfrac{v_q}{q}$ et $q = p$. De plus, pour $n > p$, on a

$$v_n = \frac{p}{n-p}\left(v_p+\cdots+v_{n-1}\right).$$

Quitte à multiplier v par le scalaire $\dfrac{u_p}{v_p}$, on peut supposer $u_p = v_p$ et alors on a $u = v$; on en déduit que $\mathrm{Ker}\left(\mathrm{T}-\dfrac{1}{p}\,\mathrm{Id}\right) \subset \mathbb{C}u$.

Réciproquement, si on pose $u_0 = \cdots = u_{p-1} = 0$ et $u_p = 1$, il y a une unique suite $\mathrm{U}(p) = (u_n)_{n\geqslant 1}$ vérifiant la relation de récurrence ci-dessus. Essayons de trouver son terme général. On a $u_{p+1} = p$, $u_{p+2} = \dfrac{p(p+1)}{2}$ puis $u_{p+3} = \dfrac{p}{3}\left(1+p+\dfrac{p(p+1)}{2}\right) = \dfrac{p(p+1)(p+2)}{6}$. On devine alors que $u_{p+k} = \mathrm{C}^k_{p+k-1}$ pour tout $k \in \mathbb{N}$. Montrons le par récurrence forte sur k. C'est vrai pour les premiers indices. Si la formule est vraie jusqu'au rang k on a

$$\begin{aligned} u_{p+k+1} &= \frac{p}{k+1}(\mathrm{C}^0_{p-1} + \mathrm{C}^1_p + \cdots + \mathrm{C}^k_{p+k-1}) = \frac{p}{k+1}\,\mathrm{C}^k_{p+k} \\ &= \frac{p}{k+1}\,\frac{(p+k)!}{p!k!} = \frac{(p+k)!}{(p-1)!(k+1)!} = \mathrm{C}^{k+1}_{p+k} \end{aligned}$$

ce qui prouve la formule au rang $k+1$ (la seconde égalité qui donne la somme d'une diagonale descendante dans le triangle de Pascal est facile à démontrer). La suite $\mathrm{U}(p)$ ainsi définie appartient à $\mathrm{Ker}\left(\mathrm{T} - \dfrac{1}{p}\,\mathrm{Id}\right)$.

Conclusion. Le spectre de T est constitué des éléments de la suite $\left(\dfrac{1}{p}\right)_{p\geqslant 1}$ et le sous-espace propre $\mathrm{Ker}\left(\mathrm{T} - \dfrac{1}{p}\,\mathrm{Id}\right)$ est la droite engendrée par la suite $\mathrm{U}(p) = (\mathrm{C}^{n-1}_{n-p})_{n\geqslant 1}$ avec la convention que le coefficient binomial est nul lorsque $n - p < 0$.

2. Soit $p \geqslant 1$, $\lambda = \dfrac{1}{p}$ et $u \in \mathrm{Ker}\left(\mathrm{T} - \dfrac{1}{p}\,\mathrm{Id}\right)^2$ non nulle. On note $v = \left(\mathrm{T} - \dfrac{1}{p}\,\mathrm{Id}\right)^2(u) = \mathrm{T}^2(u) - 2\lambda\mathrm{T}(u) + \lambda^2 u$. Pour $n \geqslant 1$, on a

$$\begin{aligned} v_n &= \frac{1}{n}\left(\sum_{k=1}^{n}\frac{1}{k}\sum_{\ell=1}^{k} u_\ell\right) - 2\lambda\frac{1}{n}\left(\sum_{k=1}^{n} u_k\right) + \lambda^2 u_n, \\ &= \frac{1}{n}\left(\sum_{\ell=1}^{n}\sum_{k=\ell}^{n}\frac{u_\ell}{k}\right) - 2\lambda\frac{1}{n}\left(\sum_{k=1}^{n} u_k\right) + \lambda^2 u_n, \\ &= \frac{1}{n}\left(\sum_{\ell=1}^{n} u_\ell\left(\sum_{k=\ell}^{n}\frac{1}{k} - 2\lambda\right)\right) + \lambda^2 u_n = 0. \end{aligned}$$

Si on note q le plus petit entier k tel que $u_k \neq 0$, on a

$$v_q = 0 = \frac{1}{q}\left(\frac{1}{q} - 2\lambda\right) + \lambda^2 = \left(\frac{1}{q} - \lambda\right)^2 = \left(\frac{1}{q} - \frac{1}{p}\right)^2.$$

On en déduit que $q = p$. Ensuite pour $n \geqslant p+1$, le coefficient de u_n dans la relation

$$\frac{1}{n}\left(\sum_{\ell=1}^{n} u_\ell\left(\sum_{k=\ell}^{n}\frac{1}{k} - 2\lambda\right)\right) + \lambda^2 u_n = 0,$$

est $\left(\frac{1}{n} - \lambda\right)^2 \neq 0$, autrement dit, u_n est uniquement déterminé par la valeur de u_p. Comme dans la première question, on en déduit que $\mathrm{Ker}\left(\mathrm{T} - \frac{1}{p}\,\mathrm{Id}\right)^2$ est contenu dans une droite. On conclut donc que

$$\mathrm{Ker}\left(\mathrm{T} - \frac{1}{p}\,\mathrm{Id}\right)^2 = \mathrm{Ker}\left(\mathrm{T} - \frac{1}{p}\,\mathrm{Id}\right) = \mathrm{Vect}(\mathrm{U}(p)).$$

3. Soit F un sous-espace de E stable par T, de dimension finie. On note $\tilde{\mathrm{T}}$ l'endomorphisme de F induit par T. Comme $\mathbb{C}$ est algébriquement clos, on peut écrire $\chi_{\tilde{\mathrm{T}}} = \prod_{i=1}^{r} (\mathrm{X} - \lambda_i)^{n_i}$ avec les λ_i deux à deux distincts dans $\mathbb{C}$ et $n_i \in \mathbb{N}^*$. Les λ_i sont des valeurs propres de $\tilde{\mathrm{T}}$, donc de T. D'après le théorème de Cayley-Hamilton, $\chi_{\tilde{\mathrm{T}}}(\tilde{\mathrm{T}}) = 0$ et en appliquant le théorème de décomposition des noyaux :

$$\mathrm{F} = \bigoplus_{i=1}^{r} \mathrm{Ker}(\tilde{\mathrm{T}} - \lambda_i\,\mathrm{Id})^{n_i}.$$

Soit λ une des valeurs propres λ_i. On a $\mathrm{Ker}(\tilde{\mathrm{T}} - \lambda\,\mathrm{Id})^2 = \mathrm{Ker}(\tilde{\mathrm{T}} - \lambda\,\mathrm{Id})$ d'après la question précédente.

Il s'ensuit que la suite des noyaux itérés $\mathrm{Ker}(\tilde{\mathrm{T}} - \lambda\,\mathrm{Id})^n$ est stationnaire à partir de $n = 1$. En particulier, on a $\mathrm{Ker}(\tilde{\mathrm{T}} - \lambda_i\,\mathrm{Id})^{n_i} = \mathrm{Ker}(\tilde{\mathrm{T}} - \lambda_i\,\mathrm{Id})$ pour tout i et

$$\mathrm{F} = \bigoplus_{i=1}^{r} \mathrm{Ker}(\tilde{\mathrm{T}} - \lambda_i\,\mathrm{Id}),$$

ce qui traduit que $\tilde{\mathrm{T}}$ est diagonalisable et F possède une base de vecteurs propres du type $\mathrm{U}(p)$ avec $p \in \mathbb{N}^*$.

Réciproquement, pour toute famille finie $p_1 < p_2 < \cdots < p_r$ de $\mathbb{N}^*$, $\mathrm{Vect}(\mathrm{U}(p_1), \ldots, \mathrm{U}(p_r))$ est un sous-espace de dimension finie stable par T, ce qui achève la description demandée. ◁

Considérons un endomorphisme trigonalisable u d'un espace de dimension finie et $\chi_u = (\mathrm{X} - \lambda_1)^{n_1} \ldots (\mathrm{X} - \lambda_r)^{n_r}$, où les λ_i sont les valeurs propres deux à deux distinctes de u. Le théorème de décomposition des noyaux permet d'écrire E *comme somme des sous-espaces caractéristiques*

$$\mathrm{E} = \bigoplus_{i=1}^{r} \mathrm{Ker}(u - \lambda_i\,\mathrm{Id}_{\mathrm{E}})^{n_i}.$$

Comme $\mathrm{F}_i = \mathrm{Ker}(u - \lambda_i\,\mathrm{Id}_{\mathrm{E}})^{n_i}$ est le noyau d'un polynôme en u, ce sous-espace est stable par u. L'induit v_i de $u - \lambda_i\,\mathrm{Id}_{\mathrm{E}}$ à F_i est

un endomorphisme nilpotent, de polynôme caractéristique $X^{\dim F_i}$. *L'induit de* u *sur* F_i *admet comme polynôme caractéristique* $(X-\lambda_i)^{\dim F_i}$. *Comme* E *est somme directe des sous-espaces caractéristiques* F_i, *on a* $\chi_u = \prod_{i=1}^{r} (X-\lambda_i)^{\dim F_i}$ *ce qui impose par unicité de la décomposition en produit de facteurs irréductibles que* $n_i = \dim F_i$ *: l'ordre de* λ *comme racine du polynôme caractéristique est la dimension du sous-espace caractéristique relatif à* λ.

On peut également rechercher une interprétation de l'ordre de λ *comme racine du polynôme minimal. Il est classique de montrer qu'en dimension finie la suite de sous-espaces* $(\operatorname{Ker} u^k)_{k\geqslant 0}$ *est strictement croissante jusqu'à un certain rang* p *puis stationnaire. Le lecteur trouvera cela dans l'exercice 6.14 du premier tome d'algèbre. Dans l'énoncé suivant on constate que cet indice* p *n'est autre que la valuation du polynôme minimal de* u. *On en déduit que l'ordre d'une valeur propre comme racine du polynôme minimal est l'indice* r *à partir duquel la suite* $(\operatorname{Ker}(u-\lambda \operatorname{Id}_E)^k)_{k\geqslant 0}$ *devient stationnaire.*

2.38. Ordre d'une valeur propre dans le polynôme minimal

Soit E un K-espace vectoriel de dimension n, u un endomorphisme de E. On note P le polynôme minimal de u et p la valuation de P (autrement dit l'ordre de 0 comme racine de P).

1. On suppose $p = 0$. Que peut-on dire de u ?

2. On suppose $p = 1$. Montrer que $E = \operatorname{Ker} u \oplus \operatorname{Im} u$.

3. Dans le cas général montrer que $E = \operatorname{Ker} u^p \oplus \operatorname{Im} u^p$ et que p est le plus petit entier pour lequel on a cette propriété.

4. En déduire que si λ est une valeur propre de u, l'ordre de λ comme racine de P n'est autre que le plus petit entier k tel que $\operatorname{Ker}(u-\lambda \operatorname{Id}_E)^k = \operatorname{Ker}(u-\lambda \operatorname{Id}_E)^{k+1}$.

(École polytechnique)

▷ **Solution.**

1. Dire que 0 n'est pas racine de P revient à dire que 0 n'est pas valeur propre de u et donc que u est inversible.

2. Si $p = 1$, on écrit $P = XQ(X)$ avec $Q(0) \neq 0$. D'après le théorème de décomposition des noyaux on a $\operatorname{Ker} u \oplus \operatorname{Ker} Q(u) = E$. D'autre part, on a $P(u) = Q(u) \circ u = 0$ d'où il vient que $\operatorname{Im} u \subset \operatorname{Ker} Q(u)$. En vertu du théorème du rang, $\dim \operatorname{Im} u = \dim E - \dim \operatorname{Ker} u = \dim \operatorname{Ker} Q(u)$, d'après ce qui précède. On a donc l'égalité $\operatorname{Im} u = \operatorname{Ker} Q(u)$ et finalement

$$\boxed{E = \operatorname{Ker} u \oplus \operatorname{Im} u}.$$

3. On suppose dans toute cette question $p \geqslant 1$, c'est-à-dire u non inversible. On procède comme précédemment et on écrit $P = X^p Q(X)$

avec $Q(0) \neq 0$. Par le théorème de décomposition des noyaux on obtient $E = \operatorname{Ker} u^p \oplus \operatorname{Ker} Q(u)$. Comme $Q(u) \circ u^p = 0$, on a $\operatorname{Im} u^p \subset \operatorname{Ker} Q(u)$ et les deux espaces ayant la même dimension, ils sont égaux. On a donc

$$\boxed{E = \operatorname{Ker} u^p \oplus \operatorname{Im} u^p}.$$

Notons que si $q \geqslant p$, le polynôme $X^q Q(X)$ annule aussi u et on a $E = \operatorname{Ker} u^q \oplus \operatorname{Ker} Q(u)$. Comme $\operatorname{Ker} u^p \subset \operatorname{Ker} u^q$ on a forcément égalité (les deux espaces ont la même dimension). Ainsi, la suite $(\operatorname{Ker} u^k)_{k\geqslant 0}$ est constante à partir du rang p. Comme $\operatorname{Im} u^q \subset \operatorname{Im} u^p$ le théorème du rang montre que la suite $(\operatorname{Im} u^k)_{k\geqslant 0}$ est aussi constante à partir du rang p. En particulier, on a $E = \operatorname{Ker} u^q \oplus \operatorname{Im} u^q$ pour tout $q \geqslant p$. On va maintenant montrer que ce n'est plus le cas pour $q < p$. En effet supposons $E = \operatorname{Ker} u^q \oplus \operatorname{Im} u^q$. Les sous-espaces $\operatorname{Ker} u^q$ et $\operatorname{Im} u^q$ sont stables par u et si v désigne l'endomorphisme induit par u dans $\operatorname{Im} u^q$ il est facile de voir que le polynôme $X^q \chi_v(X)$ annule u. Comme v est inversible (car $\operatorname{Ker} u \subset \operatorname{Ker} u^q$), 0 n'est pas racine de χ_v et on a donc nécessairement $q \geqslant p$.

Conclusion. L'entier p est le plus petit entier naturel pour lequel on a $E = \operatorname{Ker} u^p \oplus \operatorname{Im} u^p$. C'est aussi l'entier à partir duquel les suites $(\operatorname{Ker} u^k)_{k\geqslant 1}$ et $(\operatorname{Im} u^k)_{k\geqslant 1}$ sont stationnaires car si $1 \leqslant q < p$, le polynôme $X^q Q(X)$ n'annule pas u (par définition du polynôme minimal) et on a donc $\operatorname{Ker} u^q \oplus \operatorname{Ker} Q(u) \neq E$. Ainsi, $\operatorname{Ker} u^q$ est strictement inclus dans $\operatorname{Ker} u^p$.

4. Il suffit d'appliquer ce qui précède à l'endomorphisme $v = u - \lambda \operatorname{Id}_E$ dont le polynôme minimal est $Q = P(X + \lambda)$. La valuation de Q n'est autre que l'ordre de λ comme racine de P et c'est à partir de cet entier que la suite $(\operatorname{Ker} v^k)_{k\geqslant 0} = (\operatorname{Ker}(u - \lambda \operatorname{Id}_E)^k)_{k\geqslant 0}$ devient stationnaire. ◁

Dans ces conditions, si on note n_λ l'ordre de λ comme racine de χ_u et p_λ son ordre comme racine de μ_u, comme $p_\lambda \leqslant n_\lambda$ en vertu du théorème de Cayley-Hamilton, on a $\operatorname{Ker}(u - \lambda \operatorname{Id}_E)^{n_\lambda} = \operatorname{Ker}(u - \lambda \operatorname{Id}_E)^{p_\lambda}$. Autrement dit, la suite $(\operatorname{Ker}(u - \lambda \operatorname{Id}_E)^k)_{k\geqslant 0}$ stationne à F_λ, où F_λ désigne le sous-espace caractéristique relatif à λ :

$$F_\lambda = \operatorname{Ker}(u - \lambda \operatorname{Id}_E)^{n_\lambda} = \operatorname{Ker}(u - \lambda \operatorname{Id}_E)^{p_\lambda} = \operatorname{Ker}(u - \lambda \operatorname{Id}_E)^r.$$

L'exercice suivant, qui est relativement difficile, met en scène les sous-espaces caractéristiques pour dégager des sous-espaces stables par une famille d'endomorphismes afin de se ramener à une dimension strictement inférieure.

2.39. Familles de matrices anticommutantes

Déterminer le cardinal maximum d'une famille de matrices de $\mathrm{GL}_n(\mathbb{C})$ qui anticommutent deux à deux.

(École normale supérieure)

▷ **Solution.**

On considérera qu'une famille de $\mathrm{GL}_n(\mathbb{C})$ réduite à un seul élément répond au problème. On notera $r_n \in \mathbb{N}^* \cup \{\infty\}$ le cardinal maximal en dimension n.

Il est clair que $r_n = 1$ si n est impair : en effet, si A et B sont deux matrices de $\mathcal{M}_n(\mathbb{C})$ telles que $AB = -BA$, il vient, en passant au déterminant,

$$\det A \det B = \det AB = \det(-AB) = (-1)^n \det A \det B = -\det A \det B,$$

d'où $\det A \det B = 0$ et A ou B n'est pas inversible.

Passons au cas n pair. Si A et B sont dans $\mathrm{GL}_n(\mathbb{C})$ et anticommutent, A est semblable à son opposée puisque $AB = -BA$ entraîne $A = B(-A)B^{-1}$. Comme

$$\chi_{-A} = \det(XI_n + A) = (-1)^n \det(-XI_n - A) = \chi_A(-X)$$

et puisque A est semblable à $-A$, $\chi_A = \chi_{-A} = \chi_A(-X)$ ce qui entraîne que si λ est valeur propre de A, alors $-\lambda$ est valeur propre avec la même multiplicité.

• Le candidat passant l'oral prouvera son bon sens scientifique en essayant de voir ce qu'il en est pour les petites dimensions. Étudions donc le cas $n = 2$.

Prenons une famille anticommutante $(A_1, \dots, A_r)$ de $\mathrm{GL}_2(\mathbb{C})$. On suppose $r \geqslant 2$. Soit $\lambda \in \mathrm{Sp}\, A_1$ ($\mathbb{C}$ est algébriquement clos). Alors $-\lambda \in \mathrm{Sp}\, A_1$. Comme λ est non nul, la matrice A_1 a deux valeurs propres distinctes et est donc diagonalisable. Soit $P \in \mathrm{GL}_2(\mathbb{C})$ tel que $P^{-1}A_1P = \begin{pmatrix} \lambda & 0 \\ 0 & -\lambda \end{pmatrix}$. Alors la famille des $P^{-1}A_kP$ est aussi anticommutante et on a donc construit une nouvelle famille de même cardinal avec la première matrice diagonale. On la notera encore $(A_1, \dots, A_r)$. Quitte à toutes les diviser par λ on peut supposer $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Écrivons $A_k = \begin{pmatrix} a_k & b_k \\ c_k & d_k \end{pmatrix}$ $(2 \leqslant k \leqslant r)$. On a alors

$$A_1A_k = \begin{pmatrix} a_k & b_k \\ -c_k & -d_k \end{pmatrix} = -A_kA_1 = \begin{pmatrix} -a_k & b_k \\ -c_k & d_k \end{pmatrix}$$

D'où $a_k = d_k = 0$ pour tout $k \geqslant 2$. Supposons $r \geqslant 4$. Les complexes b_k et c_k sont non nuls puisque les matrices A_k sont inversibles. On a

$$A_2A_3 = \begin{pmatrix} b_2c_3 & 0 \\ 0 & b_3c_2 \end{pmatrix} = -A_3A_2 = \begin{pmatrix} -b_3c_2 & 0 \\ 0 & -b_2c_3 \end{pmatrix}$$

D'où $b_2c_3 = -b_3c_2$ et $\dfrac{b_2}{c_2} = -\dfrac{b_3}{c_3}$. En écrivant $A_2A_4 = -A_4A_2$, il vient $\dfrac{b_2}{c_2} = -\dfrac{b_4}{c_4}$. De même avec $A_3A_4 = -A_4A_3$, $\dfrac{b_3}{c_3} = -\dfrac{b_4}{c_4}$. C'est impossible! On a donc nécessairement $r \leqslant 3$. En fait, on peut exhiber une famille anticommutante de trois matrices :

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Ainsi on vient de prouver que $r_2 = 3$.

• On va essayer de trouver une relation de récurrence entre r_n pour n pair et des valeurs r_k plus petites. Posons $n = 2p$ avec $p \in \mathbb{N}^*$. On *fait l'hypothèse* que le cardinal *maximum* d'une famille *anticommutante* de matrices de $\mathrm{GL}_p(\mathbb{C})$ est fini et vaut r_p. Soit $(A_1, \ldots, A_r)$ une famille anticommutante de $\mathrm{GL}_n(\mathbb{C})$ avec $r \geqslant 3$. Comme dans le cas $n = 2$, on va montrer qu'on peut se ramener à une famille de matrices diagonalisables. Comme on l'a vu plus haut pour $n = 2$, A_1 est semblable à $-A_1$. Soit λ une valeur propre de A_1. Nous allons noter $F_\lambda = \mathrm{Ker}(A_1 - \lambda I_n)^m$ et $F_{-\lambda} = \mathrm{Ker}(A_1 + \lambda I_n)^m$ les sous-espaces caractéristiques correspondants de A_1.

Lemme. *Pour $k \geqslant 2$, $A_k(F_\lambda) \subset F_{-\lambda}$ et $A_k(F_{-\lambda}) \subset F_\lambda$.*

Démonstration. Soit $X \in F_\lambda$, $A = A_1$ et $B = A_k$. On a

$$\begin{aligned}(A - \lambda I_n)^m BX &= (A - \lambda I_n)^{m-1} B(-A - \lambda I_n)X = \cdots = B(-A - \lambda I_n)^m X \\ &= (-1)^m B(A + \lambda I_n)^m X = (-1)^m B0 = 0.\end{aligned}$$

Par conséquent $BX \in F_{-\lambda}$. Cela montre la première inclusion. L'autre s'obtient en remplaçant λ par $-\lambda$. ♢

Lemme. *Pour $k \geqslant 2$, $E_\lambda = F_\lambda \oplus F_{-\lambda}$ est stable par A_k.*

Démonstration. Pour $k = 1$, c'est parce que les sous-espaces caractéristiques de A_1 sont stables par A_1.

La stabilité par A_k pour $k \geqslant 2$ résulte du lemme précédent. Le fait que F_λ et $F_{-\lambda}$ soient en somme directe est une conséquence du théorème de décomposition des noyaux puisque $(X-\lambda)^m$ et $(X+\lambda)^m$ sont premiers entre eux. ♢

Nous avons vu que χ_{A_1} pouvait s'écrire

$$\chi_{A_1} = (X-\lambda_1)^{n_1}(X+\lambda_1)^{n_1}(X-\lambda_2)^{n_2}(X+\lambda_2)^{n_2}\dots(X-\lambda_s)^{n_s}(X+\lambda_s)^{n_s}$$

où les λ_k et $-\lambda_k$ sont deux à deux distincts et les $n_k \in \mathbb{N}^*$. On obtient

$$\begin{aligned}\mathbb{C}^n &= F_{\lambda_1} \oplus F_{-\lambda_1} \oplus F_{\lambda_2} \oplus F_{-\lambda_2} \oplus \dots \oplus F_{\lambda_s} \oplus F_{-\lambda_s} \\ &= E_{\lambda_1} \oplus E_{\lambda_2} \oplus \dots \oplus E_{\lambda_s}.\end{aligned}$$

Nous identifions les matrices et les endomorphismes canoniquement associés. D'après les lemmes précédents, pour $1 \leqslant \ell \leqslant s$, E_{λ_ℓ} est stable par chaque A_k. Prenons $\mathcal{B}_\ell$ une base de F_{λ_ℓ} et $\mathcal{B}'_\ell$ une base de $F_{-\lambda_\ell}$. Pour $k \geqslant 2$, la matrice de A_k restreinte à E_{λ_ℓ} dans la base $(\mathcal{B}_\ell, \mathcal{B}'_\ell)$ est de la forme

$$\left(\begin{array}{c|c} 0 & B \\ \hline C & 0 \end{array}\right).$$

Puisque $\mathbb{C}^n$ est somme directe des E_{λ_ℓ}, il existe un unique endomorphisme u de $\mathbb{C}^n$ dont la restriction à chaque E_{λ_ℓ} admet comme matrice dans la base $(\mathcal{B}_\ell, \mathcal{B}'_\ell)$

$$\left(\begin{array}{c|c} I & 0 \\ \hline 0 & -I \end{array}\right).$$

On remarque que

$$\left(\begin{array}{c|c} 0 & B \\ \hline C & 0 \end{array}\right)\left(\begin{array}{c|c} I & 0 \\ \hline 0 & -I \end{array}\right) = \left(\begin{array}{c|c} 0 & -B \\ \hline C & 0 \end{array}\right) = -\left(\begin{array}{c|c} 0 & B \\ \hline C & 0 \end{array}\right)\left(\begin{array}{c|c} I & 0 \\ \hline 0 & -I \end{array}\right)$$

Autrement dit, pour $k \geqslant 2$, A_k et u anticommutent sur chaque E_{λ_ℓ} et donc sur $\mathbb{C}^n$ tout entier. Si on note D la matrice de u dans la base canonique, $(D, A_2, \dots, A_r)$ est une famille anticommutante de $\mathrm{GL}_n(\mathbb{C})$ de même cardinal dont la première matrice D est semblable à

$$\left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & -I_p \end{array}\right)$$

En remplaçant A_1 par D et les matrices par des matrices semblables, on peut supposer que $A_1 = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & -I_p \end{array}\right)$. Ainsi, s'il existe une famille anticommutante de cardinal $r \geqslant 2$, il en existe une que nous noterons encore $(A_1, A_2, \dots, A_r)$ dont la première matrice est $A_1 = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & -I_p \end{array}\right)$.

• Si on pose $E_1 = \mathrm{Vect}(e_1, \dots, e_p) = \mathrm{Ker}(A_1 - I_n) = \mathrm{Ker}(A_1 - I_n)^p$ et $E_2 = \mathrm{Vect}(e_{p+1}, \dots, e_n) = \mathrm{Ker}(A_1 + I_n) = \mathrm{Ker}(A_1 + I_n)^p$, on a d'après le premier lemme $A_k(E_1) \subset E_2$ et $A_k(E_2) \subset E_1$ pour $2 \leqslant k \leqslant r$. Il existe donc B_k et C_k dans $\mathrm{GL}_p(\mathbb{C})$ tels que

$$A_k = \left(\begin{array}{c|c} 0 & B_k \\ \hline C_k & 0 \end{array}\right)$$

Pour $2 \leqslant k, \ell \leqslant r$ distincts, on a

$$A_k A_\ell = \left(\begin{array}{c|c} B_k C_\ell & 0 \\ \hline 0 & C_k B_\ell \end{array}\right) = -A_\ell A_k = \left(\begin{array}{c|c} -B_\ell C_k & 0 \\ \hline 0 & -C_\ell B_k \end{array}\right)$$

On en déduit que $\begin{cases} B_k C_\ell = -B_\ell C_k \\ C_k B_\ell = -C_\ell B_k \end{cases}$. Comme les A_k sont inversibles, les B_k et les C_k sont inversibles. Posons, pour $k \geqslant 3$, $B = B_2$, $C = C_2$ et $D_k = B_k B^{-1}$.

Lemme. *$D_3, \ldots, D_r$ anticommutent.*

Démonstration. Pour $3 \leqslant k, \ell \leqslant r$, distincts, on a $B_k C = -B C_k$ ou $B^{-1} B_k = -C_k C^{-1}$. Par conséquent,

$$\begin{aligned} D_k D_\ell &= B_k B^{-1} B_\ell B^{-1} = B_k (B^{-1} B_\ell) B^{-1} \\ &= -B_k (C_\ell C^{-1}) B^{-1} = -(B_k C_\ell) C^{-1} B^{-1} \\ &= B_\ell C_k C^{-1} B^{-1} = B_\ell (C_k C^{-1}) B^{-1} \\ &= -B_\ell B^{-1} B_k B^{-1} = -D_\ell D_k \end{aligned}$$

Par hypothèse, le cardinal d'une famille anticommutante de matrices inversibles de taille p est inférieur ou égal à r_p. On a donc $r - 2 \leqslant r_p$ et $r \leqslant r_p + 2$. On en déduit que r_n est fini et $r_n \leqslant r_p + 2$.

• Nous allons montrer qu'en fait $r_n = r_p + 2$. Par hypothèse, il existe une famille anticommutante de $\mathrm{GL}_p(\mathbb{C})$ de cardinal r_p. On la note $D_3, \ldots, D_r$ avec $r = r_p + 2$. On pose

$$A_1 = \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & -I_p \end{array}\right), \quad A_2 = \left(\begin{array}{c|c} 0 & I_p \\ \hline -I_p & 0 \end{array}\right) \quad \text{et} \quad A_k = \left(\begin{array}{c|c} 0 & D_k \\ \hline D_k & 0 \end{array}\right),$$

pour $3 \leqslant k \leqslant r$. On vérifie aisément que A_1 anticommute avec $A_2, \ldots, A_r$ et que A_2 anticommute avec tous les A_k pour $k \geqslant 3$. Enfin, pour $k, \ell \geqslant 3$, distincts

$$A_k A_\ell = \left(\begin{array}{c|c} D_k D_\ell & 0 \\ \hline 0 & D_k D_\ell \end{array}\right) = -\left(\begin{array}{c|c} D_\ell D_k & 0 \\ \hline 0 & D_\ell D_k \end{array}\right) = -A_\ell A_k.$$

On a donc construit une famille anticommutante de cardinal $r = r_p + 2$. Le cardinal maximal des familles anticommutantes de $\mathrm{GL}_n(\mathbb{C})$ existe et vaut donc $r_p + 2$.

• Pour terminer, montrons par récurrence sur n que le cardinal maximum des familles anticommutantes de $\mathrm{GL}_n(\mathbb{C})$ est fini et vaut $r_n = 2\nu_2(n) + 1$ où $\nu_2(n)$ est l'exposant de 2 dans la décomposition de n en produit de facteurs premiers.

∗ C'est vu si n est impair ou $n = 2$ (la formule convient bien).

∗ Si $n \geqslant 3$, n pair, on pose $p = \dfrac{n}{2}$. Par hypothèse de récurrence, le cardinal maximum des familles anticommutantes de $\mathrm{GL}_p(\mathbb{C})$ est $2\nu_2(p)+1$. D'après ce qui précède, le cardinal maximum des familles anticommutantes de $\mathrm{GL}_n(\mathbb{C})$ est $2\nu_2(p)+1+2 = 2(\nu_2(p)+1)+1 = 2\nu_2(n)+1$.

Conclusion. Le cardinal maximal des familles de matrices de $\mathrm{GL}_n(\mathbb{C})$ qui anticommutent existe et vaut $\boxed{r_n = 2\nu_2(n)+1}$. ◁

L'exercice suivant fait démontrer l'importante décomposition de Dunford.

2.40. Décomposition de Dunford

Soit E un $\mathbb{C}$-espace vectoriel de dimension $n \geqslant 1$ et u un endomorphisme de E. Montrer l'existence d'un unique couple (d, n) d'endomorphismes de E tel que

(i) $u = d + n$,

(ii) d et n commutent,

(iii) d est diagonalisable et u est nilpotent.

Vérifier en outre que d et n sont des polynômes en u.

(École polytechnique)

▷ **Solution.**

• Traitons d'abord l'existence du couple (d, n). Écrivons

$$\chi_u = (\mathrm{X}-\lambda_1)^{m_1} \dots (\mathrm{X}-\lambda_r)^{m_r},$$

avec les $\lambda_i \in \mathbb{C}$ deux à deux distincts et les entiers $m_i \geqslant 1$. Par le théorème de décomposition des noyaux et le théorème de Cayley-Hamilton, E s'écrit comme somme directe des sous-espaces caractéristiques $\mathrm{F}_i = \mathrm{Ker}(u-\lambda_i \mathrm{Id})^{m_i}$:

$$\mathrm{E} = \bigoplus_{i=1}^{r} \mathrm{F}_i.$$

Soit d l'endomorphisme de E dont la restriction à chaque F_i est $\lambda_i \mathrm{Id}_{\mathrm{F}_i}$: autrement dit, d est diagonalisable et admet chaque sous-espace F_i comme espace propre pour la valeur propre λ_i. Posons $n = u - d$. Il ne reste plus qu'à vérifier que n est nilpotent et commute avec d. Comme chaque sous-espace caractéristique de E est stable par u et par d (donc aussi par n), il suffit de le vérifier pour les restrictions aux F_i. Notons avec un indice i les restrictions à F_i. On a $u_i = d_i + n_i$ et $d_i = \lambda_i \mathrm{Id}_{\mathrm{F}_i}$. Par définition de F_i, $(u_i - \lambda_i \mathrm{Id}_{\mathrm{F}_i})^{m_i} = 0$ donc n_i est nilpotent et commute clairement avec l'homothétie d_i. D'où le résultat. Le couple (d, n) convient.

Avant de prouver l'unicité, vérifions que d et n sont des polynômes en u. Cela se voit dans la démonstration du théorème de décomposition des noyaux : pour tout i la projection π_i sur F_i parallèlement à la somme des autres sous-espaces caractéristiques est un polynôme en u. Or, par construction, on a simplement pris

$$d = \sum_{i=1}^{r} \lambda_i \pi_i \in \mathbb{C}[u].$$

Bien entendu $n = u - d$ est alors aussi dans $\mathbb{C}[u]$.

• Supposons l'existence d'un autre couple (d', n') répondant au problème. On a alors $d' - d = n - n'$. Comme d' commute avec n', il commute aussi avec u, donc avec tout polynôme en u. En particulier d' commute avec d. Ainsi d et d' sont codiagonalisables (cf. exercice 2.21) et donc $d' - d$ est diagonalisable. De même n commute avec n'. Il en découle que $n - n'$ est nilpotent. Le seul endomorphisme diagonalisable et nilpotent étant 0 on a $d = d'$ et $n = n'$. ◁

Cette décomposition peut être effectuée sur un corps quelconque K *dès lors que le polynôme caractéristique de u est scindé sur* K *i.e. dès lors que u est trigonalisable. C'est une condition nécessaire puisque si $u = d + n$, d diagonalisable, n nilpotente, $nd = dn$, alors d et n sont cotrigonalisables (cf. exercice 2.30), donc u est également trigonalisable.*

À partir de la décomposition de Dunford, on est fondamentalement ramené à l'étude des classes de similitude des matrices nilpotentes. Les exercices qui suivent sont consacrés aux éléments nilpotents.

2.41. Matrices nilpotentes (1)

Soient A, B dans $\mathcal{M}_n(\mathbb{C})$ et $\lambda_1, \ldots, \lambda_{n+1}$ des nombres complexes deux à deux distincts. On suppose que $A + \lambda_i B$ est nilpotente pour tout i. Montrer que A et B sont nilpotentes.

(**École polytechnique**)

▷ **Solution.**

Considérons la matrice $M = (A + XB)^n$ à coefficients dans $\mathbb{C}[X]$. Chaque coefficient de cette matrice est un polynôme de degré inférieur ou égal à n qui s'annule en $n + 1$ valeurs distinctes : chaque coefficient est donc nul. Si on développe M selon les puissances de X, il vient

$$M = A^n + \cdots + X^n B^n.$$

Chaque coefficient matriciel est nécessairement nul et on a en particulier $A^n = B^n = 0$. ◁

2.42. Matrices nilpotentes (2)

Soient A, B dans $\mathcal{M}_n(\mathbb{R})$. On suppose A nilpotente et $B = AP(A)$ où $P \in \mathbb{R}[X]$ vérifie $P(0) = 1$. Montrer qu'il existe $Q \in \mathbb{R}[X]$ tel que $Q(0) = 1$ et $A = BQ(B)$.

(École polytechnique)

▷ **Solution.**

Notons $p \geqslant 1$ l'indice de nilpotence de A. Le polynôme minimal de A est donc X^p. On a $B^p = A^p(P(A))^p = 0$ donc B est aussi nilpotente. Comme $B^{p-1} = A^{p-1}(P(A))^{p-1} \neq 0$ (le polynôme $X^{p-1}P(X)^{p-1}$ n'est pas divisible par X^p) l'indice de nilpotence de B est aussi p. On sait alors que l'algèbre $\mathbb{R}[B]$ est de dimension p avec pour base $(I_n, B, \ldots, B^{p-1})$ et de même en remplaçant B par A. Par hypothèse on a $B \in \mathbb{R}[A]$ donc $\mathbb{R}[B] \subset \mathbb{R}[A]$ et finalement $\mathbb{R}[B] = \mathbb{R}[A]$ par égalité des dimensions. Comme $B^k \in \mathrm{Vect}(A^k, A^{k+1}, \ldots, A^{p-1})$ pour tout k on a, toujours par égalité des dimensions, $\mathrm{Vect}(A^2, A^3, \ldots, A^{p-1}) = \mathrm{Vect}(B^2, B^3, \ldots, B^{p-1})$. L'hypothèse $B = AP(A)$ avec $P(0) = 1$ signifie exactement que $B - A \in \mathrm{Vect}(A^2, A^3, \ldots, A^{p-1})$. On a donc $A - B \in \mathrm{Vect}(B^2, B^3, \ldots, B^{p-1})$ et le résultat demandé. ◁

Le thème suivant porte sur l'application adjointe d'un endomorphisme (ou d'une matrice). Si u, v sont des endomorphismes d'un espace vectoriel E, on définit le crochet de Lie de u et v par $[u, v] = u \circ v - v \circ u$. L'application adjointe de f est l'endomorphisme de $\mathcal{L}(E)$ défini par $u \longmapsto [f, u]$. On va constater que l'application adjointe hérite du caractère diagonalisable ou nilpotent de f le cas échéant.

2.43. Application adjointe (1)

Soit E un $\mathbb{C}$-espace vectoriel de dimension n. $f \in \mathcal{L}(E)$ et A_f l'application linéaire $g \in \mathcal{L}(E) \mapsto f \circ g - g \circ f$.

1. Montrer que si f est nilpotente alors A_f est nilpotente.

2. Montrer que si f est diagonalisable alors A_f est diagonalisable. Préciser les valeurs propres de A_f en fonction de celles de f.

3. Montrer réciproquement que si A_f est diagonalisable alors f est diagonalisable.

(École normale supérieure)

▷ **Solution.**

1. On note $u : g \mapsto f \circ g$ et $v : g \longmapsto g \circ f$. Alors $A_f = u - v$ et u et v commutent. D'après la formule du binôme dans $\mathcal{L}(\mathcal{L}(E))$, on a pour $m \geqslant 0$,

$$\mathrm{A}_f^m = \sum_{k=0}^{m} (-1)^k \mathrm{C}_m^k u^{m-k} v^k,$$

ce qui donne en appliquant en g,

$$\Lambda_f^m(g) = \sum_{k=0}^{m} (-1)^k \mathrm{C}_m^k f^{m-k} \circ g \circ f^k.$$

Si f est nilpotente, $f^n = 0$ et $2n-k$ ou k est supérieur ou égal à n pour $0 \leqslant k \leqslant 2n$. Donc $\mathrm{A}_f^{2n} = 0$ et Λ_f est nilpotente.

2. Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base de vecteurs propres pour f, supposée diagonalisable. On écrit $f(e_i) = \lambda_i e_i$ pour tout $1 \leqslant i \leqslant n$. On considère la base canonique $(\mathrm{E}_{ij})_{1 \leqslant i,j \leqslant n}$ de $\mathcal{M}_n(\mathbb{C})$. On appelle u_{ij} l'endomorphisme de E dont la matrice est E_{ij} dans la base $\mathcal{B}$ $(1 \leqslant i, j \leqslant n)$. Si D désigne la matrice $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, on a

$$\mathrm{DE}_{ij} - \mathrm{E}_{ij}\mathrm{D} = (\lambda_i - \lambda_j)\mathrm{E}_{ij},$$

si bien que $\mathrm{A}_f(u_{ij}) = (\lambda_i - \lambda_j)u_{ij}$. On en déduit que A_f est diagonalisable et que les u_{ij} constituent une base de vecteurs propres associés aux valeurs propres $\lambda_i - \lambda_j$.

3. Nous allons utiliser la décomposition de Dunford (on se reportera à l'exercice 2.40). On écrit $f = d + n$ avec d et n dans $\mathcal{L}(\mathrm{E})$, d diagonalisable, n nilpotent et $d \circ n = n \circ d$. On peut écrire $\mathrm{A}_f = \mathrm{A}_d + \mathrm{A}_n$. D'après les questions précédentes, A_d est diagonalisable et A_n est nilpotente. De plus, si $g \in \mathcal{L}(\mathrm{E})$, on a

$$\begin{aligned} \mathrm{A}_d \circ \mathrm{A}_n(g) &= \mathrm{A}_d(ng - gn) = d(ng - gn) - (ng - gn)d \\ &= dng - dgn - ngd + gnd \\ \text{et} \quad \mathrm{A}_n \circ \mathrm{A}_d(g) &= \mathrm{A}_n(dg - dg) = n(dg - gd) - (dg - gd)n \\ &= ndg - ngd - dgn + gdn. \end{aligned}$$

Comme $dn = nd$, on a $\mathrm{A}_d \circ \mathrm{A}_n(g) = \mathrm{A}_n \circ \mathrm{A}_d(g)$. On en déduit que $\Lambda_f = \mathrm{A}_d + \mathrm{A}_n$ est la décomposition de Dunford de A_f.

Si on suppose A_f diagonalisable, par unicité de la décomposition, A_n est nulle, autrement dit n est dans le centre de $\mathcal{L}(\mathrm{E})$. Il est classique de vérifier que n est alors une homothétie (en écrivant que n commute avec un projecteur sur $\mathbb{C}x$ avec $x \neq 0$, on constate que $n(x) \in \mathbb{C}x$). Comme n est nilpotente, n est nulle et $f = d$ est diagonalisable. ◁

Dans l'exercice suivant, on s'intéresse encore à l'application adjointe, notamment pour approfondir le cas où elle est nilpotente. La solution donnée est plus élémentaire puisqu'elle ne fait pas appel à la décomposition de Dunford.

2.44. Application adjointe (2)

Soit V un espace vectoriel de dimension finie sur $\mathbb{C}$. Pour a, b dans $\mathcal{L}(V)$, on pose $\varphi_a(b) = ab - ba$.

1. Calculer $\varphi_a^n(b)$ pour a, b dans $\mathcal{L}(V)$ et $n \in \mathbb{N}$.

2. Montrer que $a^{n+1}b - ba^{n+1} = \sum_{k=0}^{n} a^k(ab - ba)a^{n-k}$.

3. Montrer que si 0 est valeur propre de $a \in \mathcal{L}(V)$, alors a est nilpotent si et seulement si φ_a est nilpotent.

4. Montrer que si $a \in \mathcal{L}(V)$ n'a qu'une valeur propre, alors φ_a est nilpotent. Étudier la réciproque.

(École polytechnique)

▷ **Solution.**

1. L'application φ_a se décompose sous la forme $\varphi_a = g - d$ où g (resp. d) est la composition à gauche (resp. à droite) par a. Comme g et d commutent, on peut appliquer la formule du binôme de Newton dans l'algèbre $\mathcal{L}(\mathcal{L}(V))$. Il vient

$$\varphi_a^n = \sum_{k=0}^{n} C_n^k g^{n-k}(-1)^k d^k.$$

Autrement dit, pour $b \in \mathcal{L}(V)$, $\varphi_a^n(b) = \sum_{k=0}^{n} (-1)^k C_n^k a^{n-k} b a^k$.

2. Pour a et b dans $\mathcal{L}(V)$, on peut écrire

$$\begin{aligned}\sum_{k=0}^{n} a^k(ab - ba)a^{n-k} &= \sum_{k=0}^{n} a^{k+1}ba^{n-k} - \sum_{k=0}^{n} a^k b a^{n+1-k} \\ &= \sum_{k=1}^{n+1} a^k b a^{n+1-k} - \sum_{k=0}^{n} a^k b a^{n+1-k} \\ &= a^{n+1}b - ba^{n+1}.\end{aligned}$$

3. • Supposons que a est nilpotent et prenons $n \geqslant 0$ tel que $a^n = 0$. Si $k \in [\![0, 2n]\!]$, $2n - k$ ou k est supérieur ou égal à n, si bien que pour tout $b \in \mathcal{L}(V)$,

$$\varphi_a^{2n}(b) = \sum_{k=0}^{2n} C_{2n}^k (-1)^k a^{2n-k} b a^k = 0.$$

Donc φ_a est nilpotent.

• Réciproquement, on suppose que 0 est valeur propre de a et que a n'est pas nilpotent. Nous allons montrer que φ_a n'est pas nilpotent en

construisant un vecteur propre de φ_a correspondant à une valeur propre non nulle.

Comme a n'est pas nilpotent, il existe $\lambda \in \mathrm{Sp}(a) \setminus \{0\}$. On a donc $\mathrm{Ker}(a-\lambda \mathrm{Id}_V) \neq \{0\}$. D'autre part, 0 est valeur propre de a, donc a n'est pas un isomorphisme et $\mathrm{Im}\, a \neq V$. Soit $F \neq \{0\}$ un supplémentaire de $\mathrm{Im}\, a$. On peut construire un endomorphisme b nul sur $\mathrm{Im}\, a$ et qui envoie F sur une droite de $\mathrm{Ker}(a-\lambda \mathrm{Id})$ (avec $b_{|F} \neq 0$). Par construction, l'image de b est contenue dans $\mathrm{Ker}(a-\lambda \mathrm{Id})$. Si $x \in V$, on a

$$(ab-ba)(x) = ab(x) - 0 = \lambda b(x), \text{ et donc } \varphi_a(b) = \lambda b.$$

4. Si λ est l'unique valeur propre de a, on constate que $\varphi_{a-\lambda \mathrm{Id}} = \varphi_a$ et $a - \lambda \mathrm{Id}$ est nilpotente. Donc φ_a est nilpotente d'après la question précédente.

Réciproquement, supposons φ_a nilpotent. Soit λ une valeur propre de a. Alors $\varphi_{a-\lambda \mathrm{Id}} = \varphi_a$ est nilpotent et 0 est valeur propre de $a - \lambda \mathrm{Id}$. D'après la question précédente, $a - \lambda \mathrm{Id}$ est nilpotente et finalement, a n'a qu'une valeur propre. La réciproque est donc vraie. ◁

2.45. Application adjointe (3)

Si $(X, Y) \in (\mathcal{M}_n(\mathbb{R}))^2$, on note $\langle X, Y\rangle = \mathrm{Tr}({}^tXY)$.

1. Vérifier que $\langle\ ,\ \rangle$ est un produit scalaire sur $\mathcal{M}_n(\mathbb{R})$.

On considère pour $A \in \mathcal{M}_n(\mathbb{R})$ l'endomorphisme de $\mathcal{M}_n(\mathbb{R})$ défini par $\mathrm{ad}_A : X \longmapsto AX - XA$.

2. Calculer l'adjoint de ad_A.

3. Montrer que A est nilpotente si et seulement si $A \in \mathrm{Im}\, \mathrm{ad}_A$.

4. Montrer que A est nilpotente si et seulement si A est semblable à 2A.

(École normale supérieure)

▷ **Solution.**

1. Si $X = (x_{ij})_{1 \leqslant i,j \leqslant n}$ et $Y = (y_{ij})_{1 \leqslant i,j \leqslant n}$, le calcul de $\mathrm{Tr}({}^tXY)$ donne

$$\langle X, Y\rangle = \sum_{1 \leqslant i,j \leqslant n} x_{ij} y_{ij}.$$

et on reconnaît le produit scalaire canonique de $\mathcal{M}_n(\mathbb{R})$.

2. Soit $(X, Y) \in (\mathcal{M}_n(\mathbb{R}))^2$. On a

$$\begin{aligned}\langle \mathrm{ad}_A(X), Y\rangle &= \mathrm{Tr}({}^t(AX-XA)Y) = \mathrm{Tr}({}^tX\,{}^tAY) - \mathrm{Tr}({}^tA\,{}^tXY)\\ &= \mathrm{Tr}({}^tX\,{}^tAY) - \mathrm{Tr}({}^tXY\,{}^tA) = \mathrm{Tr}({}^tX(\mathrm{ad}_{{}^tA}(Y))\\ &= \langle X, \mathrm{ad}_{{}^tA}(Y)\rangle.\end{aligned}$$

On en déduit que $(\mathrm{ad}_A)^* = \mathrm{ad}_{{}^tA}$.

3. • Supposons que $A \in \mathrm{Im}\,\mathrm{ad}_A$. Alors il existe $X \in \mathcal{M}_n(\mathbb{R})$ tel que $A = AX - XA$. On peut alors montrer par récurrence sur $k \geqslant 1$ que $kA^k = A^kX - XA^k$: c'est vu pour $k = 1$. Si $k \geqslant 2$, par hypothèse de récurrence,

$$(k-1)A^{k-1} = A^{k-1}X - XA^{k-1},$$

ce qui donne en multipliant par A à gauche,

$$(k-1)A^k = A^kX - AXA^{k-1}.$$

En multipliant $A = AX - XA$ par A^{k-1} à droite, on obtient

$$A^k = AXA^{k-1} - XA^k.$$

En sommant avec ce qui précède, on obtient bien

$$kA^k = A^kX - XA^k.$$

On en déduit que pour tout $k \geqslant 1$, la trace de A^k est nulle, ce qui implique que A est nilpotente (cf. exercice 2.28).

• Supposons A nilpotente. Montrons que

$$A \in \mathrm{Im}\,\mathrm{ad}_A = (\mathrm{Ker}(\mathrm{ad}_A)^*)^\perp = (\mathrm{Ker}\,\mathrm{ad}_A)^\perp.$$

Il s'agit de montrer que pour tout $C \in \mathrm{Ker}\,\mathrm{ad}_A$ on a $\langle A, C\rangle = 0$. Soit $C \in \mathrm{Ker}\,\mathrm{ad}_A$. On a alors ${}^tAC = C\,{}^tA$. Dans ces conditions, la matrice ${}^tAC$ est nilpotente et donc sa trace est nulle : $\mathrm{Tr}({}^tAC) = \langle A, C\rangle = 0$. On conclut que $A \in \mathrm{Im}\,\mathrm{ad}_A$.

4. • Supposons A semblable à 2A. Prenons λ une racine de χ_A. Alors 2λ est une valeur propre de 2A et donc de A. Par une récurrence immédiate, les $2^k\lambda$ sont valeurs propres de A. Comme A n'a qu'un nombre fini de valeurs propres, $\lambda = 0$. Donc, $\chi_A = X^n$ et A est nilpotente.

• Supposons réciproquement A nilpotente. D'après la question précédente, on peut écrire $A = AX - XA$ avec $X \in \mathcal{M}_n(\mathbb{R})$. On a alors

$$AX = (I_n + X)A,$$

et pour $k \geqslant 0$, $AX^k = (I_n + X)^kA$. Soit $a \in \mathbb{R}$. En multipliant par $\dfrac{a^k}{k!}$, il vient

$$A\frac{(aX)^k}{k!} = \frac{(aI_n + aX)^k}{k!}A,$$

et en sommant sur k, on obtient

$$Ae^{aX} = A\left(\sum_{k=0}^{+\infty}\frac{(aX)^k}{k!}\right) = \left(\sum_{k=0}^{+\infty}\frac{(aI + aX)^k}{k!}\right)A = e^{aI_n+aX}A.$$

On en déduit $A = e^{aX}(e^aA)e^{-aX}$. En prenant $a = \ln 2$, on démontre que A et 2A sont semblables. ◁

Les énoncés qui suivent concernent des questions de sous-espaces stables. Nous avons déjà largement utilisé le fait que $\operatorname{Ker} P(u)$ *est stable par* u *pour tout* $P \in K[u]$.

2.46. Sous-espaces stables par un endomorphisme nilpotent (1)

Soit E un K-espace vectoriel de dimension n et u un endomorphisme nilpotent de E de rang $n-1$. Montrer que E admet exactement $n+1$ sous-espaces vectoriels stables par u, qui sont les $\operatorname{Ker} u^k$ pour $k \in [\![0, n]\!]$.

(École polytechnique)

▷ **Solution.**

Bien entendu les sous-espaces $\operatorname{Ker} u^k$ pour $k \in [\![0, n]\!]$ sont tous stables par u. On va d'abord montrer que $\dim \operatorname{Ker} u^k = k$ pour tout $k \in [\![0, n]\!]$ ce qui prouvera que ces $n+1$ sous-espaces sont deux à deux distincts. Le théorème du rang appliqué à la restriction de u à $\operatorname{Im} u^k$ montre que

$$\operatorname{rg} u^k = \operatorname{rg} u^{k+1} + \dim(\operatorname{Ker} u \cap \operatorname{Im} u^k) \leqslant \operatorname{rg} u^{k+1} + 1.$$

On en déduit que $\dim \operatorname{Ker} u^{k+1} \leqslant \dim \operatorname{Ker} u^k + 1$: la dimension des noyaux itérés augmente d'une unité au plus à chaque fois. Mais comme $\operatorname{Ker} u^0 = \{0\}$ et $\operatorname{Ker} u^n = E$, on a nécessairement $\dim \operatorname{Ker} u^k = k$ pour tout $k \in [\![0, n]\!]$.

Considérons maintenant un sous-espace quelconque F de E stable par u. Notons p sa dimension. La restriction de u à F est encore nilpotente avec un indice forcément $\leqslant p$. On a donc $(u|_F)^p = 0$ ce qui veut dire que $F \subset \operatorname{Ker} u^p$. Comme les deux espaces ont la même dimension, ils sont égaux. ◁

On vient de retrouver dans cet exercice un cas particulier d'un résultat bien connu sur la suite des noyaux itérés : elle est croissante, mais croît de moins en moins vite (cf. tome 1, exercice 6.14).

2.47. Sous-espaces stables par un endomorphisme nilpotent (2)

Soit E un K-espace vectoriel de dimension finie, u un endomorphisme nilpotent de E et S un sous-espace vectoriel stable par u tel que $E = \operatorname{Im} u + S$. Montrer que $S = E$.

(École polytechnique)

▷ **Solution.**

Soit p est l'indice de nilpotence de u : $u^p = 0$ et $u^{p-1} \neq 0$. Montrons le résultat par récurrence sur p. Si $p = 1$, $u = 0$ et le résultat est immédiat.

Supposons $p \geqslant 2$. Notons $\widetilde{u}$ l'endomorphisme de $\operatorname{Im} u$ induit par u. On applique u sur l'égalité $\mathrm{E} = \operatorname{Im} u + \mathrm{S}$ et il vient $\operatorname{Im} u = \operatorname{Im} u^2 + u(\mathrm{S})$. On a $\operatorname{Im} u^2 = \operatorname{Im} \widetilde{u}$ et $u(\mathrm{S})$ est stable par $\widetilde{u}$. De plus, l'indice de nilpotence de $\widetilde{u} \in \mathcal{L}(\operatorname{Im} u)$ est $p-1$. Par hypothèse de récurrence, on a $\operatorname{Im} u = u(\mathrm{S}) \subset \mathrm{S}$. On conclut que $\mathrm{S} = \mathrm{E}$. ◁

2.48. Endomorphisme nilpotent semi-simple

Soit E un espace vectoriel de dimension finie. Trouver les endomorphismes nilpotents de E tels que tout sous-espace stable admette un supplémentaire stable.

(École polytechnique)

▷ **Solution.**

Soit u un endomorphisme nilpotent de E, tel que tout sous-espace stable admette un supplémentaire stable. Le noyau de u admet donc un supplémentaire stable F. Comme $\chi_u = \mathrm{X}^n$ (n étant la dimension de E), le polynôme caractéristique de $u_{|\mathrm{F}}$ est X^r avec $r = \dim \mathrm{F}$. Or, comme $\mathrm{F} \cap \operatorname{Ker} u = \{0\}$, $u_{|\mathrm{F}}$ est injective *i.e.* 0 n'est pas valeur propre de $u_{|\mathrm{F}}$. Mais alors, r est nul et $\operatorname{Ker} u = \mathrm{E}$. L'endomorphisme u est donc nul. Réciproquement, l'endomorphisme nul convient. ◁

Il s'agissait de trouver les endomorphismes u nilpotents semi-simples (voir l'exercice 2.50). Comme χ_u est scindé, cela revient à trouver, d'après l'exercice mentionné, les endomorphismes à la fois nilpotents et diagonalisables : seul 0 convient.

L'exercice suivant est plus difficile. Il précise quels sont les sous-espaces stables par un endomorphisme nilpotent qui ont un supplémentaire stable.

2.49. Existence d'un supplémentaire stable par un endomorphisme nilpotent

Soit E un K-espace vectoriel de dimension finie, $f \in \mathcal{L}(\mathrm{E})$ nilpotent et V un sous-espace de E stable par f. Montrer que V admet un supplémentaire stable si et seulement si $f^k(\mathrm{V}) = f^k(\mathrm{E}) \cap \mathrm{V}$, pour tout $k \in \mathbb{N}$.

(École polytechnique)

▷ **Solution.**

• La nécessité de la condition est facile et n'utilise pas le caractère nilpotent de f. Supposons que V admette un supplémentaire stable W.

Soit k un entier naturel. On a toujours $f^k(\mathrm{V}) \subset f^k(\mathrm{E})$ et comme V est stable on a aussi $f^k(\mathrm{V}) \subset \mathrm{V}$. On a donc $f^k(\mathrm{V}) \subset f^k(\mathrm{E}) \cap \mathrm{V}$.

Soit $x \in f^k(\mathrm{E}) \cap \mathrm{V}$. Il existe un vecteur $y \in \mathrm{E}$ tel que $x = f^k(y)$. Ce vecteur s'écrit $y = y_\mathrm{V} + y_\mathrm{W}$ avec $y_\mathrm{V} \in \mathrm{V}$ et $y_\mathrm{W} \in \mathrm{W}$. Ainsi, par stabilité de V et W, il vient

$$x = f^k(y_\mathrm{V}) + f^k(y_\mathrm{W}) \quad \text{avec } f^k(y_\mathrm{V}) \in \mathrm{V} \text{ et } f^k(y_\mathrm{W}) \in \mathrm{W}.$$

Comme V et W sont en somme directe et $x \in \mathrm{V}$, on a nécessairement $x = f^k(y_\mathrm{V}) \in f^k(\mathrm{V})$. On conclut

$$\boxed{f^k(\mathrm{V}) = f^k(\mathrm{E}) \cap \mathrm{V}}.$$

• Réciproquement, supposons que l'on ait $f^k(\mathrm{V}) = f^k(\mathrm{E}) \cap \mathrm{V}$, pour tout $k \in \mathbb{N}$. Nous allons procéder par récurrence sur l'indice de nilpotence p de f (*i.e.* le plus petit entier $p \geqslant 1$ tel que $f^p = 0$).

Si $p = 1$, $f = 0$ et n'importe quel supplémentaire W de V est stable par f.

Supposons $p \geqslant 2$. *Alors f induit un endomorphisme $\widetilde{f}$ nilpotent de $f(\mathrm{E})$.* L'indice de nilpotence de l'endomorphisme $\widetilde{f}$ est $p-1$ puisque, pour tout $k \geqslant 0$, $\widetilde{f}^k(f(\mathrm{E})) = f^{k+1}(\mathrm{E})$. Considérons $\mathrm{E}' = f(\mathrm{E})$ et le sous-espace $\mathrm{V}' = f(\mathrm{V})$ de E'. *Pour appliquer l'hypothèse de récurrence au* sous-espace V' de E', vérifions que pour $k \geqslant 0$, $\widetilde{f}^k(\mathrm{V}') = \widetilde{f}^k(\mathrm{E}') \cap \mathrm{V}'$, *i.e.* que $f^{k+1}(\mathrm{V}) = f^{k+1}(\mathrm{E}) \cap f(\mathrm{V})$. L'inclusion $f^{k+1}(\mathrm{V}) \subset f^{k+1}(\mathrm{E}) \cap f(\mathrm{V})$ est immédiate. Si $x \in f^{k+1}(\mathrm{E}) \cap f(\mathrm{V})$, alors $x \in f^{k+1}(\mathrm{E}) \cap \mathrm{V} = f^{k+1}(\mathrm{V})$ par hypothèse. L'égalité des sous-espaces est donc vérifiée.

Par hypothèse de récurrence, nous pouvons donc affirmer l'existence d'un sous-espace W de $\mathrm{E}' = f(\mathrm{E})$, stable par f tel que

$$f(\mathrm{E}) = f(\mathrm{V}) \oplus \mathrm{W} = (f(\mathrm{E}) \cap \mathrm{V}) \oplus \mathrm{W}.$$

Les sous-espaces W et V sont en somme directe puisque si $x \in \mathrm{W} \cap \mathrm{V}$, alors $x \in f(\mathrm{E}) \cap \mathrm{V}$ et $x = 0$ car $(f(\mathrm{E}) \cap \mathrm{V}) \cap \mathrm{W} = \{0\}$. Construisons à partir de W un sous-espace W' stable par f, supplémentaire de V. Il est naturel de penser à $\mathrm{W}_1 = f^{-1}(\mathrm{W})$ qui a le mérite d'être stable par f. Malheureusement, il n'a aucune raison d'être un supplémentaire de V (il contient $\mathrm{Ker}\, f$ et $\mathrm{Ker}\, f \cap \mathrm{V} \neq \{0\}$ si V n'est pas réduit à $\{0\}$ puisque $f_{|\mathrm{V}}$ est nilpotent). Nous allons corriger cela : le sous-espace W est clairement contenu dans W_1. Les sous-espace V et W étant en somme directe, la somme $(\mathrm{V} \cap \mathrm{W}_1) + \mathrm{W}$ est directe et contenue dans W_1. Prenons W_2 un supplémentaire de $(\mathrm{V} \cap \mathrm{W}_1) \oplus \mathrm{W}$ dans W_1. On a

$$\mathrm{W}_1 = f^{-1}(\mathrm{W}) = (\mathrm{V} \cap \mathrm{W}_1) \oplus \mathrm{W} \oplus \mathrm{W}_2 = (\mathrm{V} \cap \mathrm{W}_1) \oplus \mathrm{W}',$$

en ayant posé $\mathrm{W}' = \mathrm{W} \oplus \mathrm{W}_2$. Le sous-espace W' est bien stable par f puisque $f(\mathrm{W}') \subset \mathrm{W} \subset \mathrm{W}'$.

Montrons que W′ est bien un supplémentaire de V. Soit $x \in V \cap W'$. Alors $x \in V \cap W_1$ et comme $V \cap W_1$ est en somme directe avec W′, on a $x = 0$: la somme $V \oplus W'$ est directe.

Soit $z \in E$. Alors $f(z) \in f(E) = f(V) \oplus W$ et il existe donc $v \in V$ et $w \in W$ tel que $f(z) = f(v) + w$. Dans ces conditions, $f(z - v) = w \in W$ et $z - v \in W_1$. Le vecteur $z - v$ s'écrit $v' + w'$ avec $v' \in V \cap W_1 \subset V$ et $w' \in W'$. Ainsi, on a $z = (v + v') + w' \in V \oplus W'$. On conclut que $E = V \oplus W'$ et que W′ constitue bien un supplémentaire de V stable par f. ◁

Le lecteur pourra ainsi vérifier que dans la situation de l'exercice 2.46 les seuls sous-espaces stables par u qui ont un supplémentaire stable sont $\{0\}$ *et* E.

On dit qu'un endomorphisme u de E *est semi-simple si tout sous-espace de* E *stable par u admet un supplémentaire stable. L'exercice suivant montre qu'un endomorphisme est diagonalisable s'il est semi-simple et si son polynôme caractéristique est scindé. Il en résulte qu'un endomorphisme d'un $\mathbb{C}$-espace vectoriel est semi-simple si, et seulement si, il est diagonalisable.*-

2.50. Endomorphismes semi-simples (1)

Soit E un K-espace vectoriel de dimension finie, $u \in \mathcal{L}(E)$. Montrer l'équivalence des deux conditions suivantes :

(*i*) u est diagonalisable ;

(*ii*) χ_u est scindé et tout sous-espace de E stable par u admet un supplémentaire stable par u.

(**École normale supérieure**)

▷ **Solution.**

• Supposons (*i*). Alors χ_u étant le polynôme caractéristique d'une matrice diagonale, il est scindé. Soit F un sous-espace stable pour u. On peut supposer F non nul et distinct de E. Posons $n = \dim E$ et $p = \dim F$. Choisissons une base $B_F = (f_1, \dots, f_p)$ de F et $B = (e_1, \dots, e_n)$ une base de vecteurs propres de E. Le théorème de la base incomplète permet de compléter la base B_F en une base de E, en lui rajoutant des vecteurs de B, disons $e_{i_1}, \dots, e_{i_{n-p}}$. Il est clair que le sous-espace engendré par les vecteurs que l'on rajoute est un supplémentaire de F qui est stable par u.

• Supposons (*ii*). Considérons $F = \bigoplus_{\lambda \in Sp\, u} \operatorname{Ker}(u - \lambda \mathrm{Id}_E)$, la somme des espaces propres de u. C'est un sous-espace stable par u, donc par hypothèse il admet un supplémentaire stable G. Supposons $\dim G \geqslant 1$ et notons u_G l'endomorphisme induit par u sur G. Comme χ_u est scindé,

il en va de même pour χ_{u_G} (qui divise χ_u). Il admet donc une racine λ puisque son degré n'est pas nul. Il existe donc x non nul dans G tel que $u(x) = u_G(x) = \lambda x$. Mais alors x est dans $F \cap G$ qui n'est donc pas réduit à $\{0\}$. On aboutit à une contradiction.

On en déduit que la dimension de G ne peut valoir que 0 et que E est donc somme de ses sous-espaces propres pour u : u est diagonalisable. $\lhd$

Dans le cas d'un corps quelconque, l'endomorphisme u est semi-simple si et seulement si son polynôme minimal est produit de polynômes irréductibles unitaires distincts. C'est l'objet de l'exercice suivant.

Si l'on s'intéresse aux sous-espaces stables par un endomorphisme u, on peut considérer des noyaux $\operatorname{Ker} P(u)$ *avec* P *polynôme de* $K[X]$ *comme cela a déjà été fait à plusieurs reprises. Une autre manière très simple de construire des sous-espaces stables par un endomorphisme u d'un espace* E *est de prendre un vecteur x quelconque de* E *et de considérer le sous-espace vectoriel* E_x *engendré par la famille* $(u^k(x))_{k\in\mathbb{N}}$. *Il s'agit bien entendu du plus petit sous-espace u-stable qui contient x. Si* E_x *est de dimension finie, en notant r le plus grand entier $k \geqslant 0$ tel que la famille* $(x, u(x), \dots, u^{k-1}(x))$ *soit libre, il apparaît que* $(x, u(x), \dots, u^{r-1}(x))$ *est une base de* E_x *et en particulier* $r = \dim E_x$. *En effet, par définition de r, d'une part,* $(x, u(x), \dots, u^{r-1}(x))$ *est libre et d'autre part, il existe un polynôme* $P \in K[X]$ *de degré r tel que* $P(u)(x) = 0$. *Si* $y \in E_x$, *il s'écrit* $y = Q(u)(x)$ *avec* $Q \in K[X]$. *En opérant la division euclidienne de* Q *par* P, $Q = AP + B$ *avec* A *et* B *dans* $K[X]$, $\deg B < r$, *il vient* $y = A(u) \circ P(u)(x) + B(u)(x) = B(u)(x) \in \operatorname{Vect}(x, u(x), \dots, u^{r-1}(x))$.

2.51. Endomorphismes semi-simples (2)

Soit K un corps, E un K-espace vectoriel de dimension finie et $u \in \mathcal{L}(E)$. Montrer qu'il y a équivalence entre :

(i) tout sous-espace vectoriel de F stable par u admet un supplémentaire stable par u ;

(ii) le polynôme minimal μ_u de u est sans facteur carré.

(**École normale supérieure**)

$\triangleright$ **Solution.**

• Supposons (ii) et écrivons $\mu_u = P_1 \dots P_r$ où les P_i sont des polynômes unitaires irréductibles deux à deux distincts dans $K[X]$. Soit F un sous espace stable. Notons u_F l'endomorphisme de F induit par u. On a $\mu_u(u_F) = 0$ et d'après le théorème de décomposition des noyaux,

$$F = \bigoplus_{i=1}^{r} \operatorname{Ker} P_i(u_F).$$

Or le noyau $\mathrm{Ker}(\mathrm{P}_i(u_{\mathrm{F}}))$ n'est autre que $\mathrm{F} \cap \mathrm{Ker}\,\mathrm{P}_i(u)$, que l'on note F_i. Si l'on construit G_i un supplémentaire de F_i dans $\mathrm{Ker}\,\mathrm{P}_i(u)$, stable par u, alors la somme $\mathrm{G}_1 \oplus \cdots \oplus \mathrm{G}_r$ sera stable par u et ce sera bien un supplémentaire de F.

Faisons le travail pour F_1. Si $\mathrm{F}_1 = \mathrm{Ker}\,\mathrm{P}_1(u)$, $\mathrm{G}_1 = \{0\}$ convient. Sinon, on choisit $x_1 \in \mathrm{Ker}\,\mathrm{P}_1(u)$ en dehors de F_1 et on considère E_{x_1} le sous-espace vectoriel engendré par la famille $(u^k(x_1))_{k\in\mathbb{N}}$. C'est un sous-espace de $\mathrm{Ker}\,\mathrm{P}_1(u)$ non réduit à $\{0\}$ car $x_1 \neq 0$. On a $\mathrm{P}_1(u)(x_1) = 0$ donc la dimension de E_{x_1} est inférieure à $\deg \mathrm{P}_1$ d'après la remarque précédant l'exercice. Par ailleurs, si u_1 désigne l'endomorphisme induit par u sur E_{x_1}, son polynôme minimal divise P_1 puisque $\mathrm{P}_1(u_1)$ est nul. En fait, ce polynôme minimal n'est autre que P_1 puisque P_1 est irréductible et $\mathrm{E}_{x_1} \neq \{0\}$. Or d'après le théorème de Cayley-Hamilton, le degré du polynôme minimal est inférieur à celui du polynôme caractéristique et donc à la dimension de l'espace. On en arrive donc à $\dim \mathrm{E}_{x_1} = \deg \mathrm{P}_1$.

Enfin E_{x_1} est en somme directe avec F_1. En effet, si on suppose que $\mathrm{H} = \mathrm{E}_{x_1} \cap \mathrm{F}_1$ n'est pas réduit à $\{0\}$, en prenant un vecteur x non nul dans H, on montrerait comme précédemment que E_x est de dimension $\deg \mathrm{P}_1$ et finalement $\mathrm{E}_x = \mathrm{E}_{x_1} \subset \mathrm{F}_1$ ce qui est impossible car $x_1 \notin \mathrm{F}_1$.

Si $\mathrm{E}_{x_1} \oplus \mathrm{F}_1$ est égal à $\mathrm{Ker}\,\mathrm{P}_1(u)$, E_{x_1} fait l'affaire. Sinon, on réitère le procédé en prenant x_2 en dehors du sous-espace stable $\mathrm{E}_{x_1} \oplus \mathrm{F}_1$, puis x_3 etc... Comme $\mathrm{Ker}\,\mathrm{P}_1(u)$ est de dimension finie, il existe un nombre fini de vecteurs $x_1, \ldots, x_q$ dans $\mathrm{Ker}\,\mathrm{P}_1(u)$ tels que

$$\mathrm{F}_1 \oplus \mathrm{E}_{x_1} \oplus \cdots \oplus \mathrm{E}_{x_q} = \mathrm{Ker}\,\mathrm{P}_1(u).$$

La somme directe $\mathrm{E}_{x_1} \oplus \cdots \oplus \mathrm{E}_{x_q}$ est stable par u et c'est un supplémentaire de F_1 dans $\mathrm{Ker}\,\mathrm{P}_1(u)$. C'est ce qu'on voulait.

• Supposons (i) et écrivons $\mu_u = \mathrm{P}_1^{\alpha_1} \ldots \mathrm{P}_r^{\alpha_r}$ où les P_i sont des polynômes unitaires irréductibles deux à deux distincts dans $\mathrm{K}[\mathrm{X}]$ et les $\alpha_i \in \mathbb{N}^*$. Raisonnons par l'absurde et supposons par exemple $\alpha_1 > 1$. On considère

$$\mathrm{F} = \mathrm{Ker}\,\mathrm{P}_1(u) \oplus \left(\bigoplus_{i=2}^{r} \mathrm{Ker}\,\mathrm{P}_i^{\alpha_i}(u) \right),$$

qui est stable par u. Alors, par hypothèse, F admet un supplémentaire stable G. On note u_{G} l'endomorphisme de G induit par u. On a $\mu_u(u_{\mathrm{G}}) = 0$ et d'après le théorème de décomposition des noyaux,

$$\mathrm{G} = \bigoplus_{i=1}^{r} \mathrm{Ker}\,\mathrm{P}_i^{\alpha_i}(u_{\mathrm{G}}).$$

Or le noyau $\mathrm{Ker}(\mathrm{P}_i^{\alpha_i}(u_{\mathrm{G}}))$ est $\mathrm{G} \cap \mathrm{Ker}\,\mathrm{P}_i^{\alpha_i}(u)$. Comme $\mathrm{F} \cap \mathrm{G} = \{0\}$ et $\mathrm{Ker}\,\mathrm{P}_i^{\alpha_i}(u) \subset \mathrm{F}$ pour $i \geqslant 2$, on a donc $\mathrm{G} \cap \mathrm{Ker}\,\mathrm{P}_i^{\alpha_i}(u) = \{0\}$ pour tout $i \neq 2$. On en déduit que $\mathrm{G} \subset \mathrm{Ker}\,\mathrm{P}_1^{\alpha_1}(u)$. Il apparaît alors que G est

un supplémentaire de $\operatorname{Ker} P_1(u)$ dans $\operatorname{Ker} P_1^{\alpha_1}(u)$. Dans ces conditions, $P_1(u_{|})^{\alpha_1} = P_1^{\alpha_1}(u_{|}) = 0$ et $P_1(u_{|})$ ne saurait être injectif. On en déduit que $G \cap \operatorname{Ker} P_1(u) = \operatorname{Ker} P_1(u_{|}) \neq \{0\}$, ce qui est contradictoire. On a donc $\alpha_1 = 1$ et pour les mêmes raisons $\alpha_2 = \cdots = \alpha_r = 0$. ◁

On en déduit par exemple qu'une matrice de $\mathcal{M}_n(\mathbb{R})$ est semi-simple si et seulement si elle est diagonalisable dans $\mathcal{M}_n(\mathbb{C})$.

Le cas où l'on peut trouver un vecteur x tel que $E_x = E$ conduit à la notion importante d'endomorphisme cyclique. L'énoncé qui suit en donne une caractérisation simple à l'aide du polynôme minimal.

2.52. Endomorphismes cycliques (1)

Soit E un K-espace vectoriel de dimension finie, $u \in \mathcal{L}(E)$ et $x \subset E$. On définit

$$I = \{P \in K[X],\ P(u) = 0\} \quad \text{et} \quad I_x = \{P \in K[X],\ P(u)(x) = 0\}.$$

1. Montrer l'existence de polynômes unitaires non nuls μ et μ_x tels que $I = \mu K[X]$ et $I_x = \mu_x K[X]$. Montrer que μ_x divise μ.

2. Montrer qu'il existe $x \in E$ tel que $\mu_x = \mu$.

3. On dit que l'endomorphisme u est *cyclique* s'il existe $x \in E$ tel que $E = \operatorname{Vect}(u^k(x))_{k \in \mathbb{N}}$. Montrer l'équivalence

$$u \text{ est cyclique} \iff \mu = \chi_u.$$

(École normale supérieure)

▷ **Solution.**

1. L'ensemble I est le noyau du morphisme d'algèbres

$$\varphi : P \in K[X] \longmapsto P(u) \in \mathcal{L}(E).$$

C'est donc un idéal de E. L'algèbre $\mathcal{L}(E)$ étant de dimension finie, φ ne peut être injectif et $I \neq \{0\}$. Nous savons alors qu'il existe un unique polynôme unitaire non nul μ, appelé polynôme minimal de u, tel que $I = \mu K[X]$.

Montrons directement que I_x est un idéal. Soit P et Q dans I_x et A dans K[X]. Le polynôme nul est dans I_x, et

$$(P + Q)(u)(x) = (P(u) + Q(u))(x) = P(u)(x) + Q(u)(x) = 0,$$
$$(AP)(u)(x) = A(u) \circ P(u)(x) = A(u)(P(u)(x)) = A(u)(0) = 0,$$

et on a bien $P + Q \in I_x$ et $AP \in I_x$. L'ensemble I_x est donc un idéal de K[X], non réduit à $\{0\}$ puisque $I \subset I_x$: il existe donc μ_x, polynôme unitaire non nul, tel que $I_x = \mu_x K[X]$. Comme $\mu \in I_x$, μ_x divise μ.

Le polynôme unitaire μ_x est appelé polynôme minimal ponctuel de u en x.

2. Supposons $n = \dim \mathrm{E} \geqslant 1$ (si $\mathrm{E} = \{0\}$, $x = 0$ convient puisque $\mu = \mu_0 = 1$). Dans ces conditions, on a $\deg \mu \geqslant 1$. On va utiliser la décomposition de μ en facteurs irréductibles.

• Si μ est irréductible, alors pour $x \neq 0$, μ_x est un diviseur de μ distinct de 1, c'est nécessairement μ.

• Supposons que μ soit de la forme P^r avec P irréductible. Considérons une base $(e_1, \ldots, e_n)$ de E. Pour $1 \leqslant i \leqslant n$, μ_{e_i} est de la forme P^{r_i} avec $1 \leqslant r_i \leqslant n$: si on note $s = \max_{1 \leqslant i \leqslant n} r_i$, alors $s \leqslant r$ et $\mathrm{P}^s(u)$ est nul en chaque e_i et donc $\mathrm{P}^s(u) = 0$. Autrement dit $\mu | \mathrm{P}^s$. Cela entraîne $r \leqslant s$, et finalement $r = s$: autrement dit, il existe $i \in [\![1, n]\!]$ tel que $\mu_{e_i} = \mathrm{P}^r = \mu$.

• Supposons μ quelconque : on peut écrire

$$\mu = \mathrm{P}_1^{r_1} \ldots \mathrm{P}_s^{r_s},$$

avec $s \geqslant 1$, les P_i irréductibles deux à deux distincts, et les r_i entiers naturels non nuls. Posons pour $1 \leqslant i \leqslant s$, $\mathrm{Q}_i = \mathrm{P}_i^{r_i}$ et $\mathrm{E}_i = \operatorname{Ker} \mathrm{Q}_i(u)$. Comme $\mu(u) = 0$, le théorème de décomposition des noyaux nous assure que

$$\mathrm{E} = \operatorname{Ker} \mathrm{Q}_1(u) \oplus \cdots \oplus \operatorname{Ker} \mathrm{Q}_s(u) = \mathrm{E}_1 \oplus \cdots \oplus \mathrm{E}_s.$$

Chaque E_i est stable par u ; nous noterons u_i l'endomorphisme de E_i induit par u. Comme pour tout $x \in \mathrm{E}_i$, $\mathrm{Q}_i(u)(x) = 0$, on a $\mathrm{Q}_i(u_i) = 0$. Ainsi, le polynôme minimal de u_i, que nous noterons μ_i est un diviseur de $\mathrm{Q}_i = \mathrm{P}_i^{r_i}$. Il est en fait égal à Q_i. En effet, si on avait $\mu_i = \mathrm{P}_i^{r_i'}$ avec $r_i' < r_i$, alors le polynôme

$$\mathrm{P} = \mathrm{P}_1^{r_1} \ldots \mathrm{P}_{i-1}^{r_{i-1}} \mathrm{P}_i^{r_i'} \mathrm{P}_{i+1}^{r_{i+1}} \ldots \mathrm{P}_s^{r_s}$$

annulerait u (en effet $\mathrm{P}(u)$ s'annule sur chaque E_j pour $j \neq i$, puisque si $x \in \mathrm{E}_j$, $\mathrm{P}_j^{r_j}(u)(x) = 0$ et s'annule aussi sur E_i, puisque si $x \in \mathrm{E}_i$, $\mathrm{P}_i^{r_i'}(u)(x) = 0$; comme E est somme directe des sous-espaces E_i, $\mathrm{P}(u)$ est l'endomorphisme nul). Cela contredirait la minimalité de μ.

D'après ce qui précède, il existe $x_i \in \mathrm{E}_i$ tel que le polynôme minimal ponctuel de u_i en x_i soit égal au polynôme minimal de u_i, *i.e.* Q_i. Or le polynôme minimal ponctuel de u_i en x_i est aussi le polynôme minimal ponctuel de u en x_i puisque si $\mathrm{P} \in \mathrm{K}[\mathrm{X}]$, $\mathrm{P}(u)(x) = \mathrm{P}(u_i)(x)$. On a donc $\mu_{x_i} = \mathrm{Q}_i$.

Considérons $x = x_1 + \cdots + x_s$. On va montrer que $\mu_x = \mu$. Soit $\mathrm{P} \in \mathrm{K}[\mathrm{X}]$ tel que $\mathrm{P}(u)(x) = 0$. On a alors

$$0 = \mathrm{P}(u)(x) = \mathrm{P}(u)(x_1) + \mathrm{P}(u)(x_2) + \cdots + \mathrm{P}(u)(x_s).$$

Chaque E_i étant stable par u, on a $P(u)(x_i) \in E_i$ pour tout i et, comme la somme de ces sous-espaces est directe, $P(u)(x_i) = 0$ pour tout $1 \leqslant i \leqslant s$. On en déduit que $Q_i = \mu_{x_i}$ divise P et comme ces polynômes sont premiers entre eux deux à deux, $\mu = Q_1 \dots Q_s$ divise P. En particulier, μ divise donc μ_x, et finalement $\mu_x = \mu$.

3. Pour x vecteur de E, on pose $E_x = \mathrm{Vect}(u^k(x))_{k\in\mathbb{N}}$. C'est le plus petit sous-espace de E stable par u, contenant x. Montrons pour commencer que la dimension de E_x n'est rien d'autre que le degré r du polynôme minimal ponctuel μ_x. En effet, par définition la famille $\mathrm{Vect}(x, u(x), \dots, u^{r-1}(x))$ est libre. Notons W le sous-espace de dimension r qu'elle engendre. C'est un sous-espace de E_x. On va montrer en fait que $W = E_x$. Comme $\mu_x(u)(x) = 0$ la famille $(x, u(x), \dots, u^r(x))$ est liée. Donc $u^r(x) \in \mathrm{Vect}(x, u(x), \dots, u^{r-1}(x)) = W$. Cela prouve que W est stable par u (l'image d'un vecteur quelconque de la base de W est encore dans W). Comme W contient x on a donc $E_x \subset W$ et finalement $W = E_x$. En particulier E_x est bien de dimension r.

La solution en découle aisément :

• Si u est cyclique, il existe $x \in E$ tel que $E_x = E$ donc tel que $\deg \mu_x = n$ où $n = \dim E$. On a alors $\deg \mu \geqslant n$ d'après la première question. Mais on sait par le théorème de Cayley-Hamilton que μ divise le polynôme caractéristique χ_u de u. Ce dernier étant de degré n on a donc $\mu = \chi_u$.

• Réciproquement, supposons $\chi_u = \mu$. D'après la question précédente, il existe $x \in E$ tel que $\mu_x = \mu = \chi_u$. Pour un tel x on a

$$\dim E_x = \deg \mu_x = \deg \mu = \deg \chi_u = n$$

et donc $E_x = E$. Ainsi u est cyclique. ◁

Il découle directement de cet exercice qu'un endomorphisme nilpotent d'indice de nilpotence égal à la dimension n de l'espace E *est cyclique. En effet, son polynôme minimal vaut* X^n *et est égal au polynôme caractéristique. C'est un exercice classique de démontrer cela directement : on sait que pour tout vecteur x en dehors de* $\mathrm{Ker}\, u^{n-1}$*, la famille* $(x, u(x), \dots, u^{n-1}(x))$ *est une base de* E.

2.53. Endomorphismes cycliques (2)

Soit E un $\mathbb{C}$-espace vectoriel de dimension finie et $u \in \mathcal{L}(E)$. Montrer qu'il y a équivalence entre :

(*i*) les sous-espaces propres de u sont de dimension 1 ;

(*ii*) il existe $x \in E$ tel que $(x, u(x), \dots, u^{n-1}(x))$ soit une base de E.

(**École polytechnique**)

▷ **Solution.**

• Supposons (ii). Notons $\mathcal{B}$ la base $(x, u(x), \dots, u^{n-1}(x))$. On peut écrire $u^n(x) = a_0 x + a_1 u(x) + \cdots + a_{n-1} u^{n-1}(x)$ avec $(a_0, \dots, a_{n-1}) \in \mathbb{C}^n$. La matrice de u dans la base $\mathcal{B}$ est donc

$$\mathrm{M} = \begin{pmatrix} 0 & 0 & \dots & \dots & a_0 \\ 1 & 0 & & & a_1 \\ 0 & 1 & \ddots & & a_2 \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \dots & \dots & 1 & a_{n-1} \end{pmatrix}.$$

Si λ est une valeur propre de u, $\mathrm{M} - \lambda \mathrm{I}_n$ est au moins de rang $n-1$ puisque le bloc $(n-1) \times (n-1)$ en bas à gauche est inversible. Donc $\dim \mathrm{Ker}(\mathrm{M} - \lambda \mathrm{I}_n) \leqslant 1$ d'après le théorème du rang. En fait il y a égalité puisque λ est une valeur propre : $\mathrm{Ker}(u - \lambda \mathrm{I}_n)$ est de dimension 1.

• Supposons (i). Notons $\lambda_1, \dots, \lambda_r$ les valeurs propres deux à deux distinctes de u et $n_1, \dots, n_r$ leurs ordres respectifs comme racines de χ_u. En vertu du théorème de Cayley-Hamilton, $\chi_u(u) = 0$ et d'après le théorème de décomposition des noyaux, on a

$$\mathrm{E} = \bigoplus_{i=1}^{r} \mathrm{F}_i, \text{ avec } \mathrm{F}_i = \mathrm{Ker}(u - \lambda_i \mathrm{I}_n)^{n_i}.$$

Les F_i sont les sous-espaces caractéristiques de u. Ils sont stables par u et on note u_i l'endomorphisme de F_i induit par u et $v_i = u_i - \lambda_i \mathrm{I}_{n_i}$. L'endomorphisme v_i est alors nilpotent. Si on note m_i la dimension de F_i, on a donc $\chi_{v_i} = \mathrm{X}^{m_i}$, $\chi_{u_i} = (\mathrm{X} - \lambda_i)^{m_i}$ et en écrivant la matrice de u dans une base obtenue par recollement de bases de $\mathrm{F}_1, \dots, \mathrm{F}_r$, on obtient

$$\chi_u = \chi_{u_1} \cdots \chi_{u_r} = (\mathrm{X} - \lambda_1)^{m_1} \dots (\mathrm{X} - \lambda_r)^{m_r}.$$

Par unicité de la décomposition en facteurs irréductibles, on en déduit que $\dim \mathrm{F}_i = m_i = n_i$. De plus, on a $\dim \mathrm{Ker}\, v_i = \dim \mathrm{Ker}(u - \lambda_i \mathrm{I}_n) \geqslant 1$, par hypothèse. On en déduit, classiquement, que v_i est nilpotent d'indice maximal. En effet, on a $\dim \mathrm{Ker}\, v_i^{n_i - 1} \leqslant n_i - 1 < n_i$. C'est une conséquence du lemme suivant :

Lemme. *Soit* E *un* K*-espace vectoriel de dimension finie* n, $u, v \in \mathcal{L}(\mathrm{E})$. *Alors* $\dim \mathrm{Ker}\, v \circ u \leqslant \dim \mathrm{Ker}\, u + \dim \mathrm{Ker}\, v$.

Démonstration. On considère l'application linéaire

$$f : x \in \mathrm{Ker}\, v \circ u \longmapsto u(x) \in \mathrm{Ker}\, v.$$

Elle est bien définie et son noyau est $\mathrm{Ker}\, f = \mathrm{Ker}\, u$. D'après le théorème du rang, $\dim \mathrm{Ker}\, v \circ u = \dim \mathrm{Ker}\, u + \dim \mathrm{Im}\, f \leqslant \dim \mathrm{Ker}\, u + \dim \mathrm{Ker}\, v$. ◇

Il existe donc pour tout i, $x_i \in \mathrm{F}$, tel que $v_i^{n_i-1}(x_i) \neq 0$. Posons alors $x = x_1 + \cdots + x_r$. Montrons que $(x, u(x), \ldots, u^{n-1}(x))$ est libre, ce qui achèvera la démonstration. Une relation de liaison se traduit par $\mathrm{P}(u)(x) = 0$, avec $\mathrm{P} \in \mathbb{C}[\mathrm{X}]$ et $\deg \mathrm{P} \leqslant n-1$. On a donc, par linéarité, $0 = \mathrm{P}(u)(x) = \mathrm{P}(u)(x_1) + \cdots + \mathrm{P}(u)(x_r)$ avec $\mathrm{P}(u)(x_i) \in \mathrm{F}_i$ pour tout i. Comme la somme des F_i est directe, $\mathrm{P}(u)(x_i) = 0$, c'est-à-dire $\mathrm{P}(u_i)(x_i) = 0$ pour tout i. Comme $(x_i, v_i(x_i), \ldots, v_i^{n_i-1}(x_i))$ est une base de F_i on en déduit que $\mathrm{P}(u_i) = 0$. Or le polynôme minimal de u_i est $(\mathrm{X}-\lambda_i)^{n_i}$ puisque celui de v_i est X^{n_i} (en effet, v_i est nilpotent et $v_i^{n_i-1} \neq 0$). On en déduit que $(\mathrm{X}-\lambda_i)^{n_i}$ divise P. Finalement χ_u divise P et pour des raisons de degré, il vient nécessairement $\mathrm{P} = 0$ ce qui prouve la liberté de $(x, u(x), \ldots, u^{n-1}(x))$. ◁

En reprenant les notations précédentes lorsque u est supposé cyclique avec $\mathrm{E} = \mathrm{E}_x = \mathrm{Vect}(x, \ldots, u^{n-1}(x))$ et $u^n(x) = a_0 x + \cdots + a_{n-1}u^{n-1}(x)$, il apparaît que la matrice de u dans la base $(x, u(x), \ldots, u^{n-1}(x))$ est

$$\begin{pmatrix} 0 & 0 & \ldots & \ldots & a_0 \\ 1 & 0 & & & a_1 \\ 0 & 1 & \ddots & & a_2 \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \ldots & \ldots & 1 & a_{n-1} \end{pmatrix}$$

qui est la matrice compagnon du polynôme $\mathrm{X}^n - a_{n-1}\mathrm{X}^{n-1} - \cdots - a_1\mathrm{X} - a_0$. Dans ces conditions, d'après l'exercice 1.17, on a

$$\chi_u = \mu_u = \mathrm{X}^n - a_{n-1}\mathrm{X}^{n-1} - \cdots - a_1\mathrm{X} - a_0.$$

Ce calcul de polynôme caractéristique est la clef d'une preuve du théorème de Cayley-Hamilton : on se donne u un endomorphisme d'un espace vectoriel E de dimension finie. Soit $x \in \mathrm{E}$ non nul. On désire montrer que $[\chi_u(u)](x) = 0$. On considère $\mathrm{E}_x = \underset{k\in\mathbb{N}}{\mathrm{Vect}}\, u^k(x)$ et u_x l'endomorphisme induit par u sur E_x. Nous savons que χ_{u_x} divise χ_u. Si on note $r = \dim \mathrm{E}_x$, $(x, u(x), \ldots, u^{r-1}(x))$ est une base de E_x. Si on écrit $u^r(x) = a_0 x + a_1 u(x) + \cdots + a_{r-1}u^{r-1}(x)$, on a alors $\chi_{u_x} = \mathrm{X}^r - a_{r-1}\mathrm{X}^{r-1} - \cdots - a_1\mathrm{X} - a_0$ et

$$\chi_{u_x}(u)(x) = u^r(x) - a_{r-1}u^{r-1}(x) - \cdots - a_1 u(x) - a_0 x = 0.$$

On conclut que $\chi_u(u)(x) = 0$.

Voici une première application de la notion d'endomorphisme cyclique pour caractériser les endomorphismes simples, c'est-à-dire ceux qui n'ont aucun sous-espace stable non trivial.

2.54. Endomorphismes simples

Soit E un K-espace vectoriel de dimension finie $n \geqslant 1$, u un endomorphisme de E. Montrer l'équivalence des conditions suivantes :
(i) les seuls sous-espaces de E stables par u sont $\{0\}$ et E ;
(ii) le polynôme caractéristique χ_u de u est irréductible sur K.
(**École normale supérieure**)

▷ **Solution.**

• $(ii) \Longrightarrow (i)$.

Supposons χ_u irréductible. Soit F un sous-espace de E stable par u. Nous savons que χ_{u_F} divise χ_u. De l'hypothèse, on déduit que $\chi_{u_{|F}}$ est égal à 1 ou χ_u. Il en résulte que $\dim F = \deg \chi_{u_{|F}}$ vaut 0 ou n. Donc, F est réduit à $\{0\}$ ou est égal à E tout entier.

• $(i) \Longrightarrow (ii)$.

Réciproquement, supposons que les seuls sous-espaces stables par u soient $\{0\}$ et E. Soit $x \neq 0$ et considérons E_x le sous-espace engendré par les $u^k(x)$ où k parcourt $\mathbb{N}$. Par hypothèse, $E_x = E$ et comme nous l'avons vu plus haut, $(x, u(x), \ldots, u^{n-1}(x))$ est une base de E. Si $P \in K[X]$ et $P(u)(x) = 0$ avec $\deg P \leqslant n-1$, la liberté de $(x, u(x), \ldots, u^{n-1}(x))$ impose $P = 0$. En particulier, μ_u ne peut être de degré inférieur ou égal à $n-1$. Comme d'après le théorème de Cayley-Hamilton, μ_u divise χ_u, on en déduit que $\mu_u = \chi_u$. Il suffit donc de démontrer que μ_u est irréductible. Soit P un diviseur irréductible unitaire de μ_u. Alors $\operatorname{Ker} P(u)$ est stable par u. S'il était égal à $\{0\}$, $P(u)$ serait injective et si on écrit $\mu_u = PQ$ avec $Q \in K[X]$, alors $P(u) \circ Q(u) = 0$ entraîne $Q(u) = 0$ ce qui est exclu car $\deg Q < \deg \mu_u$. On a donc par hypothèse $\operatorname{Ker} P(u) = E$, ce qui signifie que $P(u) = 0$ et $\mu_u | P$. Ainsi $\chi_u = \mu_u = P$ est irréductible. ◁

Nous venons de montrer en particulier qu'un endomorphisme simple est cyclique. L'exercice qui suit propose une généralisation de ce point. On trouvera une preuve différente de celle de l'exercice 2.52 du fait que le polynôme minimal est le polynôme minimal ponctuel en un certain vecteur, preuve valable sur tout corps infini.

2.55. Nombre fini de sous-espaces stables

Soit K un corps infini. E un K-espace vectoriel de dimension finie n et $u \in \mathcal{L}(E)$. Montrer que l'ensemble des sous-espaces stables par u est fini si et seulement si le polynôme minimal de u est de degré n.
(**École normale supérieure**)

↪ **Solution.**

Il convient d'être familiarisé avec la notion de polynôme minimal ponctuel présentée à l'exercice 2.52.

• Supposons que le polynôme minimal de u soit de degré n. En reprenant le travail fait à l'exercice 2.52, on prouve qu'il existe $x \in \mathrm{E}$ tel que $\mu_x = \mu_u$, autrement dit $(x, u(x), \dots, u^{n-1}(x))$ est une base de E.

Nous allons en donner une autre preuve, utilisant le caractère infini de K et qui est sans doute celle qu'aurait amené l'examinateur lors de l'oral. Considérons l'ensemble $\mathcal{D}$ des diviseurs unitaires de μ_u. Cet ensemble est fini. Pour tout $x \in \mathrm{E}$, le polynôme minimal ponctuel μ_x de x pour u est dans $\mathcal{D}$. Le vecteur x est dans le noyau de $\operatorname{Ker} \mu_x(u)$, si bien que

$$\mathrm{E} = \bigcup_{x \in \mathrm{E}} \operatorname{Ker} \mu_x(u).$$

Or, ces sous-espaces $\operatorname{Ker} \mu_x(u)$ sont en nombre fini puisque $\mathcal{D}$ est fini. Un résultat classique affirme que, lorsque le corps de base est infini, E ne peut être réunion finie de sous-espaces stricts (voir l'exercice 6.2 de notre tome 1 d'algèbre). Il existe donc un vecteur x tel que $\operatorname{Ker} \mu_x(u) = \mathrm{E}$. Il s'ensuit que $\mu_u = \mu_x$.

La preuve que nous venons de rédiger montre que, si le corps de base est $\mathbb{R}$ ou $\mathbb{C}$, l'ensemble des vecteurs x tels que $\mu = \mu_x$ est en fait très gros : c'est le complémentaire d'une réunion finie de sous-espaces stricts, donc clairement un ouvert dense de l'espace.

Ainsi, on peut écrire $\mathrm{E} = \{\mathrm{P}(u)(x)\}_{\mathrm{P} \in \mathrm{K}[\mathrm{X}]}$. Soit F un sous-espace stable par u. Notons $\mathrm{J} = \{\mathrm{P} \in \mathrm{K}[\mathrm{X}], \mathrm{P}(u)(x) \in \mathrm{F}\}$. Alors J est un idéal de $\mathrm{K}[\mathrm{X}]$. En effet $0 \in \mathrm{J}$ et si $(\mathrm{P}, \mathrm{Q}) \in \mathrm{J}^2$,

$$(\mathrm{P} + \mathrm{Q})(u)(x) = \mathrm{P}(u)(x) + \mathrm{Q}(u)(x) \in \mathrm{F},$$

car $\mathrm{P}(u)(x)$ et $\mathrm{Q}(u)(x)$ sont dans F. De plus, si $\mathrm{R} \in \mathrm{K}[\mathrm{X}]$, on a $(\mathrm{RP}(u)(x) = \mathrm{R}(u)(\mathrm{P}(u)(x)) \in \mathrm{F}$ car F est stable par u. Donc $\mathrm{P} + \mathrm{Q}$ et RP sont dans F. L'idéal J n'est pas réduit à $\{0\}$, car $\mu_u \in \mathrm{J}$, et il est principal : il existe $\mathrm{A} \in \mathrm{K}[\mathrm{X}]$ unitaire non nul tel que $\mathrm{J} = \mathrm{AK}[\mathrm{X}]$. Ainsi, pour $y \in \mathrm{E}$, on a

$$y \in \mathrm{F} \iff \exists \mathrm{P} \in \mathrm{K}[\mathrm{X}],\ y = (\mathrm{PA})(u)(x) = \mathrm{P}(u)(\mathrm{A}(u)(x)).$$

Si on pose $x_{\mathrm{F}} = \mathrm{A}(u)(x)$, on a donc $\mathrm{E}_{x_{\mathrm{F}}} = \underset{k \in \mathbb{N}}{\operatorname{Vect}}(u^k(x_{\mathrm{F}})) = \mathrm{F}$. Or, le polynôme A est un diviseur unitaire de μ_u car $\mu_u \in \mathrm{J}$. Donc A ne prend qu'un nombre fini de valeurs et x_{F} également. Il y a donc qu'un nombre fini de sous-espaces stables.

On a démontré au passage que l'endomorphisme induit par un endomorphisme cyclique sur un sous-espace stable est également cyclique.

• Réciproquement, supposons que E ne possède qu'un nombre fini de sous-espaces stables par u. On note toujours $\mathrm{E}_x = \underset{k \in \mathbb{N}}{\operatorname{Vect}}(u^k(x))$:

c'est un sous-espace stable. Les sous-espaces E_x sont en nombre fini par hypothèse. Or, la réunion des E_x pour $x \in E$ est évidemment égale à E. Toujours d'après l'exercice 6.2 de notre tome 1 d'algèbre, tous les sous-espaces E_x ne peuvent être des sous-espaces stricts. Il existe donc $x \in E$ tel que $E_x = E$ et dans ces conditions, $\deg \mu_x = \dim E = n$. Comme on a $\deg \mu_u \geqslant \deg \mu_x$ et en même temps par le théorème de Cayley-Hamilton, $\deg \mu_u \leqslant \dim E$, on obtient $\deg \mu_u = n$. ◁

Voici trois exercices qui reviennent sur la notion de polynôme minimal ponctuel. Lorsqu'on se place en dimension infinie, il peut arriver que tout vecteur x admette un polynôme minimal ponctuel non nul sans que l'endomorphisme u n'ait de polynôme minimal. Prenons par exemple la dérivation D dans $\mathbb{C}[X]$: pour tout polynôme P la famille $(D^k(P))_{k\in\mathbb{N}}$ est liée mais D n'a pas de polynôme annulateur non nul. En effet, le degré du polynôme minimal ponctuel μ_P d'un polynôme donné P est égal à $1 + \deg P$ (c'est tout simplement $X^{1+\deg P}$). Un polynôme annulateur de D doit être multiple de tous les X^k, $k \in \mathbb{N}$, et est donc nul. L'exercice suivant montre que le caractère majoré de la famille des $(\deg \mu_x)_{x\in E}$ est en fait suffisant pour avoir l'existence d'un polynôme minimal global. Il est important de faire l'exercice 2.52 avant de l'aborder.

2.56. Polynôme minimal ponctuel (1)

Soit V un espace vectoriel quelconque, f un endomorphisme de V et $n \in \mathbb{N}^*$. On suppose que, pour tout $x \in V$, la famille $(x, f(x), \dots, f^n(x))$ est liée. Montrer que la famille $(\mathrm{Id}_V, f, \dots, f^n)$ est liée.

(École normale supérieure)

▷ **Solution.**

On peut remarquer que lorsque n vaut 1, on retrouve le lemme célèbre qui affirme que si pour x dans V, x et $f(x)$ sont colinéaires, alors f est une homothétie.

Pour $x \in V$ on pose $I_x = \{Q \in K[X],\ Q(f)(x) = 0\}$. Comme il a été vu dans l'exercice 2.52, I_x est un idéal de $K[X]$, non nul par hypothèse, dont le générateur unitaire sera noté μ_x (on l'appelle le polynôme minimal ponctuel de x). Par hypothèse on a $\deg \mu_x \leqslant n$ pour tout x de V. Quitte à prendre n plus petit, on peut supposer qu'il existe un vecteur a de E tel que $\deg \mu_a = n$.

Nous allons montrer que μ_a annule f, ce qui répondra à la question. Prenons un vecteur x quelconque dans V. Le sous-espace

$$\begin{aligned} \mathrm{W} &= \sum_{k\in\mathbb{N}} \mathrm{Vect}(f^k(x)) + \sum_{k\in\mathbb{N}} \mathrm{Vect}(f^k(a)) \\ &= \mathrm{Vect}(x, f(x), \ldots, f^n(x)) + \mathrm{Vect}(a, f(a), \ldots, f^n(a)). \end{aligned}$$

est stable par f (car somme de deux sous-espaces stables) et de dimension finie. Notons f' l'endomorphisme induit par f sur W. Par hypothèse le polynôme minimal ponctuel pour f' de tout vecteur y de W est de degré $\leqslant n$ (ce polynôme n'est autre que μ_y). Or nous savons d'après la question 2 de l'exercice 2.52 qu'il existe $y \in \mathrm{W}$ tel que $\mu' = \mu_y$, où μ' désigne le polynôme minimal de f'. Comme $a \in \mathrm{W}$ le polynôme μ_a divise μ' et comme ils ont tous les deux le même degré n, ils sont égaux. On a donc $\mu' = \mu_a$. En particulier, comme $x \in \mathrm{W}$ le polynôme μ_x divise $\mu' = \mu_a$ et on a $\mu_a(f')(x) = \mu_a(f)(x) = 0$. Comme cela est valable pour tout $x \in \mathrm{V}$, on a $\mu_a(f) = 0$ et $(\mathrm{Id}_\mathrm{V}, f, \ldots, f^n)$ est lié. $\triangleleft$

L'exercice suivant fournit une caractérisation des matrices cycliques.

2.57. Polynôme minimal ponctuel (2)

Soit K un corps commutatif infini et $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$. Montrer que le polynôme minimal de A est de degré n si et seulement si l'application

$$\begin{array}{rccc} \Phi : & (\mathrm{K}^n)^2 & \longrightarrow & \mathrm{K}^n \\ & (\mathrm{X}, \mathrm{Y}) & \longmapsto & ({}^t\mathrm{XY}, {}^t\mathrm{XAY}, \ldots, {}^t\mathrm{XA}^{n-1}\mathrm{Y}) \end{array}$$

est surjective.

(**École normale supérieure**)

▷ **Solution.**

• Si $\deg \mu_\mathrm{A} < n$, il existe donc $a_0, a_1, \ldots, a_{n-1}$ dans K, non tous nuls tels que

$$a_0\mathrm{I}_n + a_1\mathrm{A} + \cdots + a_{n-1}\mathrm{A}^{n-1} = 0.$$

Dans ces conditions, pour tout $(\mathrm{X}, \mathrm{Y}) \in (\mathrm{K}^n)^2$, on a

$$\begin{aligned} 0 &= {}^t\mathrm{X}(a_0\mathrm{I}_n + a_1\mathrm{A} + \cdots + a_{n-1}\mathrm{A}^{n-1})\mathrm{Y} \\ &= a_0({}^t\mathrm{XY}) + a_1({}^t\mathrm{XAY}) + \cdots + a_{n-1}({}^t\mathrm{XA}^{n-1}\mathrm{Y}). \end{aligned}$$

Par conséquent, tout élément $(\alpha_0, \ldots, \alpha_{n-1})$ dans l'image de Φ est dans l'hyperplan de K^n d'équation

$$a_0\alpha_0 + a_1\alpha_1 + \cdots + a_{n-1}\alpha_{n-1} = 0.$$

Il s'ensuit que Φ n'est pas surjective.

• Supposons $\deg \mu_\mathrm{A} = n$. Comme pour tout $\mathrm{P} \in \mathrm{K}[\mathrm{X}]$, $\mathrm{P}(\mathrm{A}) = 0$ équivaut à $\mathrm{P}({}^t\mathrm{A}) = 0$, le polynôme minimal de ${}^t\mathrm{A}$ est μ_A. D'après l'exercice

2.52, ou mieux en reprenant la preuve donnée dans l'exercice 2.55 (c'est sans doute la preuve qu'attend l'examinateur ici), il existe $X \in K^n$ tel que le polynôme minimal ponctuel de X pour ${}^tA$ soit μ_A. En particulier, la famille $(X, {}^tAX, \dots, {}^tA^{n-1}X)$ est libre. La famille des vecteurs transposés $({}^tX, {}^tXA, \dots, {}^tXA^{n-1})$ l'est tout autant. Les n formes linéaires $Y \in K^n \longmapsto {}^tXA^kY$ pour $0 \leqslant k \leqslant n-1$ sont donc libres dans l'espace dual et le système (S) : ${}^tXA^kY = \alpha_k$, $0 \leqslant k \leqslant n-1$ d'inconnue Y est un système de Cramer pour tout second membre $(\alpha_0, \alpha_1, \dots, \alpha_{n-1}) \in K^n$. Il en résulte que la fonction Φ est surjective. $\lhd$

Notons que lorsqu'on se place sur $\mathbb{C}$, ou plus généralement sur un corps algébriquement clos, on peut aussi prouver ce dernier résultat par récurrence sur la dimension : l'existence d'une valeur propre pour ${}^tu$ donne un hyperplan stable par u qui permet de se ramener à une dimension inférieure. Le lecteur est invité à rédiger les détails.

Enfin, voici une inégalité liant le rang et le degré du polynôme minimal.

2.58. Rang d'un endomorphisme et degré du polynôme minimal

Soit K un sous-corps de $\mathbb{C}$, E un K-espace vectoriel de dimension finie, u un endomorphisme de E, μ_u son polynôme minimal. Démontrer l'inégalité : $\deg \mu_u \leqslant 1 + \operatorname{rg} u$.

(École normale supérieure)

$\rhd$ **Solution.**

La clef de cet exercice consiste encore à dire qu'il existe un vecteur x tel que le polynôme minimal ponctuel en x noté μ_x soit égal à μ_u, ce qui est démontré dans les exercices 2.52 et 2.55. Alors si $d = \deg \mu_u = \deg \mu_x$, le sous-espace $E_x = \operatorname{Vect}(x, u(x), \dots, u^{d-1}(x))$ est de dimension d. En particulier, $(u(x), u^2(x), \dots, u^{d-1}(x))$ est une famille libre de $\operatorname{Im} u$. Ainsi, $d - 1 \leqslant \operatorname{rg} u$, ce qui donne l'inégalité voulue. $\lhd$

L'exercice suivant est un pas vers le théorème de décomposition de Frobenius qui règle la question des classes de similitude (voir le commentaire qui suit l'exercice).

2.59. Réduction d'un endomorphisme en somme d'endomorphismes cycliques

Soit E un K-espace vectoriel de dimension finie n et u un endomorphisme de E. On note π le polynôme minimal de u. Pour tout $x \in E$, on note $E_x = \{P(u)(x),\ P \in K[X]\}$.

1. On suppose π irréductible. Soit F un sous-espace de E stable par u et $x \in \mathrm{E}$. Montrer que $\mathrm{E}_x \subset \mathrm{F}$ ou $\mathrm{E}_x \cap \mathrm{F} = \{0\}$. Montrer qu'il existe des vecteurs $x_1, \dots, x_p$ de E tels que $\mathrm{E} = \bigoplus_{i=1}^{p} \mathrm{E}_{x_i}$.

2. Montrer le même résultat en supposant que la décomposition de π en facteurs irréductibles est sans facteur carré.

3. Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{Q})$ telle que $\mathrm{A}^4 = \mathrm{I}_n$. Montrer qu'il existe $\mathrm{P} \in \mathrm{GL}_n(\mathbb{Q})$ tel que $\mathrm{P}^{-1}\mathrm{AP}$ appartienne à $\mathcal{M}_n(\mathbb{Z})$.

(**École normale supérieure**)

▷ **Solution.**

Nous supposerons connus les résultats de l'exercice 2.52.

1. • Supposons $\mathrm{E}_x \cap \mathrm{F} \neq \{0\}$ (en particulier $x \neq 0$). Nous allons montrer que $\mathrm{E}_x \subset \mathrm{F}$. Notons $v = u_{|\mathrm{E}_x}$ la restriction de u à E_x ; v est alors cyclique. On a donc $\mu_v = \chi_v$ mais comme μ_v divise $\mu_u = \pi$ qui est supposé irréductible, on a $\chi_v = \mu_v = \pi$. Prenons alors un vecteur y non nul quelconque dans $\mathrm{E}_x \cap \mathrm{F}$. On a $\mathrm{E}_y \subset \mathrm{E}_x$ car E_y est le plus petit sous-espace contenant y, stable par u. Comme $\dim \mathrm{E}_y \geqslant 1$, $\chi_{v_{|\mathrm{E}_y}}$ est un polynôme de degré supérieur ou égal à 1 qui divise $\chi_v = \pi$. Comme celui-ci est irréductible, on a forcément $\chi_{v_{|\mathrm{E}_y}} = \pi$. En particulier, la dimension de E_y est égale à $\deg \pi = \dim \mathrm{E}_x$. Ainsi $\mathrm{E}_x = \mathrm{E}_y$. D'autre part $\mathrm{E}_y \subset \mathrm{F}$ puisque F est stable par u et contient y. Donc, $\mathrm{E}_x = \mathrm{E}_y \subset \mathrm{F}$.

• Montrons l'existence de $x_1, \dots, x_p$ tels que $\mathrm{E} = \bigoplus_{i=1}^{p} \mathrm{E}_{x_i}$.

On considère les familles $(x_1, \dots, x_p)$ d'éléments de E telles que la somme des E_{x_i} soit directe (il en existe, (0) par exemple). On en choisit une de telle manière que $\dim \bigoplus_{i=1}^{p} \mathrm{E}_{x_i}$ soit maximale (une telle famille existe car la dimension des sous-espaces de E est majorée par n). Alors $\mathrm{F} = \bigoplus_{i=1}^{p} \mathrm{E}_{x_i}$ est stable par u. Imaginons un instant que F soit distinct de E. Alors, il existe $x \in \mathrm{E}$ tel que $x \notin \mathrm{F}$. Par conséquent, E_x n'est pas contenu dans F, et d'après ce qui précède, $\mathrm{E}_x \cap \mathrm{F} = \{0\}$. Mais alors, la somme $\mathrm{F} + \mathrm{E}_x$ est directe et

$$\dim(\mathrm{E}_{x_1} \oplus \cdots \oplus \mathrm{E}_{x_p} \oplus \mathrm{E}_x) > \dim(\mathrm{E}_{x_1} \oplus \cdots \oplus \mathrm{E}_{x_p})$$

ce qui contredit le choix de la famille des x_i. Donc $\bigoplus_{i=1}^{p} \mathrm{E}_{x_i} = \mathrm{E}$.

2. Écrivons $\pi = \mathrm{P}_1 \dots \mathrm{P}_r$ où les P_i sont irréductibles et deux à deux distincts. D'après le théorème de décomposition des noyaux on a

$$\mathrm{E} = \mathrm{Ker}\,\pi = \bigoplus_{k=1}^{r} \mathrm{Ker}\,\mathrm{P}_k(u).$$

On note $\mathrm{F}_k = \mathrm{Ker}\,\mathrm{P}_k(u)$ pour $1 \leqslant k \leqslant r$. Alors u induit un endomorphisme u_k sur F_k et $\mu_{u_k} = \mathrm{P}_k$ (en effet, $\mathrm{P}_k(u_k) = 0$ d'où $\mu_{u_k} | \mathrm{P}_k$; comme $\mathrm{F}_k \neq \{0\}$, sinon $\dfrac{\pi}{\mathrm{P}_k}$ annulerait u, on a $\mathrm{P}_k \neq 1$, d'où l'égalité $\mu_{u_k} = \mathrm{P}_k$, puisque P_k est irréductible). On sait alors que F_k est somme directe de sous-espaces du type :

$$\{\mathrm{P}(u_k)(x),\ \mathrm{P} \in \mathrm{K}[\mathrm{X}]\} = \{\mathrm{P}(u)(x),\ \mathrm{P} \in \mathrm{K}[\mathrm{X}]\} = \mathrm{E}_x.$$

Par associativité de la somme directe, on a le résultat pour E tout entier.

3. On prend dans cette question $\mathrm{E} = \mathbb{Q}^n$ et u l'endomorphisme canoniquement associé à la matrice A. Le polynôme minimal μ_A divise $\mathrm{X}^4 - 1 = (\mathrm{X} - 1)(\mathrm{X} + 1)(\mathrm{X}^2 + 1)$. Donc μ_A est produit d'irréductibles de $\mathbb{Q}[\mathrm{X}]$ deux à deux distincts. D'après ce qui précède, $\mathbb{Q}^n$ est somme directe de sous-espaces cycliques du type E_x avec $\mu_{u_{|\mathrm{E}_x}} = \mathrm{X} - 1$, $\mathrm{X} + 1$ ou $\mathrm{X}^2 + 1$. Dans une base bien choisie, la matrice de $u_{|\mathrm{E}_x}$ est (1), (-1) ou $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Par conséquent, A est semblable dans $\mathcal{M}_n(\mathbb{Q})$ à une matrice B diagonale par blocs, avec sur la diagonale des blocs du type (1), (-1) ou $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. On a alors $\mathrm{B} \in \mathcal{M}_n(\mathbb{Z})$. ◁

Le cas général est sans doute un peu long pour faire l'objet d'un exercice d'oral. Donnons tout de même le résultat qui répondra enfin à la question posée tout au début du chapitre de savoir quand deux endomorphismes sont semblables. On se donne E *un* K*-espace de dimension* $n \geqslant 1$ *et* u *un endomorphisme de* E *dont on note* μ *le polynôme minimal. Le but est de décomposer* E *en somme directe de sous-espaces stables par* u *dans lesquels* u *induit des endomorphismes cycliques. On sait qu'on peut trouver un vecteur* x *de* E *tel que* $\mu_x = \mu$ *où* μ_x *désigne le polynôme minimal ponctuel en* x *(voir l'exercice 2.52). Le lemme fondamental consiste à prouver que* $\mathrm{E}_x = \mathrm{Vect}(u^k(x))_{k \geqslant 0}$ *admet un supplémentaire* F *stable par* u. *Il suffit alors de recommencer le travail avec la restriction* v *de* u *à* F. *On notera que comme* μ *annule* v, *le polynôme minimal de* v *divise* μ. *On obtient donc le résultat suivant : l'espace* E *se décompose sous la forme* $\mathrm{E} = \mathrm{E}_{x_1} \oplus \mathrm{E}_{x_2} \oplus \cdots \oplus \mathrm{E}_{x_s}$ *et, en notant* P_k *le polynôme minimal (ou caractéristique) de la restriction de* u *à* E_{x_k}, $\mathrm{P}_s | \mathrm{P}_{s-1} | \ldots | \mathrm{P}_2 | \mathrm{P}_1 = \mu$. *On a donc* $\chi_u = \mathrm{P}_1 \mathrm{P}_2 \ldots \mathrm{P}_s$. *Les sous-espaces* E_{x_i} *ne sont pas uniques, mais on démontre que la suite* $(\mathrm{P}_1, \ldots, \mathrm{P}_s)$ *l'est. On appelle cette suite de polynômes la suite des invariants de similitude de* u. *En effet, le résultat essentiel est alors que deux endomorphismes* u *et* v *sont semblables si et seulement si ils ont les mêmes invariants de similitude.*

Nous commençons maintenant une série d'exercices consacrés à diverses équations portant sur des endomorphismes (ou des matrices).

2.60. Équation du second degré matricielle

Trouver les matrices $X \in \mathcal{M}_2(\mathbb{R})$ telles que $X^2 + X = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$.

(École polytechnique)

▷ **Solution.**

La matrice $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ est symétrique réelle, donc diagonalisable et ses valeurs propres sont 0 et 2. On obtient facilement que $\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right)$ est une base de vecteurs propres si $P = \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}$, $P^{-1}AP = \mathrm{Diag}(0,2)$.

Si X est une matrice solution, X commute avec A et laisse donc stable les sous-espaces propres de A : la matrice $X' = P^{-1}XP$ est donc diagonale. On recherche les solutions sous la forme $X = P\,\mathrm{Diag}(\alpha,\beta)P^{-1}$. La matrice X répond au problème si et seulement si

$$\mathrm{Diag}(\alpha,\beta)^2 + \mathrm{Diag}(\alpha,\beta) = \mathrm{Diag}(0,2).$$

autrement dit si et seulement si $\alpha^2+\alpha = 0$ et $\beta^2+\beta = 2$. On trouve alors quatre couples solutions correspondant à $\alpha = -1$ ou 0 et $\beta = 1$ ou -2. En calculant $P\,\mathrm{Diag}(\alpha,\beta)P^{-1}$, on obtient les quatre solutions suivantes

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \frac{1}{2}\begin{pmatrix} -3 & -1 \\ -1 & -3 \end{pmatrix},\ \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \text{ et } \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}. \quad \triangleleft$$

Les deux exercices suivants sont consacrés à l'étude du commutant : la recherche du commutant de u correspond à la résolution de l'équation $ux = xu$.

2.61. Commutant d'une matrice carrée de taille 2

Soit $A \in \mathcal{M}_2(\mathbb{C})$. L'algèbre C des matrices de $\mathcal{M}_2(\mathbb{C})$ commutant avec A peut-elle être un corps ?

(École polytechnique)

▷ **Solution.**

La réponse est non. Si A est une homothétie, son commutant est $\mathcal{M}_2(\mathbb{C})$

tout entier et ce n'est pas un corps. Sinon le commutant de A, qui contient tous les polynômes en A, contient au moins le plan $\mathrm{Vect}(\mathrm{I}_2, \mathrm{A})$. Comme le corps de base est $\mathbb{C}$, la matrice A admet au moins une valeur propre λ. Alors $\mathrm{A} - \lambda \mathrm{I}_2$ n'est pas inversible et est dans le commutant de A. $\triangleleft$

Notons que le résultat n'est plus vrai sur le corps des réels par exemple. Si on prend $\mathrm{A} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, *il est facile de vérifier que le commutant de* A *est le plan* $\mathrm{Vect}(\mathrm{I}_2, \mathrm{A})$ *formé des matrices* $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$. *Or, il est aisé de vérifier que ce commutant est un corps isomorphe à* $\mathbb{C}$.

On sait que le commutant d'une matrice A *contient la sous-algèbre* $\mathrm{K}[\mathrm{A}]$ *des polynômes en* A. *L'exercice suivant montre qu'il y a égalité si et seulement si* A *est cyclique.*

2.62. Dimension du commutant

Soit $\mathrm{A} \in \mathcal{M}_n(\mathrm{K})$. On suppose A diagonalisable et on note $\lambda_1, \ldots, \lambda_r$ les valeurs propres deux à deux distinctes de A et $n_1, \ldots, n_r$ leurs multiplicités respectives.

On note $\mathcal{C}(\mathrm{A})$ le commutant de A et $\mathrm{K}[\mathrm{A}] = \{\mathrm{P}(\mathrm{A}),\ \mathrm{P} \in \mathrm{K}[\mathrm{X}]\}$.

1. Calculer les dimensions de $\mathcal{C}(\mathrm{A})$ et $\mathrm{K}[\mathrm{A}]$.

2. Montrer les équivalences suivantes :

$$\dim \mathcal{C}(\mathrm{A}) = n \iff \dim \mathrm{K}[\mathrm{A}] = n \iff r = n \iff \mathcal{C}(\mathrm{A}) = \mathrm{K}[\mathrm{A}].$$

3. On ne suppose plus A diagonalisable. Trouver une condition nécessaire et suffisante sur A pour que $\mathrm{K}[\mathrm{A}] = \mathcal{C}(\mathrm{A})$.

(École normale supérieure)

▷ **Solution.**

1. • Calculons la dimension de $\mathrm{K}[\mathrm{A}]$. Le polynôme minimal de A est $\mu_\mathrm{A} = (\mathrm{X} - \lambda_1) \ldots (\mathrm{X} - \lambda_r)$ puisque A est diagonalisable. Si on considère l'application linéaire

$$\varphi : \mathrm{P} \in \mathrm{K}[\mathrm{X}] \longmapsto \mathrm{P}(\mathrm{A}) \in \mathrm{K}[\mathrm{A}],$$

elle est surjective et son noyau est $\mu_\mathrm{A}\mathrm{K}[\mathrm{X}]$. Sa restriction à $\mathrm{K}_{r-1}[\mathrm{X}]$ est donc injective, et même surjective puisque si $\mathrm{P} \in \mathrm{K}[\mathrm{X}]$, $\mathrm{P}(\mathrm{A}) = \mathrm{R}(\mathrm{A})$, où R est le reste de P modulo μ_A. Ainsi, la dimension de $\mathrm{K}[\mathrm{A}]$ est égale à celle de $\mathrm{K}_{r-1}[\mathrm{X}]$, c'est-à-dire à r.

• Identifions les matrices aux endomorphismes de K^n canoniquement associés et notons $\mathrm{E}_1, \ldots, \mathrm{E}_r$ les sous-espaces propres de A associés respectivement à $\lambda_1, \ldots, \lambda_r$. Si B commute avec A, alors B laisse stable

chaque sous-espace E_i et induit sur E_i un endomorphisme B_i. Considérons l'application Ψ qui à B dans le commutant de A associe $(B_1, \dots, B_r)$ dans $\mathcal{L}(E_1) \times \cdots \times \mathcal{L}(E_r)$. Cette application est clairement linéaire et injective (car si B est nulle sur chaque E_i, $B = 0$). Elle est également surjective : en effet, si on se donne $(B_1, \dots, B_r)$ dans $\mathcal{L}(E_1) \times \cdots \times \mathcal{L}(E_r)$, l'endomorphisme B qui coïncide avec B_i sur le sous-espace E_i est bien dans le commutant de B, puisque chaque $B_{|E_i} = B_i$ commute avec $A_{|E_i} = \lambda_i \operatorname{Id}_{E_i}$. Il en résulte que $\mathcal{C}(A)$ est isomorphe à $\mathcal{L}(E_1) \times \cdots \times \mathcal{L}(E_r)$ et

$$\dim \mathcal{C}(A) = n_1^2 + n_2^2 + \cdots + n_r^2,$$

puisque n_i est la dimension de E_i.

2. Notons que comme pour tout $i \in [\![1.r]\!]$, on a $n_i^2 \geqslant n_i \geqslant 1$, on a l'inégalité

$$\dim \mathcal{C}(A) = n_1^2 + \cdots + n_r^2 \geqslant n_1 + \cdots + n_r = n.$$

Rappelons aussi que $K[A] \subset \mathcal{C}(A)$. Pour avoir égalité il suffit donc que les deux espaces aient la même dimension.

• Supposons que $\dim \mathcal{C}(A) = n$. L'inégalité précédente montre que nécessairement $n_i^2 = n_i$ pour tout i. On a donc $n_i = 1$ pour tout i et par suite $r = n$. Ainsi, $\dim K[A] = n = \dim \mathcal{C}(A)$ et $\mathcal{C}(A) = K[A]$.

• Supposons que $\dim K[A] = n$ c'est-à-dire que $r = n$. Comme la somme $n_1 + \cdots + n_n$ vaut n, chaque n_i est nécessairement égal à 1, donc on a $\dim \mathcal{C}(A) = n_1^2 + \cdots + n_n^2 = n = \dim K[A]$. De l'inclusion $K[A] \subset \mathcal{C}(A)$, on déduit l'égalité des deux sous-espaces : $\mathcal{C}(A) = K[A]$.

• Supposons enfin que $K[A] = \mathcal{C}(A)$. On a donc les inégalités suivantes

$$\dim K[A] = r \leqslant n = n_1 + \cdots + n_r \leqslant n_1^2 + \cdots + n_r^2 = \dim \mathcal{C}(A).$$

Comme $\dim K[A] = \dim \mathcal{C}(A)$, on en déduit que $\mathcal{C}(A)$ est de dimension n.

3. La dimension de K[A] est toujours égale au $\deg \mu_A$. Cette dimension est donc d'après le théorème de Cayley-Hamilton inférieure ou égale à n. Établissons le lemme suivant sur celle du commutant :

Lemme. *La dimension du commutant de* A *est toujours supérieure ou égale à* n.

Démonstration. Il s'agit en fait de démontrer [1] que $\dim S \geqslant n$. où S est l'espace des solutions du système linéaire

$$AX - XA = 0,$$

d'inconnue $X = (x_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(K)$.

1. Le preuve que nous allons donner est celle de notre collègue Yves Duval. Elle a été publiée dans la RMS 114 de janvier 2004.

Plaçons-nous d'abord dans le cas où A est trigonalisable. Quitte à se placer dans une nouvelle base idoine, on peut supposer $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ triangulaire supérieure. Cherchons les solutions $X = (x_{ij})_{1 \leqslant i,j \leqslant n}$ triangulaires supérieures. Si on restreint le système à $T_n(K)$, il reste $\frac{n(n+1)}{2}$ inconnues. Comme $AX - XA$ est triangulaire supérieure, dire que X est solution revient à écrire $\frac{n(n+1)}{2}$ équations correspondant à la nullité des coefficients de $AX - XA$ dans la partie supérieure. Mais de ces équations, on peut en retirer n, puisque les équations traduisant la nullité des coefficients diagonaux sont triviales : pour $1 \leqslant i \leqslant n$, $a_{ii}x_{ii} - x_{ii}a_{ii} = 0$. Ce système homogène a donc seulement $\frac{n(n+1)}{2} - n$ équations pour $\frac{n(n+1)}{2}$ inconnues. Nous savons alors que l'espace des solutions est au moins de dimension n. Comme $\dim S \cap T_n(K) \geqslant n$, *a fortiori*, la dimension de S est au moins n.

Le résultat reste valable lorsque A n'est pas trigonalisable. La dimension de l'espace des solutions du système $AX - XA = 0$ reste inchangée si on remplace K par un sur-corps commutatif L de K (c'est un corollaire direct du lemme de l'exercice 2.70). Il suffit alors de prendre L de telle manière que χ_A soit scindé. Dans ce cas-là, A est trigonalisable et la dimension de l'espace des solutions est supérieure ou égale à n, que ce soit sur L ou sur K. On conclut que $\dim \mathcal{C}(A) \geqslant n$. $\diamondsuit$

Dans le cas où $K[A] = \mathcal{C}(A)$, le lemme entraîne que la dimension de ce sous-espace est n. En particulier, on a $\dim K[A] = \deg \mu_A = n$. Comme μ_A est un diviseur de χ_A, on obtient que

$$\boxed{\mu_A = \chi_A}.$$

Réciproquement, supposons que $\chi_A = \mu_A$. On sait alors que A est cyclique (se reporter à l'exercice 2.52), autrement dit qu'il existe un vecteur e de K^n tel que $(e, Ae, A^2e, \ldots, A^{n-1}e)$ soit une base de K^n. Considérons alors l'application

$$f : B \in \mathcal{C}(A) \longmapsto Be \in K^n.$$

Cette application est linéaire. Elle est également injective, car si $Be = 0$, on a

$$BA^k e = A^k Be = 0 \quad \text{pour tout } k \geqslant 0.$$

L'endomorphisme B s'annule sur une base de K^n donc est nul. On en déduit que $\dim \mathcal{C}(A) \leqslant \dim K^n = n$. Compte tenu du lemme, on a donc $\dim \mathcal{C}(A) = \dim K[A] = \deg \mu_A = n$. De l'inclusion $K[A] \subset \mathcal{C}(A)$, on déduit l'égalité $\mathcal{C}(A) = K[A]$.

Conclusion. $\boxed{\mathcal{C}(A) = K[A] \iff A \text{ cyclique} \iff \chi_A = \mu_A}$. $\lhd$

Notons que le résultat de la dernière question est immédiat dès lors que l'on dispose du théorème de décomposition en sommes de sous-espaces cycliques exposé page 158. En effet, si u n'est pas cyclique, alors sa liste de facteurs invariants $(P_1, \dots, P_s)$ contient au moins 2 polynômes. Notons $E = E_{x_1} \oplus \cdots \oplus E_{x_s}$ une décomposition de E en sous-espaces cycliques correspondants aux polynômes P_i. Il est clair que la projection sur E_{x_s} parallèlement à la somme des autres sous-espaces commute avec u. Mais elle ne peut pas être un polynôme en u car P_s divise P_1.

Voici une équation toujours construite avec le crochet de Lie, mais avec un second membre. Lorsqu'il y a des solutions (g étant l'inconnue), il s'agit bien entendu d'un sous-espace affine dirigé par le commutant de f. La solution est courte car la plupart des arguments ont déjà été rencontrés dans les exercices précédents.

2.63. Équation avec un crochet de Lie (1)

Soit E un K-espace vectoriel de dimension finie n (K sous-corps de $\mathbb{C}$) et f et g des endomorphismes de E tels que $fg - gf = f$.

1. Montrer que f est nilpotent.

2. On suppose que g est diagonalisable et que $\dim \operatorname{Ker} f = 1$. Déterminer g.

(École polytechnique)

▷ **Solution.**

1. Nous avons donné trois solutions différentes de cette question dans l'exercice 2.32. Notons que c'est pour cette question que sert l'hypothèse faite sur le corps K.

2. L'endomorphisme f est nilpotent de rang $n-1$. On a vu dans l'exercice 2.46 qu'on a alors $\dim \operatorname{Ker} f^k = k$ pour tout $k \in [\![0, n]\!]$ et, comme nous l'avons signalé page 149 dans la remarque qui suit l'exercice 2.52, f est cyclique : pour tout vecteur $e \notin \operatorname{Ker} f^{n-1}$ la famille $(e, f(e), \dots, f^{n-1}(e))$ est une base de E.

On va expliciter g dans une telle base en choisissant pour e un vecteur propre de g : c'est possible puisque, si g est diagonalisable, elle possède forcément un vecteur propre e qui n'appartient pas à $\operatorname{Ker} f^{n-1}$. On note λ la valeur propre associée. On a alors

$$f(e) = fg(e) - gf(e) = \lambda f(e) - g(f(e))$$

de sorte que $g(f(e)) = (\lambda - 1)f(e)$. Autrement dit, $f(e)$ est vecteur propre de g pour la valeur propre $\lambda - 1$. En itérant, on voit de même que $f^k(e)$ est un vecteur propre de g relativement à la valeur propre $\lambda - k$

pour tout $k \in [\![0, n-1]\!]$. On en déduit que $(e, f(e), \dots, f^{n-1}(e))$ est une base de vecteurs propres de g.

Réciproquement, si e est un vecteur quelconque de E qui n'appartient pas à $\operatorname{Ker} f^{n-1}$, la famille $(e, f(e), \dots, f^{n-1}(e))$ est une base de E et pour tout $\lambda \in \mathrm{K}$, l'endomorphisme g défini par $g(f^k(e)) = (\lambda - k)f^k(e)$ pour tout $k \in [\![0, n-1]\!]$, est solution de $fg - gf = f$. ◁

Voici une équation du même type où il n'y a pas de solution.

2.64. Équation avec un crochet de Lie (2)

Soit E un K-espace vectoriel de dimension finie, $(f, h) \in \mathcal{L}(\mathrm{E})^2$. On suppose χ_f irréductible et $\operatorname{rg} h = 1$. Montrer que l'équation $fg - gf = h$ d'inconnue $g \in \mathcal{L}(\mathrm{E})$ n'a pas de solution.

(École polytechnique)

▷ **Solution.**

On raisonne bien entendu par l'absurde en supposant qu'il y a une solution g. On va essayer de trouver un sous-espace F stable par f et non trivial (*i.e.* de dimension comprise entre 1 et $\dim \mathrm{E} - 1$), ce qui contredira l'irréductibilité de χ_f : en effet, le polynôme caractéristique de la restriction de f à F sera un diviseur non trivial de χ_f (on utilise le sens facile de l'exercice 2.54).

Soit e une base de $\operatorname{Im} h$ et $\mathrm{F} = \operatorname{Vect}(f^k(e))_{k \in \mathbb{N}}$. C'est un sous-espace vectoriel de E stable par f, de dimension non nulle puisqu'il contient e. Montrons que $\mathrm{F} \subset \operatorname{Ker} h$, ce qui montrera que $\dim \mathrm{F} \leqslant n - 1$. Il suffit de montrer que, pour tout $k \in \mathbb{N}$, $f^k(e) \in \operatorname{Ker} h$. Il existe $x \in \mathrm{E}$ tel que $e = h(x)$ et $f^k(e) = (f^k h)(x)$. On a $\operatorname{rg}(f^k h) \leqslant \operatorname{rg} h \leqslant 1$. Soit $f^k h = 0$ et il n'y a rien à démontrer, soit $\operatorname{rg}(f^k h) = \operatorname{rg} h = 1$. On considère alors une base $\mathcal{B} = (e_1, \dots, e_n)$ de E telle que $\operatorname{Vect}(e_1, \dots, e_{n-1}) = \operatorname{Ker}(f^k h)$ et soit A la matrice de $f^k h$ dans cette base. On a $\operatorname{Tr}(f^k h) = \operatorname{Tr} \mathrm{A} = a_{n,n}$ et par ailleurs $\operatorname{Tr}(f^k h) = \operatorname{Tr}(f^{k+1} g) - \operatorname{Tr}(f^k g f) = 0$, donc $a_{n,n} = 0$. Cela montre que $\operatorname{Im}(f^k h) \subset \operatorname{Ker}(f^k h)$. Il est clair que $\operatorname{Ker} h \subset \operatorname{Ker}(f^k h)$. Comme ces sous-espaces ont même dimension, ils sont égaux. On en déduit que $f^k(e) = (f^k h)(x) \in \operatorname{Ker} h$ et $\mathrm{F} \subset \operatorname{Ker} h$. ◁

Le thème des exercices suivants est l'équation $\mathrm{AX} - \mathrm{XB} = \mathrm{Y}$, *où* $\mathrm{A}, \mathrm{B}, \mathrm{Y}$ *sont des matrices carrées fixées et* X *l'inconnue, appelée équation de Sylvester. On retrouve le problème de la recherche du commutant en prenant* $\mathrm{A} = \mathrm{B}$ *et* $\mathrm{Y} = 0$.

2.65. Équation de Sylvester (1)

Soient A et B dans $\mathcal{M}_n(\mathbb{C})$. À quelle condition l'équation

$$AX - XB = Y$$

a-t-elle une solution $X \in \mathcal{M}_n(\mathbb{C})$, quelle que soit $Y \in \mathcal{M}_n(\mathbb{C})$?

(École polytechnique)

▷ **Solution.**

On cherche à quelle condition l'endomorphisme Φ de $\mathcal{M}_n(\mathbb{C})$ défini par $\Phi(X) = AX - XB$ est surjectif. Comme $\mathcal{M}_n(\mathbb{C})$ est de dimension finie, il faut que Φ soit injectif. Montrons que ceci est réalisé si et seulement si A et B n'ont pas de valeur propre commune.

• Supposons que A et B n'ont pas de valeur propre commune. Comme $\mathbb{C}$ est algébriquement clos, leurs polynômes caractéristiques χ_A et χ_B sont premiers entre eux. D'après le théorème de Bezout, il existe U et V dans $\mathbb{C}[X]$ tels que $U\chi_A + V\chi_B = 1$. On a donc $U(A)\chi_A(A) + V(A)\chi_B(A) = I_n$. Puisque $\chi_A(A) = 0$, d'après le théorème de Cayley-Hamilton, on en déduit que $V(A)\chi_B(A) = I_n$ et donc que $\chi_B(A)$ est inversible. Soit maintenant $X \in \mathcal{M}_n(\mathbb{C})$ tel que $\Phi(X) = 0$, c'est-à-dire $AX = XB$. On montre facilement par récurrence que pour tout $k \in \mathbb{N}$, on a : $A^kX = XB^k$. D'où l'on déduit, pour tout $P \in \mathbb{C}[X]$, $P(A)X = XP(B)$. En particulier, on a $\chi_B(A)X = X\chi_B(B)$ et donc $\chi_B(A)X = 0$, puisque $\chi_B(B) = 0$. Comme $\chi_B(A)$ est inversible, on en déduit $X = 0$; Φ est donc injective.

• Supposons qu'au contraire, A et B ont une valeur propre commune λ. On va construire une matrice non nulle dans le noyau de Φ. L'idée est de trouver une matrice non nulle M telle que $AM = \lambda M$ et $MB = \lambda M$. La première égalité a lieu si toutes les colonnes de M sont des vecteurs propres de A pour la valeur propre λ et la seconde a lieu si toutes les lignes de M sont des vecteurs propres de B toujours pour la valeur propre λ. Il suffit pour fabriquer une telle matrice de se donner un vecteur propre X pour A, un vecteur propre Y pour B et de considérer la matrice $M = X^tY$: toutes ses colonnes sont colinéaires à X et toutes ses lignes sont colinéaires à Y. De plus elle n'est pas nulle, car son terme général est x_iy_j où $X = (x_1, \ldots, x_n)$ et $Y = (y_1, \ldots, y_n)$ et chacun des deux vecteurs a au moins une coordonnée non nulle. On vient donc de construire M non nulle telle que $\Phi(M) = 0$: Φ n'est pas injective et l'équivalence voulue est établie. ◁

Voici un second énoncé, plus détaillé, où l'on retrouve ce résultat en étudiant plus précisément le spectre de l'application $X \mapsto AX - XB$.

2.66. Équation de Sylvester (2)

Soit $(A, B) \in \mathcal{M}_n(\mathbb{C})^2$.

1. Déterminer les spectres des endomorphismes de $\mathcal{M}_n(\mathbb{C})$ définis par $\alpha : M \longmapsto AM$ et $\beta : M \longmapsto MB$. Décrire les sous-espaces propres de α et β et donner leur dimension en fonction des dimensions des sous-espaces propres de A et B.

2. Montrer que si A (resp. B) est diagonalisable, α (resp. B) l'est aussi.

3. Donner une condition nécessaire ou suffisante pour que l'endomorphisme φ de $\mathcal{M}_n(\mathbb{C})$ défini par $\varphi(M) = AM - MB$ soit surjectif.

4. Soient $A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_4 \end{pmatrix}$, $B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}$ telles que A et B commutent et $\operatorname{Sp} A_1 \cap \operatorname{Sp} A_4 = \emptyset$. Montrer que $B_3 = 0$.

(École polytechnique)

▷ **Solution.**

1. Si X est un vecteur propre de A pour la valeur propre λ, la matrice M composée de n colonnes égales à X vérifie $AM = \lambda M$: ainsi, λ est valeur propre de α.

Réciproquement, soit λ est une valeur propre de α et M une matrice non nulle de colonnes $(C_1, \dots, C_n)$. On a

$$M \in \operatorname{Ker}(\alpha - \lambda \operatorname{Id}) \iff AM = \lambda M \iff \forall i \in [\![1, n]\!],\ AC_i = \lambda C_i.$$

Comme une des colonnes de M est non nulle, λ est une valeur propre de A. On a donc $\operatorname{Sp} A = \operatorname{Sp} \alpha$.

De plus, $\operatorname{Ker}(\alpha - \lambda \operatorname{Id})$ est composée des matrices $M = (C_1, \dots, C_n)$ avec C_i dans $\operatorname{Ker}(A - \lambda I_n)$ pour tout $1 \leqslant i \leqslant n$. Ainsi, $\operatorname{Ker}(\alpha - \lambda \operatorname{Id})$ est isomorphe à $(\operatorname{Ker}(A - \lambda I_n))^n$ et on en déduit que

$$\dim \operatorname{Ker}(\alpha - \lambda \operatorname{Id}) = n \dim \operatorname{Ker}(A - \lambda I_n).$$

Il est inutile de refaire le même travail pour β. En effet, on a clairement

$$\beta(M) = \lambda M \iff MB = \lambda M \iff {}^t B\, {}^t M = \lambda\, {}^t M$$

et comme la transposition est un isomorphisme de $\mathcal{M}_n(\mathbb{C})$ sur lui-même il suffit d'appliquer le résultat précédent à ${}^t B$. On a donc

$$\operatorname{Sp} \beta = \operatorname{Sp}\, {}^t B = \operatorname{Sp} B$$

et pour toute valeur propre λ de B,

$$\dim \operatorname{Ker}(\beta - \lambda \operatorname{Id}) = n \dim \operatorname{Ker}({}^t B - \lambda \operatorname{Id}) = n \dim \operatorname{Ker}(B - \lambda \operatorname{Id}),$$

car une matrice et sa transposée ont même rang.

2. Si A est diagonalisable, on peut écrire

$$\sum_{\lambda\in \mathrm{Sp}\,\alpha} \dim \mathrm{Ker}(\alpha-\lambda\,\mathrm{Id}) = \sum_{\lambda\in \mathrm{Sp}\,\mathrm{A}} n \dim \mathrm{Ker}(\mathrm{A}-\lambda\,\mathrm{Id})$$
$$= n\sum_{\lambda\in \mathrm{Sp}\,\mathrm{A}} \dim \mathrm{Ker}(\mathrm{A}-\lambda\,\mathrm{Id}) = n^2.$$

On en déduit que α est diagonalisable. De même, on montre que si B est diagonalisable, β l'est aussi.

Ce calcul prouve qu'en fait il y a équivalence entre la diagonalisabilité de A *et celle de* α. *Pour le montrer, on peut également écrire la matrice de* α *dans la base* $(E_{11}, E_{21}, \ldots, E_{n1}, E_{12}, \ldots, E_{nn})$ *: on constate qu'il s'agit d'une matrice diagonale par blocs où, sur la diagonale, on trouve* n *blocs* A. *Il en résulte que* $\mu_{\mathrm{A}} = \mu_\alpha$.

3. Remarquons que $\mathcal{M}_n(\mathbb{C})$ étant de dimension finie, dire que φ est surjective revient à dire que φ est bijective.

Les endomorphismes α et β commutent, si bien qu'il existe une base commune de trigonalisation (voir exercice 2.30). Dans une telle base la matrice de φ est triangulaire et ses coefficients diagonaux sont de la forme $\lambda-\mu$ avec $\lambda\in \mathrm{Sp}\,\mathrm{A}$ et $\mu\in \mathrm{Sp}\,\mathrm{B}$. Il s'ensuit que si le spectre de A et le spectre de B sont disjoints, les coefficients diagonaux sont tous non nuls et φ est donc inversible. Réciproquement, si A et B ont une valeur propre commune λ, on procède comme dans l'exercice précédent pour fabriquer une matrice propre à la fois pour α et β c'est-à-dire dans le noyau de φ : rappelons qu'il suffit de prendre la matrice $\mathrm{M} = \mathrm{X}\,{}^t\mathrm{Y} = (x_i y_j)_{1\leqslant i,j\leqslant n}$ où $\mathrm{X} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ est un vecteur propre de A et $\mathrm{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$ un vecteur propre de ${}^t\mathrm{B}$.

Conclusion. L'endomorphisme φ est surjectif si et seulement si

$$\boxed{\mathrm{Sp}\,\mathrm{A}\cap \mathrm{Sp}\,\mathrm{B} = \emptyset}.$$

4. En écrivant les matrices par blocs AB et BA, on constate que $\mathrm{AB} = \mathrm{BA}$ entraîne $\mathrm{A_4B_3} = \mathrm{B_3A_1}$. Autrement dit, $\mathrm{B_3}$ est dans le noyau de $\mathrm{M}\longmapsto \mathrm{A_4M}-\mathrm{MA_1}$. Or cet endomorphisme est injectif puisque les spectres de $\mathrm{A_1}$ et $\mathrm{A_4}$ sont disjoints. On a donc $\mathrm{B_3} = 0$. ◁

Dans la version suivante, formulée abstraitement, on se place sur un corps quelconque. L'existence d'une valeur propre commune n'est alors plus assurée. Elle est remplacée par une condition arithmétique sur les polynômes caractéristiques.

2.67. Équation de Sylvester (3)

Soit E un espace vectoriel de dimension finie $n \geqslant 1$, f et g deux endomorphismes de E.

1. On suppose f et g non inversibles. Montrer qu'il existe h non nul dans $\mathcal{L}(\mathrm{E})$ tel que $h \circ f = g \circ h$.

2. On suppose que f et g ont une valeur propre commune. Montrer qu'il existe h non nul dans $\mathcal{L}(\mathrm{E})$ tel que $h \circ f = g \circ h$.

3. Soit $h \in \mathcal{L}(\mathrm{E})$ tel que $h \circ f = g \circ h$ et $\mathrm{rg}(h) = r \geqslant 1$. Montrer que les polynômes caractéristiques de f et g ont un facteur commun de degré r.

4. Trouver f et g tels que χ_f et χ_g ont un facteur commun de degré r, mais tels qu'il n'existe aucun h de rang supérieur ou égal à r vérifiant $h \circ f = g \circ h$.

(**École polytechnique**)

▷ **Solution.**

1. On va construire h non nul dans $\mathcal{L}(\mathrm{E})$ tel que $h \circ f = g \circ h = 0$. Géométriquement la condition $h \circ f = 0$ signifie que le noyau de h doit contenir $\mathrm{Im}\, f$ et la condition $g \circ h = 0$ que l'image de h doit être contenue dans $\mathrm{Ker}\, g$. Soit F un supplémentaire de $\mathrm{Im}\, f$ et h_1 une application linéaire non nulle de F dans $\mathrm{Ker}\, g$; cela est possible car ni F, ni $\mathrm{Ker}\, g$ ne sont nuls. L'application $h \in \mathcal{L}(\mathrm{E})$ dont la restriction à F est h_1 et la restriction à $\mathrm{Im}\, f$ est nulle vérifie bien $h \circ f = 0$ car $\mathrm{Im}\, f \subset \mathrm{Ker}\, h$ et $g \circ h = 0$ car $\mathrm{Im}\, h \subset \mathrm{Ker}\, g$.

2. Soit λ une valeur propre commune à f et g. Les endomorphismes $f - \lambda\,\mathrm{Id_E}$ et $g - \lambda\,\mathrm{Id_E}$ ne sont pas inversibles. D'après la question 1, il existe $h \in \mathcal{L}(\mathrm{E})$, non nul, tel que

$$h \circ (f - \lambda\,\mathrm{Id_E}) = (g - \lambda\,\mathrm{Id_E}) \circ h.$$

Par distributivité, on obtient $h \circ f - \lambda h = g \circ h - \lambda h$ et donc $h \circ f = g \circ h$.

3. Soit F, G, H les matrices de f, g, h dans une base quelconque. Il existe des matrices P et Q dans $\mathrm{GL}_n(\mathrm{K})$ telles que $\mathrm{H} = \mathrm{PJ}_r\mathrm{Q}$, où $\mathrm{J}_r = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$. L'égalité $h \circ f = g \circ h$ s'écrit $\mathrm{HF} = \mathrm{GH}$, c'est-à-dire

$$\mathrm{J}_r\mathrm{QFQ}^{-1} = \mathrm{P}^{-1}\mathrm{GPJ}_r.$$

On écrit par blocs $\mathrm{QFQ}^{-1} = \begin{pmatrix} \mathrm{F}_1 & \mathrm{F}_2 \\ \mathrm{F}_3 & \mathrm{F}_4 \end{pmatrix}$ et $\mathrm{P}^{-1}\mathrm{GP} = \begin{pmatrix} \mathrm{G}_1 & \mathrm{G}_2 \\ \mathrm{G}_3 & \mathrm{G}_4 \end{pmatrix}$. L'égalité précédente donne $\begin{pmatrix} \mathrm{F}_1 & \mathrm{F}_2 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} \mathrm{G}_1 & 0 \\ \mathrm{G}_3 & 0 \end{pmatrix}$, c'est-à-dire $\mathrm{F}_1 = \mathrm{G}_1$, $\mathrm{F}_2 = 0$, $\mathrm{G}_3 = 0$. On en déduit que

$$\chi_f = \chi_F = \chi_{QFQ^{-1}} = \det\begin{pmatrix} XI_r - F_1 & 0 \\ -F_3 & XI_{n-r} - F_4 \end{pmatrix}$$
$$= \det(XI_r - F_1)\det(XI_{n-r} - F_4)$$

et de même

$$\chi_g = \det(XI_r - G_1)\det(XI_{n-r} - G_4) = \det(XI_r - F_1)\det(XI_{n-r} - G_4).$$

Les polynômes caractéristiques de f et g ont un facteur de degré r en commun.

4. Soit f l'application nulle et g un endomorphisme dont la matrice dans une base quelconque est une matrice nilpotente de rang $n-1$, par exemple $\begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & 0 \end{pmatrix}$. Les endomorphismes f et g ont même polynôme caractéristique X^n. Si $h \in \mathcal{L}(E)$ vérifie $h \circ f = g \circ h$, alors $g \circ h = 0$ et $\operatorname{Im} h \subset \operatorname{Ker} g$. On en déduit que $\operatorname{rg} h \leqslant \dim(\operatorname{Ker} g) \leqslant 1$. Pour $r \in [\![2, n]\!]$, les polynômes caractéristiques de f et g ont un facteur de degré r en commun, mais il n'existe pas d'endomorphisme h de rang r tel que $h \circ f = g \circ h$. ◁

Voici une équation non linéaire. On s'intéresse aux cubes de $\mathcal{M}_3(\mathbb{R})$.

2.68. Matrices possédant une racine cubique

Déterminer l'image de $\Phi : A \in \mathcal{M}_3(\mathbb{R}) \longmapsto A^3 \in \mathcal{M}_3(\mathbb{R})$.
(École normale supérieure)

▷ **Solution.**

Commençons par quelques observations fondamentales. Si B est dans l'image de Φ, donc de la forme $B = A^3$, alors toute matrice semblable à B aussi : en effet, pour $P \in \mathrm{GL}_3(\mathbb{R})$ on a

$$P^{-1}BP = P^{-1}A^3P = (P^{-1}AP)^3.$$

On cherche donc plutôt les classes de similitude incluses dans l'image de Φ. Une autre remarque qu'on peut faire dès maintenant est que si $B = A^3$, alors A et B commutent : en effet, $AB = A^4 = BA$.

Soit $B \in \mathcal{M}_3(\mathbb{R})$ donnée. On va discuter selon la manière dont on peut réduire B. L'examinateur s'attendra sûrement à ce que le candidat commence par regarder le cas simple où B est diagonalisable. Comme tout réel admet une racine cubique, il est clair que toute matrice diagonale $D = \operatorname{Diag}(a, b, c)$ est dans l'image de Φ : il suffit de constater que

$D = \Delta^3$ avec $\Delta = \mathrm{Diag}(\sqrt[3]{a}, \sqrt[3]{b}, \sqrt[3]{b})$. Il en résulte donc que toute matrice diagonalisable est dans l'image de Φ. On va poursuivre la discussion selon le polynôme caractéristique χ de B. Celui-ci est de degré 3 donc admet au moins une racine réelle.

• Regardons pour commencer le cas où χ n'a qu'une seule racine réelle a et un facteur irréductible de degré 2 qui admet deux racines complexes conjuguées $u \pm iv$ avec $v \neq 0$. Dans ce cas B est diagonalisable dans $\mathcal{M}_3(\mathbb{C})$. En fait, on peut montrer que B est semblable dans $\mathcal{M}_3(\mathbb{R})$ à la matrice $B' = \begin{pmatrix} a & 0 & 0 \\ 0 & u & -v \\ 0 & v & u \end{pmatrix}$. En effet, on a $\chi = (X - a)((X-u)^2 + v^2)$ et, d'après le lemme de décomposition des noyaux,

$$\mathbb{R}^3 = \mathrm{Ker}(B - aI_3) \oplus \mathrm{Ker}((B - uI_3)^2 + v^2 I_3).$$

Prenons $X \in \mathrm{Ker}((B-uI_3)^2+v^2I_3)$ et $Y = \frac{1}{v}(BX - uX)$. Alors Y n'est pas colinéaire à X, sinon X serait vecteur propre de B. On a $BX = uX + vY$ et

$$BY - uY = (B - uI_3)Y = \frac{1}{v}(B - uI_3)^2 X = -vX,$$

soit $BY = -vX + uY$. Dans une base constituée d'un vecteur propre pour la valeur propre a, de X et de Y, l'endomorphisme canoniquement associé à B a pour matrice B'. Il existe donc $P \in \mathrm{GL}_3(\mathbb{R})$ telle que $B = PB'P^{-1}$.

On peut aller un peu plus vite en utilisant l'exercice 2.71. Prenons une base quelconque du plan $\mathrm{Ker}((B - uI_3)^2 + v^2 I_3)$ *et notons* C *la matrice de la restriction de* B *à ce plan dans cette base. La matrice* C *est semblable sur* $\mathbb{C}$ *à la matrice* $\begin{pmatrix} u+iv & 0 \\ 0 & u-iv \end{pmatrix}$. *Mais il en est de même de la matrice* $\begin{pmatrix} u & -v \\ v & u \end{pmatrix}$ *car ses valeurs propres sont* $u \pm iv$. *Donc* C *est semblable sur* $\mathbb{C}$ *et donc aussi sur* $\mathbb{R}$ *à cette dernière matrice. Le fait que* B *est semblable à* B' *en découle de suite.*

Comme a admet une racine cubique, il nous suffit de regarder si la matrice $\begin{pmatrix} u & -v \\ v & u \end{pmatrix}$ est un cube de $\mathcal{M}_2(\mathbb{R})$. Or, si on identifie $\mathbb{C}$ et $\mathbb{R}^2$, l'endomorphisme de $\mathbb{R}^2$ de matrice $\begin{pmatrix} u & -v \\ v & u \end{pmatrix}$ s'identifie à la similitude $\mathbb{R}$-linéaire de $\mathbb{C}$ $z \longmapsto (u+iv)z$. Il est alors clair que si on se donne s et t deux réels tels que $(s+it)^3 = (u+iv)$, on a

$$\begin{pmatrix} u & -v \\ v & u \end{pmatrix} = \begin{pmatrix} s & -t \\ t & s \end{pmatrix}^3.$$

On a alors $B' = \Phi(A')$ où $A' = \begin{pmatrix} \sqrt[3]{a} & 0 & 0 \\ 0 & s & -t \\ 0 & t & s \end{pmatrix}$ et donc $B \in \operatorname{Im}\Phi$.

• Il nous reste à regarder le cas où χ est scindé sur $\mathbb{R}$. S'il admet 3 racines distinctes alors B est diagonalisable et cela a déjà été traité. Il reste donc deux cas à regarder : A admet une valeur propre simple a et une valeur propre double b en étant non diagonalisable, ou A admet une valeur propre triple b sans être diagonalisable (c'est-à-dire sans être l'homothétie bI_3). Commençons par le premier de ces deux cas. On a alors $\mathbb{R}^3 = \operatorname{Ker}(B - aI_3) \oplus \operatorname{Ker}(B - bI_3)^2$ et $\operatorname{Ker}(B - bI_3)^2 \neq \operatorname{Ker}(B - bI_3)$ car la dimension de l'espace propre associé à b est égale à 1. Prenons $X_1 \in \operatorname{Ker}(B - aI_3)$, $X_3 \in \operatorname{Ker}(B - bI_3)^2 \setminus \operatorname{Ker}(B - bI_3)$ et $X_2 = (B - bI_3)X_3$. Dans la base (X_1, X_2, X_3), l'endomorphisme canoniquement associé à B a pour matrice $B' = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 1 \\ 0 & 0 & b \end{pmatrix}$. Il nous suffit de regarder si B' est un cube ou non. Si $A^3 = B'$, alors A et B' commutent de sorte que A stabilise forcément les sous-espaces $\operatorname{Ker}(B - aI_3)$ et $\operatorname{Ker}(B - bI_3)^2$. Elle doit donc être de la forme $A = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & \gamma \\ 0 & \delta & \varepsilon \end{pmatrix}$. Comme précédemment, on doit prendre $\alpha = \sqrt[3]{a}$ et le problème se ramène finalement à savoir si la matrice $\begin{pmatrix} b & 1 \\ 0 & b \end{pmatrix}$ est un cube de $\mathcal{M}_2(\mathbb{R})$. Essayons d'abord de chercher une solution triangulaire supérieure de la forme $\begin{pmatrix} \beta & \gamma \\ 0 & \beta \end{pmatrix}$. On est conduit aux relations $\beta^3 = b$ et $3\gamma\beta^2 = 1$. Si b n'est pas nul on a une solution obtenue en prenant $\beta = \sqrt[3]{b}$ et $\gamma = \dfrac{1}{3\beta^2}$. Montrons que dans le cas $b = 0$ il n'y a pas de solution. Supposons qu'il existe $M \in \mathcal{M}_2(\mathbb{R})$ telle que $M^3 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = J$. On a $J^2 = 0$ (J est nilpotente) donc $M^6 = 0$ et M est forcément nilpotente. Mais son indice de nilpotence est forcément $\leqslant 2$ donc $M^2 = 0$ et par suite $M^3 = 0$: contradiction. Donc J n'est pas un cube.

• Terminons par le dernier cas : B admet une valeur propre triple b sans être diagonalisable. Quitte à remplacer B par une matrice semblable, on peut supposer que $B = bI_3 + N$ où N est triangulaire supérieure de diagonale nulle (N est nilpotente et non nulle). Comme précédemment, on peut d'abord chercher une solution triangulaire supérieure de la forme $A = \beta I_3 + cN + dN^2$, où β est la racine cubique réelle de b. Un calcul facile donne $A^3 = bI_3 + 3\beta^2 cN + 3(\beta c^2 + \beta^2 d)N^2$. Si $b \neq 0$ alors $\beta \neq 0$ et pour avoir $A^3 = B$ il suffit de prendre $c = \dfrac{1}{3\beta^2}$ et $d = -\dfrac{c}{\beta}$. Montrons

pour finir que lorsque $b = 0$, $B = N$ n'est pas un cube. C'est le même argument que précédemment. Si $N = M^3$, alors M est nilpotente et donc $M^3 = N = 0$: la seule matrice nilpotente de $\mathcal{M}_3(\mathbb{R})$ qui est un cube est la matrice nulle.

Conclusion. On constate que les seules matrices $B \in \mathcal{M}_3(\mathbb{R})$ qui ne sont pas des cubes sont celles qui admettent 0 pour valeur propre double ou triple et pour lesquelles la dimension de $\operatorname{Ker} B$ n'est pas égale à la multiplicité de la valeur propre. Une autre manière le dire : une matrice B est dans l'image de Φ si et seulement si $\operatorname{Ker} B = \operatorname{Ker} B^2$. ◁

Il est classique de montrer que la condition $\operatorname{Ker} B = \operatorname{Ker} B^2$ *équivaut aussi à* $\operatorname{Im} B = \operatorname{Im} B^2$ *ou encore à* $\mathbb{R}^3 = \operatorname{Im} B \oplus \operatorname{Ker} B$ *(le lecteur se reportera à l'exercice 6.15 du tome 1 d'algèbre).*

2.69. Fonctions polynomiales Φ vérifiant $\Phi(AB) = \Phi(BA)$

On désigne par $f_1, \dots, f_n$ les n fonctions de $A \in \mathcal{M}_n(\mathbb{C})$ qui donnent les coefficients du polynôme caractéristique χ_A de A :

$$\chi_A = X^n + f_1(A)X^{n-1} + \cdots + f_{n-1}(A)X + f_n(A).$$

1. Montrer qu'on a $f_i(AB) = f_i(BA)$ pour tout $i \in [\![1, n]\!]$ et tous A et B dans $\mathcal{M}_n(\mathbb{C})$.

2. Soit Φ une fonction polynôme de $\mathcal{M}_n(\mathbb{C})$ dans $\mathbb{C}$, telle que pour tout couple $(A, B) \in \mathcal{M}_n(\mathbb{C})$, $\Phi(AB) = \Phi(BA)$. Montrer que Φ est un polynôme en $f_1, \dots, f_n$.

(École polytechnique)

▷ **Solution.**

1. Il suffit en fait de prouver que $\chi_{AB} = \chi_{BA}$. Si B est inversible, on peut écrire

$$\begin{aligned}\chi_{AB} &= \det(XI_n - AB) = \det((XB^{-1} - A)B) = \det(XB^{-1} - A)\det B \\ &= \det(B(XB^{-1} - A)) = \det(XI_n - BA) = \chi_{BA}.\end{aligned}$$

Pour A fixé, les fonctions $B \longmapsto f_i(BA) - f_i(AB)$ $(1 \leqslant i \leqslant n)$ sont des fonctions polynomiales des coefficients de $B \in \mathcal{M}_n(\mathbb{C})$: ce sont des fonctions continues de B. Comme $\chi_{AB} = \chi_{BA}$ pour B inversible, elles sont nulles sur $\mathrm{GL}_n(\mathbb{C})$ qui est dense dans $\mathcal{M}_n(\mathbb{C})$. Par continuité, elles sont nulles sur tout $\mathcal{M}_n(\mathbb{C})$.

On peut éviter de s'appuyer sur la densité topologique de $\mathrm{GL}_n(\mathbb{C})$ *par un argument algébrique valable sur un corps commutatif* K *quelconque : la matrice* $B - YI_n$*, à coefficients dans le corps* $K(Y)$ *des fractions rationnelles en* Y*, est inversible dans* $\mathcal{M}_n(K(Y))$*, puisque son déterminant est un polynôme en* Y *non nul, de degré* n*. Ainsi, on a comme précédemment*

$$\chi_{A(B-YI_n)}(X) = \chi_{(B-YI_n)A}(X).$$

En faisant $Y = 0$, *il vient* $\chi_{AB} = \chi_{BA}$. *Une troisième solution est possible en utilisant un astucieux calcul de déterminants par blocs.*

2. Remarquons que Φ est constante sur les classes de similitude : si A et B sont dans $\mathcal{M}_n(\mathbb{C})$ et $P \in GL_n(\mathbb{C})$ avec $B = P^{-1}AP$,

$$\Phi(B) = \Phi(P^{-1}AP) = \Phi(APP^{-1}) = \Phi(A).$$

On va naturellement s'intéresser au cas des matrices diagonalisables pour lesquelles les fonctions f_i sont au signe près les fonctions symétriques élémentaires des valeurs propres : si $\lambda_1, \ldots, \lambda_n$ sont les valeurs propres de A comptées avec multiplicité, on a

$$\chi_A = \prod_{i=1}^{n}(X - \lambda_i) = X^n + \sum_{k=0}^{n-1}(-1)^{n-k}\sigma_{n-k}(\lambda_1, \ldots, \lambda_n)X^k.$$

Considérons le polynôme en $(\lambda_1, \ldots, \lambda_n) \in \mathbb{C}^n$

$$P(\lambda_1, \ldots, \lambda_n) = \Phi\left(\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)\right).$$

Si σ est une permutation de $[\![1, n]\!]$, $\mathrm{Diag}(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})$ est semblable à $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ (cela correspond à la permutation σ des vecteurs de la base canonique). Il s'ensuit que

$$P(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)}) = P(\lambda_1, \ldots, \lambda_n).$$

Autrement dit, le polynôme P est symétrique en les λ_i. Il existe donc un polynôme R tel que

$$\begin{aligned} P(\lambda_1, \ldots, \lambda_n) &= R(\sigma_1(\lambda_1, \ldots, \lambda_n), \ldots, \sigma_n((\lambda_1, \ldots, \lambda_n)) \\ &= R(-f_1(D), f_2(D), \ldots, (-1)^n f_n(D)) \end{aligned}$$

où $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. Si $R' = R(-X_1, X_2, \ldots, (-1)^n X_n)$, on a donc pour toute matrice diagonale D

$$\Phi(D) = R'(f_1(D), \ldots, f_n(D)).$$

Comme Φ et les f_i prennent les mêmes valeurs sur deux matrices semblables, on a donc pour tout matrice A diagonalisable, semblable à D diagonale,

$$\Phi(A) = \Phi(D) = R'(f_1(D), \ldots, f_n(D)) = R'(f_1(A), \ldots, f_n(A)).$$

Nous savons que l'ensemble des matrices diagonalisables est dense dans $\mathcal{M}_n(\mathbb{C})$. Comme les fonctions Φ, R' et f_i sont continues car polynomiales, l'identité suivante valable pour A diagonalisable

$$\Phi(A) = R'(f_1(A), \ldots, f_n(A)),$$

reste vraie par densité pour A quelconque dans $\mathcal{M}_n(\mathbb{C})$.

Conclusion. La fonction Φ est polynomiale en $f_1, \ldots, f_n$. ◁

Nous terminons ce chapitre en regardant ce que deviennent le polynôme minimal et le caractère semblable lorsqu'on change le corps de base.

Dans l'exercice suivant on montre que le polynôme minimal d'une matrice $A \in \mathcal{M}_n(K)$ a en fait ses coefficients dans le plus petit sous-corps de K qui contient tous les coefficients de la matrice A.

2.70. Invariance du polynôme minimal par extension de corps

Soit $M \in \mathcal{M}_n(\mathbb{Q})$. On appelle $\pi_M \in \mathbb{Q}[X]$ (resp. $\mu_M \in \mathbb{R}[X]$) le polynôme minimal de M en tant que matrice de $\mathcal{M}_n(\mathbb{Q})$ (resp. $\mathcal{M}_n(\mathbb{R})$). Montrer que $\pi_M = \mu_M$.

(École normale supérieure)

▷ **Solution.**

Si $P \in \mathbb{R}[X]$ est tel que $P(M) = 0$, alors μ_M divise P (l'ensemble des polynômes annulateurs dans $\mathbb{R}[X]$ de M est un idéal engendré par μ_M). Or $\pi_M \in \mathbb{R}[X]$ et $\pi_M(M) = 0$. Ainsi μ_M divise π_M. Pour prouver que ces deux polynômes sont égaux, il suffit de montrer qu'ils ont le même degré (ils sont tous deux unitaires). Cela découle de l'invariance du rang par extension de corps :

Lemme. *Soit $D \in \mathcal{M}_{n,p}(K)$, L un sur-corps commutatif de K. On note r_K (resp. r_L) le rang de A comme matrice à coefficients dans K (resp. dans L). On a alors $r_K = r_L$.*

Démonstration.

Tout d'abord, on a l'équivalence suivante pour $B \in \mathcal{M}_m(K)$:

$$B \in \mathrm{GL}_m(K) \iff \det B \neq 0 \iff B \in \mathrm{GL}_m(L)$$

le déterminant étant le même, que le calcul se fasse dans K ou L. Par conséquent, toute sous-matrice B de D est inversible sur K si, et seulement si, elle est inversible sur L. Or, r_K est la taille maximale des sous-matrices de D inversibles sur K donc, d'après l'équivalence, la taille maximale des sous-matrices de D inversibles sur L. Ainsi $r_K = r_L$. ◇

Notons alors d le degré de π_M. Par minimalité de d, la famille $(I_n, M, M^2, \ldots, M^{d-1})$ est libre dans $\mathcal{M}_n(\mathbb{Q})$. Écrivons la matrice D de cette famille de vecteurs dans la base canonique $(E_{i,j})_{1 \leqslant i,j \leqslant n}$ de $\mathcal{M}_n(\mathbb{Q})$. Elle possède d colonnes et n^2 lignes. Les coefficients de D sont tous dans

$\mathbb{Q}$ puisque ce sont des coefficients des matrices M^k. Le rang de D sur $\mathbb{Q}$ est égal à d par hypothèse. Le lemme permet de dire qu'il s'agit aussi du rang de D sur $\mathbb{R}$. Il s'ensuit que les colonnes $(I, M, \dots, M^{d-1})$ sont libres dans $\mathcal{M}_n(\mathbb{R})$. Il n'existe donc pas de polynôme annulateur de M à coefficients réels, non nul, de degré strictement inférieur à d.

On en déduit $\deg \mu_M \geqslant d = \deg \pi_M$. Comme on a $\mu_M | \pi_M$ et comme les deux polynômes sont unitaires, on peut affirmer que $\boxed{\pi_M = \mu_M}$. ◁

Le résultat s'étend à un corps quelconque K *et une extension* L *de* K, *avec la même démonstration.*

Dans l'exercice suivant on montre que, deux matrices A *et* B *de* $\mathcal{M}_n(K)$ *étant données, si* A *et* B *sont semblables lorsqu'on se place dans un surcorps de* K, *alors elles sont semblables dans* $\mathcal{M}_n(K)$. *La preuve donnée ici est valable avec un corps infini, le cas général nécessitant le théorème de caractérisation des classes de similitude à l'aide des facteurs invariants. Notons que la première question est souvent posée comme exercice d'oral. Quant à la seconde, on reconnaîtra qu'elle n'est plus dans l'esprit des programmes actuels.*

2.71. Similitude et extension de corps

1. Montrer que deux matrices carrées réelles semblables dans $\mathcal{M}_n(\mathbb{C})$ sont également semblables dans $\mathcal{M}_n(\mathbb{R})$.

2. Soit K un corps commutatif infini, L un sur-corps de K, M et N deux matrices de $\mathcal{M}_n(K)$. On suppose M et N semblables dans $\mathcal{M}_n(L)$. Démontrer qu'elles le sont également dans $\mathcal{M}_n(K)$.

(**École polytechnique**)

▷ **Solution.**

1. Soit M et N deux matrices de $\mathcal{M}_n(\mathbb{R})$ semblables dans $\mathcal{M}_n(\mathbb{C})$: il existe donc $P \in GL_n(\mathbb{C})$ telle que $N = P^{-1}MP$, ce qu'on peut aussi écrire $PN = MP$. Écrivons $P = Q + iR$ avec $Q, R \in \mathcal{M}_n(\mathbb{R})$. On a donc

$$(Q + iR)N = M(Q + iR)$$

et en identifiant partie réelle et imaginaire, il vient

$$QN = MQ \quad \text{et} \quad RN = MR.$$

Si l'une des deux matrices, Q ou R est inversible, M et N sont bien semblables dans $\mathcal{M}_n(\mathbb{R})$. Mais il se peut qu'aucune de ces deux matrices ne soit inversible. Dans ce cas, considérons pour $\lambda \in \mathbb{C}$, la matrice $P_\lambda = Q + \lambda R \in \mathcal{M}_n(\mathbb{C})$. La fonction $\lambda \longmapsto \det(P_\lambda)$ est une fonction polynomiale en λ non identiquement nulle puisqu'en i elle vaut $\det P \neq 0$.

Il existe donc $\lambda_0 \in \mathbb{R}$ tel que $\det(P_{\lambda_0}) \neq 0$. Comme $P_{\lambda_0}N = MP_{\lambda_0}$ et P_{λ_0} est inversible dans $\mathcal{M}_n(\mathbb{R})$, M et N sont bien semblables en tant que matrices réelles.

2. Une généralisation de ce qui précède pour des corps quelconques est exclue. Considérons le système linéaire homogène (S) : $PN - MP = 0$ où l'inconnue est $P = (p_{i,j})_{1 \leqslant i,j \leqslant n}$. C'est un système à n^2 inconnues et n^2 équations et on cherche en fait à montrer qu'il admet une solution $P \in \mathrm{GL}_n(K)$ sachant qu'il en a une dans $\mathrm{GL}_n(L)$. Notons r le rang de ce système c'est-à-dire le rang de la matrice de ce système qui est de taille n^2 et à coefficients dans K. Comme (S) a une solution non nulle dans L on a forcément $r < n^2$. En effet, que le rang de (S) est le même que l'on considère le système à coefficients dans K ou dans L (cf. exercice 2.70). Posons $m = n^2 - r$. L'espace des solutions du système dans K (resp. dans L) est donc un espace de dimension m sur K (resp. sur L).

Choisissons une base $(P_1, \ldots, P_m)$ de l'espace des K-solutions du système. L'invariance du rang par extension de corps que l'on vient de rappeler montre que cette famille $(P_1, \ldots, P_m)$ est encore de rang m dans l'espace $\mathcal{M}_n(L)$. Autrement dit, l'espace des K-solutions est $KP_1 \oplus \cdots \oplus KP_m$ et l'espace des L-solutions est $LP_1 \oplus \cdots \oplus LP_m$. On sait par hypothèse qu'il existe une matrice inversible $Q \in \mathrm{GL}_n(L)$ qui est solution. Elle se décompose sous la forme $Q = \alpha_1 P_1 + \cdots + \alpha_m P_m$, les α_i étant dans L. On cherche à prouver qu'il existe des scalaires $\lambda_1, \ldots, \lambda_m$ de K tels que $P = \lambda_1 P_1 + \cdots + \lambda_m P_m$ soit encore inversible.

Considérons pour cela le polynôme $\det(X_1 P_1 + \cdots + X_m P_m)$ de $L[X_1, \ldots, X_m]$. Ce polynôme est non nul puisque sa valeur en $(\alpha_1, \ldots, \alpha_m)$ est non nulle (car Q est inversible). Comme K est infini (c'est ici seulement que sert cette hypothèse), ce polynôme ne peut être identiquement nul sur K^m : il existe donc $(\lambda_1, \ldots, \lambda_m) \in K^m$ tel que $P = \lambda_1 P_1 + \cdots + \lambda_m P_m$ ait un déterminant non nul. La matrice $P \in \mathcal{M}_n(K)$ est inversible et vérifie $QN = MQ$: les matrices M et N sont donc semblables dans $\mathcal{M}_n(K)$. ◁

Le résultat est vrai en toute généralité mais à notre connaissance il n'y en a pas de preuve élémentaire, c'est-à-dire qui ne passe pas par la détermination des classes de similitude.

Chapitre 3
Le groupe linéaire

La connaissance d'une partie génératrice d'un groupe est fondamentale comme le lecteur pourra s'en rendre compte à travers la plupart des exercices de ce chapitre. Pour un groupe donné, on cherche à avoir des éléments générateurs les plus simples possibles. Lorsque ce groupe opère naturellement sur un ensemble (comme c'est le cas du groupe symétrique ou ici du groupe linéaire), on va donc regarder les éléments qui ont le plus de points fixes. Pour le groupe symétrique on considère alors les transpositions (qui bougent seulement deux éléments) et qui en forment bien une partie génératrice. Dans le cas du groupe linéaire d'un espace vectoriel de dimension finie, on regarde les applications linéaires inversibles qui admettent un hyperplan de points fixes. Cela mène alors aux transvections (cas non diagonalisable) et aux dilatations (qui sont diagonalisables).

Le premier exercice de ce chapitre se propose de démontrer que le groupe linéaire $\mathrm{GL}_n(\mathrm{K})$ est bien engendré par l'ensemble des matrices de dilatation et des matrices de transvection. Il s'agit simplement de regarder convenablement l'algorithme du pivot de Gauss permettant de calculer l'inverse d'une matrice.

3.1. Génération du groupe linéaire

Soit n un entier $\geqslant 2$ et K un corps commutatif. On appelle matrice de transvection toute matrice de la forme $\mathrm{T}_{ij}(\lambda) = \mathrm{I}_n + \lambda \mathrm{E}_{ij}$, où $i \neq j$ et $\lambda \in \mathrm{K}$, et matrice de dilatation toute matrice diagonale $\mathrm{D}_i(\alpha) = \mathrm{I}_n + (\alpha - 1)\mathrm{E}_{ii}$, avec $\alpha \in \mathrm{K}^*$.

1. Montrer que l'ensemble des matrices de transvection engendre le groupe $\mathrm{SL}_n(\mathrm{K})$ et que l'ensemble des matrices de transvection et de dilatation engendre le groupe $\mathrm{GL}_n(\mathrm{K})$.

2. Déterminer les centres des groupes $\mathrm{GL}_n(\mathrm{K})$ et $\mathrm{SL}_n(\mathrm{K})$.

3. On suppose que $\mathrm{K} = \mathbb{R}$ ou $\mathbb{C}$. Démontrer que $\mathrm{SL}_n(\mathrm{K})$ est connexe par arcs.

(École polytechnique)

▷ **Solution.**

1. Notons que les matrices de transvection sont toutes de déterminant 1. Le groupe qu'elles engendrent est donc inclus dans $\mathrm{SL}_n(\mathrm{K})$. La multiplication à gauche (resp. à droite) par une matrice de transvection $\mathrm{T}_{ij}(\lambda)$ revient à effectuer sur les lignes (resp. les colonnes) l'opération

élémentaire $L_i \leftarrow L_i + \lambda L_j$ (resp. $C_j \leftarrow C_j + \lambda C_i$). Notons qu'il est possible de réaliser l'échange de deux lignes (ou de deux colonnes) uniquement à l'aide de transvections modulo un changement de signe : en effet, la matrice $T_{ij}(1)T_{ji}(-1)T_{ij}(1)$ a pour effet (par multiplication à gauche) de remplacer L_i par L_j et L_j par $-L_i$, les autres lignes étant invariantes. Il n'est évidemment pas possible de réaliser l'échange de deux lignes sans apparition de ce signe moins puisque cette opération change le signe du déterminant.

Soit $A \in GL_n(K)$. En appliquant l'algorithme du pivot de Gauss, nous allons transformer A en une matrice de dilatation, en utilisant uniquement des transvections. Comme A est inversible, sa première colonne n'est pas nulle. Si $a_{i1} \neq 0$ avec $i \geqslant 2$, l'opération $L_1 \leftarrow L_1 - \dfrac{a_{11}-1}{a_{i1}} L_i$ permet de mettre un coefficient 1 en position $(1,1)$. Si tous les coefficients a_{i1} pour $i \geqslant 2$ sont nuls, on effectue l'échange des lignes $L_1 \leftarrow L_2$ et $L_2 \leftarrow -L_1$ pour se ramener au cas précédent. En utilisant le coefficient $(1,1)$ comme pivot, une succession d'opérations sur les lignes puis sur les colonnes permet d'annuler tous les autres coefficients de la première ligne et de la première colonne. Autrement dit, il existe des matrices de transvection $M_1, \ldots, M_p$ et $N_1, \ldots, N_q$ telles que

$$M_p \ldots M_1 A N_1 \ldots N_q = \begin{pmatrix} 1 & 0 \\ 0 & A_1 \end{pmatrix}$$

où $A_1 \in GL_{n-1}(K)$.

On recommence le même algorithme sur la matrice A_1 et ainsi de suite. On aboutit à la fin de cet algorithme à une matrice diagonale $\mathrm{Diag}(1, 1, \ldots, 1, \alpha)$ où le scalaire α n'est autre que $\det A$. On vient donc de montrer que pour toute matrice inversible A, il existe des matrices de transvection $U_1, \ldots, U_r$ et $V_1, \ldots, V_s$ telles que

$$A = U_r \ldots U_1 D_n(\det A) V_1 \ldots V_s.$$

Cela permet de répondre à la question : toute matrice $A \in SL_n(K)$ s'écrit comme un produit de matrices de transvection et toute matrice de $GL_n(K)$ est produit de matrices de transvection et de dilatation.

2. Soit A une matrice qui commute avec toutes les matrices de transvection. Alors A commute avec toutes les matrices E_{ij}, $i \neq j$. Cela impose que A est scalaire. On en déduit que le centre de $GL_n(K)$ est l'ensemble des matrices scalaires λI_n avec $\lambda \neq 0$. Il est isomorphe au groupe $(K^*, \times)$. Et le centre de $SL_n(K)$ est l'ensemble des matrices scalaires λI_n avec $\lambda^n = 1$. Il est isomorphe au groupe des racines n-ièmes de l'unité dans le corps K. Son cardinal dépend bien entendu du corps K.

3. Montrons que toute matrice $A \in SL_n(K)$ est reliée par un arc continu à l'identité I_n. D'après la première question, il existe une partie X contenue dans l'ensemble des couples $(i,j) \in [\![1,n]\!]^2$ avec $i \neq j$ et

une famille $(\lambda_C)_{C\in X}$ de K telles que A soit le produit des transvections $T_C(\lambda_C)$:

$$A = \prod_{C\in X} T_C(\lambda_C).$$

En posant $\varphi : t \in [0,1] \longmapsto A_t = \prod_{C\in X} T_C(t\lambda_C)$, on obtient un arc continu qui relie $\varphi(0) = I_n$ à $\varphi(1) = A$ et $SL_n(K)$ est donc connexe par arcs. ◁

Le résultat de la première question est très important et sera utilisé dans plusieurs des exercices de ce chapitre, dont les deux suivants.

3.2. Groupe engendré par les matrices diagonalisables inversibles

Quel est le sous-groupe de $GL_n(\mathbb{R})$ engendré par l'ensemble des matrices inversibles diagonalisables ?

(École normale supérieure)

▷ **Solution.**

On sait (voir l'exercice précédent) que $GL_n(\mathbb{R})$ est engendré par les matrices de dilatation et les matrices de transvection. Les premières sont diagonales. Montrons que toute matrice de transvection peut s'écrire comme un produit de matrices diagonalisables. Si $M = I_n + \lambda E_{ij}$ avec $i \neq j$ et $\lambda \in \mathbb{R}$ est une telle matrice, il suffit d'écrire $M = D^{-1}(DM)$ où D est une matrice diagonale dont les coefficients diagonaux sont non nuls et deux à deux distincts (par exemple $1, 2, \ldots, n$). La matrice D^{-1} est diagonale et la matrice DM est triangulaire avec la même diagonale que D : elle est donc diagonalisable.

Conclusion. L'ensemble des matrices diagonalisables inversibles engendre tout le groupe $GL_n(\mathbb{R})$. ◁

3.3. Produits d'automorphismes de trace nulle

Soit $n \geqslant 2$. Déterminer les matrices de $GL_n(K)$ qui s'écrivent comme un produit de matrices de trace nulle.

(École normale supérieure)

▷ **Solution.**

Notons déjà que si $M \in GL_n(K)$ est décomposée comme un produit de matrices de trace nulle, alors toutes les matrices du produit sont forcément inversibles. Il est clair que l'ensemble cherché est stable par produit. Nous savons que toute matrice de $GL_n(K)$ s'écrit comme un produit de matrices de transvection et de dilatation. On s'intéresse donc pour commencer à ces matrices élémentaires.

Il est facile de trouver des matrices inversibles de trace nulle et même

de diagonale nulle : il suffit de prendre une matrice de permutation associée à une permutation sans point fixe, comme par exemple

$$P = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \dots & \dots & 0 \end{pmatrix}.$$

Si D est une matrice diagonale quelconque, alors la matrice DP est de diagonale nulle puisque la multiplication à droite par P a pour effet de réaliser le n-cycle $(1, 2, \dots, n)$ sur les colonnes. En particulier si D est une matrice de dilatation, il suffit d'écrire $D = (DP)P^{-1}$ pour voir que D est produit de deux matrices de trace nulle (puisque $P^{-1} = {}^tP$ est aussi de diagonale nulle).

Passons aux matrices de transvection. Essayons d'appliquer la même idée. Soit $T = I_n + \lambda E_{ij}$ avec $i \neq j$ une matrice de transvection. Si $n \geqslant 3$ alors l'une au moins des matrices TP ou TP^{-1} a une diagonale nulle puisque soit $j - 1$ soit $j + 1$ est distinct de i modulo n. On conclut comme avant en écrivant selon le cas $T = (TP)P^{-1}$ ou $T = (TP^{-1})P$. Il reste à traiter le cas $n = 2$. Mais c'est facile puisqu'on peut écrire

$$\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & \lambda \\ 0 & -1 \end{pmatrix}$$

si K n'est pas de caractéristique 2 et dans le cas où K est de caractéristique 2 la matrice de transvection est elle-même de trace nulle.

Conclusion. Toute matrice de $GL_n(K)$ peut s'écrire comme un produit de matrices inversibles de trace nulle. ◁

Les prochains exercices sont consacrés à des sous-groupes du groupe linéaire. Le premier demande ainsi de montrer que $GL_2(\mathbb{Q})$ ne contient pas de sous-groupe cyclique de cardinal 5.

3.4. Élément d'ordre 5 de $GL_2(\mathbb{Q})$

Montrer que le groupe $GL_2(\mathbb{Q})$ ne contient pas d'élément d'ordre 5.

(École normale supérieure)

▷ **Solution.**

Supposons par l'absurde qu'il existe $A \in GL_2(\mathbb{Q})$ d'ordre 5. Le poly-

nôme $P = X^5 - 1 = (X - 1)(X^4 + X^3 + X^2 + X + 1)$ annule A. Montrons que le second facteur Q est irréductible dans $\mathbb{Q}[X]$. En effet, Q n'a pas de racine réelle (donc *a fortiori* pas de racine rationnelle) et si $Q = AB$ où A, B sont deux polynômes unitaires du second degré alors, par unicité de la décomposition dans $\mathbb{R}[X]$, ces polynômes sont $X^2 + \frac{1}{2}(1+\sqrt{5})X + 1$ et $X^2 + \frac{1}{2}(1-\sqrt{5})X + 1$. C'est absurde car $\sqrt{5} \notin \mathbb{Q}$. Donc Q est bien irréductible dans $\mathbb{Q}[X]$.

Le polynôme minimal μ de A est un diviseur de P dans $\mathbb{Q}[X]$. Comme μ est de degré inférieur à 2, on a $\mu = X - 1$, et donc $A = I_2$. D'où le résultat. ◁

3.5. Conjugaison dans $\mathbf{SL_2(\mathbb{C})}$

Soit $A \in \mathrm{SL}_2(\mathbb{C})$ telle que $\mathrm{Tr}\, A \neq \pm 2$.

1. Montrer qu'il existe $P \in \mathrm{SL}_2(\mathbb{C})$ telle que $P^{-1}AP$ soit diagonale.

2. Soit $B \in \mathrm{SL}_2(\mathbb{C})$ telle que A et B n'aient pas de vecteurs propres communs. Montrer qu'il existe $P \in \mathrm{SL}_2(\mathbb{C})$ telle que $P^{-1}AP$ et $P^{-1}BP$ soient symétriques.

3. Soit $a_1, \ldots, a_n, b_1, \ldots, b_n$ des éléments de $\mathbb{Z}$ tels que $A^{a_1}B^{b_1} \ldots A^{a_n}B^{b_n} = I_2$. Montrer que $A^{-a_1}B^{-b_1} \ldots A^{-a_n}B^{-b_n} = I_2$.

(**École normale supérieure**)

▷ **Solution.**

1. Le polynôme caractéristique de A est $\chi_A = X^2 - \mathrm{Tr}\, A\, X + 1$. Son discriminant $(\mathrm{Tr}\, A)^2 - 4$ n'est pas nul par hypothèse, donc A possède deux valeurs propres distinctes, elle est diagonalisable. Il existe $Q \in \mathrm{GL}_2(\mathbb{C})$ telle que $Q^{-1}AQ$ soit diagonale. Soit ω une racine carrée de $\det Q$ et $P = \frac{1}{\omega}Q$. Alors P appartient à $\mathrm{SL}_2(\mathbb{C})$ et répond à la question posée car $P^{-1}AP = Q^{-1}AQ$.

2. Soit (e_1, e_2) une base de vecteurs propres de A. La matrice de l'endomorphisme u_B canoniquement associé à B dans la base (e_1, e_2) est $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$, avec $bc \neq 0$, car ni e_1, ni e_2 n'est vecteur propre de B. Soit $\alpha \in \mathbb{C}^*$. La matrice de u_B dans la base $(\alpha e_1, e_2)$ est $\begin{pmatrix} a & \frac{b}{\alpha} \\ \alpha c & d \end{pmatrix}$. Elle est symétrique si on choisit α tel que $\alpha^2 = \frac{b}{c}$. Dans la même base, la matrice de l'endomorphisme associé à A reste diagonale donc symétrique. Si $Q \in \mathrm{GL}_2(\mathbb{C})$ est la matrice de passage de la base canonique de $\mathbb{C}^2$ à la base $(\alpha e_1, e_2)$, les matrices $Q^{-1}AQ$ et $Q^{-1}BQ$ sont symétriques. Soit ω

une racine carrée de $\det Q$ et $P = \frac{1}{\omega} Q$. Alors P appartient à $SL_2(\mathbb{C})$ et répond à la question posée car $P^{-1}AP = Q^{-1}AQ$ et $P^{-1}BP = Q^{-1}BQ$.

3. Avec les notations de la question précédente, on pose $U = P^{-1}AP$ et $V = P^{-1}BP$. Ce sont des matrices symétriques. L'hypothèse équivaut à $U^{a_1}V^{b_1}\dots U^{a_n}V^{b_n} = I_2$. En prenant la transposée de chaque membre de l'égalité, on obtient $V^{b_n}U^{a_n}\dots V^{b_1}U^{a_1} = I_2$, puis en prenant l'inverse $U^{-a_1}V^{-b_1}\dots U^{-a_n}V^{-b_n} = I_2$, ce qui équivaut à l'égalité demandée. ◁

Si deux groupes G *et* G′ *sont isomorphes, tout sous-groupe de* G *doit se retrouver dans* G′. *L'exercice suivant utilise cette idée pour montrer que deux groupes linéaires* $GL_n(K)$ *et* $GL_m(L)$ *ne peuvent pas être isomorphes si* $n \neq m$.

3.6. Isomorphismes entre groupes linéaires

Soient K et L deux corps commutatifs de caractéristique différente de 2.

1. Soit G un sous-groupe fini de $GL_n(K)$ tel que pour toute matrice A de G, $A^2 = I$. Montrer que $\operatorname{Card} G \leqslant 2^n$.

2. On suppose qu'il existe un morphisme injectif du groupe $GL_n(K)$ dans le groupe $GL_m(L)$. Montrer que $n \leqslant m$.

3. Existe-t-il un isomorphisme entre $GL_n(\mathbb{Q})$ et $GL_m(\mathbb{R})$? entre $GL_n(\mathbb{R})$ et $GL_m(\mathbb{C})$?

(École normale supérieure)

▷ **Solution.**

1. Comme K est de caractéristique différente de 2, le polynôme $X^2 - 1 = (X-1)(X+1)$ est scindé à racines simples. Par hypothèse, ce polynôme annule toute matrice A de G qui est donc diagonalisable avec un spectre inclus dans $\{\pm 1\}$. Géométriquement les éléments de G sont des symétries.

Par ailleurs G est abélien : en effet, pour tout $A \in G$, on a $A = A^{-1}$ et si $(A, B) \in G^2$,

$$AB = (AB)^{-1} = B^{-1}A^{-1} = BA.$$

Les éléments de G sont alors codiagonalisables (c'est un résultat classique que le lecteur pourra retrouver dans l'exercice 2.21). Il existe donc $P \in GL_n(K)$ telle que, pour tout $A \in G$,

$$P^{-1}AP = \begin{pmatrix} \pm 1 & 0 & \dots & 0 \\ 0 & \pm 1 & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & \pm 1 \end{pmatrix}.$$

Le cardinal de G est donc inférieur à celui des familles $(\varepsilon_1, \dots, \varepsilon_n)$ de $\{-1, 1\}^n$ qui est égal à 2^n.

2. Considérons le sous-groupe G de $\mathrm{GL}_n(\mathrm{K})$ constitué des matrices diagonales avec des 1 ou -1 sur la diagonale :

$$\mathrm{G} = \{\mathrm{M} \in \mathrm{GL}_n(\mathrm{K}).\ \exists (\varepsilon_1, \dots \varepsilon_n) \in \{-1, 1\}^n,\ \mathrm{M} = \mathrm{Diag}(\varepsilon_1, \dots, \varepsilon_n)\}.$$

Le groupe G est de cardinal 2^n et pour tout $\mathrm{A} \in \mathrm{G}$, $\mathrm{A}^2 = \mathrm{I}_n$. On considère un morphisme injectif de $\mathrm{GL}_n(\mathrm{K})$ dans $\mathrm{GL}_m(\mathrm{L})$. Alors l'image G′ de G par ce morphisme est un sous-groupe fini isomorphe à G. En particulier, pour tout $\mathrm{B} \in \mathrm{G}'$, $\mathrm{B}^2 = \mathrm{I}_m$ et $\mathrm{Card}\,\mathrm{G}' = 2^n$. D'après la première question, on a $2^n \leqslant 2^m$, ce qui entraîne $n \leqslant m$.

3. La question précédente montre que si $\mathrm{GL}_n(\mathrm{K})$ est isomorphe à $\mathrm{GL}_m(\mathrm{L})$ alors nécessairement $n = m$. Le groupe $\mathrm{GL}_n(\mathbb{Q})$ ne peut pas être isomorphe à $\mathrm{GL}_n(\mathbb{R})$ (ou $\mathrm{GL}_n(\mathbb{C})$) car il est dénombrable alors que ces deux derniers sont indénombrables. Montrons pour terminer que $\mathrm{GL}_n(\mathbb{R})$ et $\mathrm{GL}_n(\mathbb{C})$ ne sont pas isomorphes. Pour $n = 1$, $\mathbb{R}^*$ et $\mathbb{C}^*$ ne sont pas isomorphes : il y a des éléments d'ordre 3 dans $\mathbb{C}^*$ mais pas dans $\mathbb{R}^*$. Ce cas suffit à conclure puisque le centre de $\mathrm{GL}_n(\mathrm{K})$ est isomorphe à K^* (voir l'exercice 3.1). Donc si $\mathrm{GL}_n(\mathbb{R})$ et $\mathrm{GL}_n(\mathbb{C})$ étaient isomorphes, leurs centres $\mathbb{R}^*$ et $\mathbb{C}^*$ le seraient aussi, ce qui n'est pas. ◁

L'exercice suivant étudie encore l'existence d'un sous-groupe fini d'un certain type dans le groupe linéaire.

3.7. Sous-groupe de $\mathrm{GL}_n(\mathbb{R})$

Pour quelles valeurs de n existe-t-il un sous-groupe de $\mathrm{GL}_n(\mathbb{R})$ isomorphe au groupe additif $(\mathbb{Z}/4\mathbb{Z})^2$?

(École normale supérieure)

▷ **Solution.**

Une première remarque s'impose : si une valeur de n convient, toute valeur plus grande convient aussi puisque si $n \leqslant m$, $\mathrm{GL}_n(\mathbb{R})$ peut s'identifier au sous-groupe de $\mathrm{GL}_m(\mathbb{R})$ formé des matrices diagonales par blocs de la forme $\begin{pmatrix} \mathrm{M} & 0 \\ 0 & \mathrm{I}_{m-n} \end{pmatrix}$ où $\mathrm{M} \in \mathrm{GL}_n(\mathbb{R})$. On cherche donc, si elle existe, la plus petite valeur de n qui convient.

Analysons un peu ce que l'on cherche. Le groupe $\mathbb{Z}/4\mathbb{Z}$ est cyclique d'ordre 4. Trouver un sous-groupe isomorphe à $\mathbb{Z}/4\mathbb{Z}$ revient à trouver un élément d'ordre 4. On peut voir le groupe $(\mathbb{Z}/4\mathbb{Z})^2$ comme deux copies « indépendantes » de $\mathbb{Z}/4\mathbb{Z}$: il est engendré par deux éléments d'ordre 4 (à savoir $(1,0)$ et $(0,1)$) qui commutent, les deux sous-groupes engendrés par ces deux éléments n'ayant en commun que l'élément neutre. Réciproquement, si A, B sont deux matrices de $\mathrm{GL}_n(\mathbb{R})$ d'ordre 4, qui commutent et telles que $\langle \mathrm{A}\rangle \cap \langle \mathrm{B}\rangle = \{\mathrm{I}_n\}$ (où $\langle \mathrm{A}\rangle = \{\mathrm{I}_n, \mathrm{A}, \mathrm{A}^2, \mathrm{A}^3\}$ désigne le groupe engendré par A), alors le groupe engendré par A et B est isomorphe à $(\mathbb{Z}/4\mathbb{Z})^2$: en effet, il est facile de vérifier que $\langle \mathrm{A}, \mathrm{B}\rangle = \{\mathrm{A}^k\mathrm{B}^\ell,\ 0 \leqslant k, \ell \leqslant 3\}$ et que $\psi : (\mathbb{Z}/4\mathbb{Z})^2 \to \langle \mathrm{A}, \mathrm{B}\rangle$ qui à (k, ℓ) associe $\mathrm{A}^k\mathrm{B}^\ell$ est bien définie et est un isomorphisme de groupes. On est donc ramené à chercher deux matrices A et B vérifiant les propriétés énumérées ci-dessus.

Pour commencer étudions à quelle condition une matrice M de $\mathrm{GL}_n(\mathbb{R})$ est d'ordre 4. Cela signifie que $\mathrm{M}^4 = \mathrm{I}$ et que $\mathrm{M}^2 \neq \mathrm{I}$. La matrice M est diagonalisable dans $\mathcal{M}_n(\mathbb{C})$ et son spectre est inclus dans l'ensemble $\{1, -1, i, -i\}$ des racines quatrièmes de l'unité. Son spectre contient nécessairement i ou $-i$ car sinon M serait d'ordre 2. Par ailleurs, comme M est à coefficients réels, si i est valeur propre alors $-i$ aussi et avec la même multiplicité. On doit donc déjà avoir $n \geqslant 2$. Dans le cas $n = 2$, la matrice réelle $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$, qui a pour polynôme caractéristique $\mathrm{X}^2 + 1$, est bien d'ordre 4, diagonalisable sur $\mathbb{C}$, et semblable à la matrice $\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}$. À partir de cette matrice il est facile de trouver une solution dans le cas $n = 4$. Il suffit en effet de prendre les deux matrices

$$\mathrm{A} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix} \quad \text{et} \quad \mathrm{B} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Elles sont toutes les deux d'ordre 4, commutent, et vérifient bien la condition $\langle \mathrm{A}\rangle \cap \langle \mathrm{B}\rangle = \{\mathrm{I}_4\}$.

Pour conclure l'exercice, on va prouver qu'on ne peut pas trouver de solution dans le cas $n = 3$. Supposons par l'absurde qu'existent A et B dans $\mathrm{GL}_3(\mathbb{R})$ vérifiant les propriétés ci-dessus. Comme on vient de le voir i et $-i$ sont valeurs propres de A et comme elle doivent avoir le même ordre de multiplicité celui-ci vaut 1. Notons $\varepsilon = \pm 1$ la troisième valeur propre de A. Le polynôme minimal de A est donc $(\mathrm{X} - \varepsilon)(\mathrm{X}^2 + 1)$ et le lemme des noyaux montre que $\mathbb{R}^3$ est somme directe de la droite $\mathrm{Ker}(\mathrm{A} - \varepsilon \mathrm{I}_3)$ et du plan $\mathrm{Ker}(\mathrm{A}^2 + \mathrm{I}_3)$. Ce qui vient d'être dit pour A est aussi valable pour B. Notons $\varepsilon' \in \{\pm 1\}$ la valeur propre réelle de B, i et $-i$ étant ses deux autres valeurs propres. Comme A et B commutent, la

droite $\mathrm{Ker}(A-\varepsilon I_3)$ est stable par B et il s'agit donc d'une droite propre, nécessairement pour la valeur propre réelle ε'. Donc $\mathrm{Ker}(A-\varepsilon I_3) = \mathrm{Ker}(B-\varepsilon' I_3)$. De même le plan $\mathrm{Ker}(A^2+I_3)$ est stable par B. Les valeurs propres complexes de l'endomorphisme induit par B dans ce plan sont forcément i et $-i$ de sorte que $\mathrm{Ker}(B^2+I_3) = \mathrm{Ker}(A^2+I_3)$. Comme $\varepsilon^2 = \varepsilon'^2 = 1$, cela implique que $A^2 = B^2$, ce qui contredit l'hypothèse $\langle A\rangle \cap \langle B\rangle = \{I_n\}$. ◁

Si G est un groupe, on appelle exposant de G le ppcm des ordres des éléments de G si celui-ci est défini ou $+\infty$ si les ordres des éléments de G n'ont aucun multiple commun. Un groupe infini peut avoir un exposant fini : c'est par exemple le cas du groupe additif $(\mathbb{Z}/2\mathbb{Z})^{\mathbb{N}}$ qui est d'exposant 2. Un célèbre problème posé par Burnside en 1902 demande si un groupe d'exposant fini et de type fini (c'est-à-dire engendré par une partie finie) est nécessairement fini. La réponse est négative et un contre-exemple a été trouvé, mais seulement en 1975. L'exercice suivant consiste à prouver qu'un sous-groupe du groupe linéaire $\mathrm{GL}_n(\mathbb{C})$ qui est d'exposant fini est fini.

3.8. Un théorème de Burnside

1. Soit $A \in \mathcal{M}_n(\mathbb{C})$ telle que $\mathrm{Tr}(A^k) = 0$ pour tout $k \in \mathbb{N}^*$. Montrer que A est nilpotente.

2. Soit G un sous-groupe de $\mathrm{GL}_n(\mathbb{C})$, $(M_i)_{1\leqslant i<m} \in G^m$ une base de $\mathrm{Vect}(G)$ et $f : G \to \mathbb{C}^m$ l'application qui à $A \in G$ associe $(\mathrm{Tr}(AM_i))_{1\leqslant i\leqslant m}$. Montrer que si $f(A) = f(B)$ alors $AB^{-1} - I$ est nilpotente.

3. On suppose que toutes les matrices de G sont diagonalisables. Montrer que f est injective.

4. En déduire qu'un sous-groupe de $\mathrm{GL}_n(\mathbb{C})$ d'exposant fini (c'est-à-dire qu'il existe un entier N tel que $A^N = I$ pour toute matrice A du groupe) est fini. *C'est le théorème de Burnside.*

(École normale supérieure)

▷ **Solution.**

1. Ce résultat est classique. On en trouvera plusieurs preuves dans l'exercice 2.28 du chapitre réduction.

2. Par linéarité de la trace, on a $\mathrm{Tr}(AM) = \mathrm{Tr}(BM)$ pour toute matrice $M \in \mathrm{Vect}(G)$ et en particulier pour toute matrice M de G. Posons $D = AB^{-1}$. Cette matrice est dans G donc, pour tout $k \in \mathbb{N}^*$,

$$\mathrm{Tr}(D^k) = \mathrm{Tr}(AB^{-1}D^{k-1}) = \mathrm{Tr}(BB^{-1}D^{k-1}) = \mathrm{Tr}(D^{k-1}).$$

Il en résulte donc que pour tout $k \in \mathbb{N}$, $\mathrm{Tr}\, D^k = \mathrm{Tr}\, I_n = n$. Ainsi, pour $k \geqslant 1$,

$$\mathrm{Tr}(\mathrm{D}-\mathrm{I}_n)^k = \mathrm{Tr}\left(\sum_{j=0}^{k} \mathrm{C}_k^j (-1)^j \mathrm{D}^{k-j}\right) = n\sum_{j=0}^{k} \mathrm{C}_k^j (-1)^j = n(1-1)^k = 0.$$

La question précédente permet de conclure que $\mathrm{D}-\mathrm{I}_n$ est nilpotente.

3. Supposons de plus que les éléments de G sont diagonalisables et reprenons les notations précédentes. La matrice $\mathrm{D} = \mathrm{AB}^{-1}$ est dans G. Elle est donc diagonalisable. Mais alors $\mathrm{D}-\mathrm{I}$ est aussi diagonalisable! Comme elle est nilpotente d'après ce qui précède, elle est nulle. Donc $\mathrm{D} = \mathrm{I}$, c'est-à-dire $\mathrm{A} = \mathrm{B}$. Conclusion : f est injective.

4. Prenons pour G un sous-groupe de $\mathrm{GL}_n(\mathbb{C})$ d'exposant fini N. Toute matrice A de G est annulée par le polynôme $\mathrm{X}^{\mathrm{N}}-1$ qui est scindé à racines simples. Donc toute matrice de G est diagonalisable. Le résultat précédent s'applique et l'application $f : \mathrm{G} \to \mathbb{C}^m$ définie dans la question 2 est injective. En réalité, l'image de f est incluse dans X^m où X est l'ensemble des traces des éléments de G. Pour conclure, il suffit donc de prouver que X est fini. Or, vu ce qui précède, les valeurs propres des éléments de G appartiennent à l'ensemble fini des racines N-ièmes de l'unité. Donc X est fini. ◁

L'exercice suivant prouve que le groupe linéaire $\mathrm{GL}_n(\mathbb{C})$ ne contient pas de sous-groupe borné non trivial arbitrairement petit.

3.9. Petits sous-groupes de $\mathrm{GL}_n(\mathbb{C})$

On munit $\mathcal{M}_n(\mathbb{C})$ de la norme triple subordonnée à une norme de $\mathbb{C}^n$. Soit G un sous-groupe borné de $\mathrm{GL}_n(\mathbb{C})$.

1. Soit $\mathrm{M} \in \mathrm{G}$. Montrer que les valeurs propres de M sont de module 1 et que M est diagonalisable.

2. On suppose que $\mathrm{G} \subset \mathrm{B}(\mathrm{I}_n, \sqrt{2})$ (boule ouverte de centre l'identité et de rayon $\sqrt{2}$). Montrer que G est réduit à $\{\mathrm{I}_n\}$.

3. Montrer que le résultat reste vrai si on remplace $\sqrt{2}$ par $\sqrt{3}$.

(École polytechnique)

▷ **Solution.**

1. On notera $\|\ \|$ la norme de $\mathbb{C}^n$ utilisée et $|||\ |||$ la norme triple de $\mathcal{M}_n(\mathbb{C})$ associée. Rappelons pour commencer que si $\mathrm{A} \in \mathcal{M}_n(\mathbb{C})$ est une matrice quelconque et λ une valeur propre de A, alors $|\lambda| \leqslant |||\mathrm{A}|||$. En effet, si X est un vecteur propre associé à la valeur propre λ on a $\mathrm{AX} = \lambda \mathrm{X}$ donc $|\lambda|\|\mathrm{X}\| = \|\lambda \mathrm{X}\| = \|\mathrm{AX}\| \leqslant |||\mathrm{A}|||\ \|\mathrm{X}\|$. Comme X n'est pas nul, on peut simplifier $\|\mathrm{X}\|$ pour obtenir l'inégalité annoncée.

Comme G est un groupe borné, il en résulte que l'ensemble des valeurs propres des éléments de G est borné. Soit $\mathrm{M} \in \mathrm{G}$ et λ une valeur propre de G. Pour tout $p \in \mathbb{Z}$, M^p est dans G et a pour valeur propre λ^p. D'après ce qu'on vient de dire, la suite $(\lambda^p)_{p\in\mathbb{Z}}$ est bornée et cela n'a lieu que si

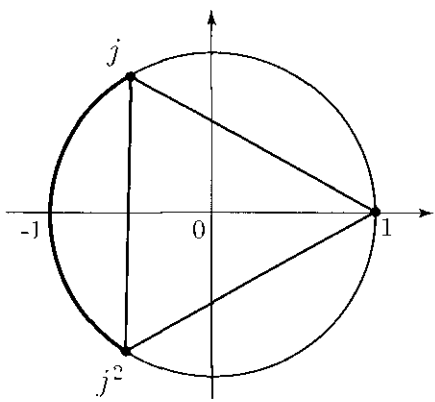
λ est de module 1 (λ n'est pas nul puisque M est inversible). Les valeurs propres de M sont donc toutes de module 1. Montrons maintenant que M est diagonalisable. Cela va encore découler du fait que la suite (M^p) est bornée et de la décomposition de Dunford (voir l'exercice 2.40). La matrice M s'écrit (de manière unique) $M = D + N$ avec D diagonalisable, N nilpotente et $ND = DN$. Soit s l'indice de nilpotence de N. On va montrer que $N = 0$, c'est-à-dire que $s = 1$. Supposons $s \geqslant 2$. Alors $\mathrm{Ker}\, N$ est strictement inclus dans $\mathrm{Ker}\, N^2$. Prenons X dans $\mathrm{Ker}\, N^2 \setminus \mathrm{Ker}\, N$. Comme D et N commutent, on a pour tout $p \geqslant s$,

$$M^p = (D+N)^p = D^p + pD^{p-1}N + C_p^2 D^{p-2}N^2 + \cdots + C_p^{s-1} D^{p-s+1} N^{s-1}.$$

En particulier, $M^pX = D^pX + pD^{p-1}NX$. Il en découle que la suite $(M^pX)_{p\in\mathbb{N}}$ n'est pas bornée dans $\mathbb{C}^n$, ce qui est absurde car la suite $(M^p)_{p\in\mathbb{N}}$ est bornée.

2. On suppose que $G \subset B(I_n, \sqrt{2})$. Soit $M \in G$. On va montrer que M est égal à la matrice identité et pour cela, il nous suffit de prouver que sa seule valeur propre est 1 (car M est diagonalisable d'après la question précédente). Soit λ dans le spectre de M. On a $|\lambda - 1| \leqslant \|M - I_n\| < \sqrt{2}$ par hypothèse. Comme $M^p \in G$ pour tout $p \in \mathbb{Z}$, on doit aussi avoir $|\lambda^p - 1| < \sqrt{2}$ pour tout $p \in \mathbb{Z}$. Il suffit de prouver que cela n'est le cas que si $\lambda = 1$. Géométriquement, cela consiste à prouver que l'une des puissances de λ au moins possède une partie réelle négative. Posons $\lambda = e^{i\theta}$ avec $0 < \theta < \dfrac{\pi}{2}$ (on peut se contenter de traiter ce cas quitte à prendre le conjugué de λ c'est-à-dire son inverse, ce qui revient juste à remplacer M par M^{-1}). Il existe clairement un entier naturel p tel que $\dfrac{\pi}{2} \leqslant p\theta < \pi$ car $\dfrac{\pi}{2\theta} > 1$. On a alors $\cos(p\theta) \leqslant 0$ et $|1 - \lambda^p| \geqslant \sqrt{2}$.

3. On raisonne comme dans la question précédente. Il faut juste changer l'argument final pour montrer que si λ est un complexe de module 1 distinct de 1, il existe au moins un entier p tel $|1 - \lambda^p| \geqslant \sqrt{3}$. Les deux complexes du cercle unité qui sont à distance $\sqrt{3}$ de 1 sont j et j^2, les racines cubiques non triviales de 1.



On doit donc montrer que l'une des puissances de λ tombe dans le petit arc de cercle délimité par j et j^2. Posons $\lambda = e^{i\theta}$ avec $0 < \theta \leqslant \pi$ (on peut se limiter à ce cas quitte à remplacer λ par son inverse). Si $\theta \geqslant \dfrac{2\pi}{3}$ c'est fini car on a directement $|\lambda - 1| \geqslant \sqrt{3}$. Supposons $0 < \theta < \dfrac{2\pi}{3}$. Il existe un entier p tel que $\dfrac{2\pi}{3} \leqslant p\theta < \dfrac{4\pi}{3}$ car $\dfrac{2\pi}{3\theta} > 1$. D'où le résultat. ◁

Si on remplace $\sqrt{3}$ par une valeur plus grande, le résultat n'est plus forcément vrai comme le montre le sous-groupe de $\mathrm{GL}_1(\mathbb{C})$ égal à $\{1, j, j^2\}$. En effet, la norme sur $\mathbb{C}$ étant le module, ce groupe est inclus dans la boule fermée de centre 1 et de rayon $\sqrt{3}$.

Si G est un groupe, on appelle commutateur de G tout élément de la forme $ghg^{-1}h^{-1}$. Le sous-groupe engendré par les commutateurs est appelé le groupe dérivé de G.

3.10. Groupe dérivé de $\mathrm{GL}_n(\mathrm{K})$

Soit $n \geqslant 2$ et K un corps possédant au moins 3 éléments. Montrer que le sous-groupe de $\mathrm{GL}_n(\mathrm{K})$ engendré par tous les commutateurs $\mathrm{ABA}^{-1}\mathrm{B}^{-1}$ avec $(\mathrm{A}, \mathrm{B}) \in \mathrm{GL}_n(\mathrm{K})^2$, est égal à $\mathrm{SL}_n(\mathrm{K})$.

(École polytechnique)

▷ **Solution.**

Notons tout d'abord que pour tout $(\mathrm{A}, \mathrm{B}) \in \mathrm{GL}_n(\mathrm{K})^2$ on a $\det(\mathrm{ABA}^{-1}\mathrm{B}^{-1}) = 1$ par multiplicativité du déterminant. Le groupe engendré par les commutateurs est donc inclus dans $\mathrm{SL}_n(\mathrm{K})$. On sait que ce dernier est engendré par les matrices de transvection (cf. exercice 3.1). On va donc essayer de voir si toute matrice de transvection est un commutateur. Pour $i \neq j$ et λ dans K on pose $\mathrm{T}_{ij}(\lambda) = \mathrm{I}_n + \lambda \mathrm{E}_{ij}$. On a alors $\mathrm{T}_{ij}(\lambda)^{-1} = \mathrm{T}_{ij}(-\lambda)$. Calculons le commutateur défini par une matrice de dilatation $\mathrm{D}_i(a)$ avec $a \notin \{0, 1\}$ (il en existe car K contient au moins 3 éléments) et une matrice de transvection $\mathrm{T}_{ij}(b) = \mathrm{I}_n + b\mathrm{E}_{ij}$. On a

$$\mathrm{D}_i(a)(\mathrm{I}_n + b\mathrm{E}_{ij})\mathrm{D}_i(a)^{-1} = \mathrm{I}_n + \mathrm{D}_i(a)b\mathrm{E}_{ij}\mathrm{D}_i(a)^{-1} = \mathrm{I}_n + ab\mathrm{E}_{ij} = \mathrm{T}_{ij}(ab),$$

et donc

$$\mathrm{D}_i(a)\mathrm{T}_{ij}(b)\mathrm{D}_i(a)^{-1}\mathrm{T}_{ij}(b)^{-1} = \mathrm{T}_{ij}(ab)\mathrm{T}_{ij}(-b) = \mathrm{T}_{ij}((a-1)b).$$

Lorsque b varie dans K, le scalaire $(a-1)b$ décrit aussi K. Donc toute matrice de transvection est un commutateur. Par suite, le groupe engendré par les commutateurs de $\mathrm{GL}_n(\mathrm{K})$ est égal à $\mathrm{SL}_n(\mathrm{K})$. ◁

Lorsque K *est le corps à deux éléments le résultat reste vrai pour* $n \geqslant 3$ *mais pas pour* $n = 2$. *On peut aussi montrer que le groupe dérivé de* $\mathrm{SL}_n(\mathrm{K})$ *est* $\mathrm{SL}_n(\mathrm{K})$ *sauf si* $n = 2$ *et* $\mathrm{Card}(\mathrm{K}) \leqslant 3$.

3.11. Commutateur égal à $-\mathrm{I}_2$

Trouver une condition nécessaire et suffisante sur le corps K pour qu'il existe $(\mathrm{A}, \mathrm{B}) \in \mathrm{SL}_2(\mathrm{K})^2$ tel que $-\mathrm{I}_2 = \mathrm{ABA}^{-1}\mathrm{B}^{-1}$.

(École normale supérieure)

▷ **Solution.**

Montrer qu'une matrice inversible M est un commutateur, c'est-à-dire de la forme $\mathrm{ABA}^{-1}\mathrm{B}^{-1}$ avec A, B inversibles, revient à prouver qu'elle s'écrit $\mathrm{M} = \mathrm{AC}$ avec C semblable à A^{-1}. Il est alors facile de montrer que la matrice $-\mathrm{I}_2$ est un commutateur de $\mathrm{GL}_2(\mathrm{K})$ pour tout corps K puisqu'il suffit d'écrire $-\mathrm{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ les deux matrices de ce produit étant clairement semblables entre elles et semblables à leurs inverses. La difficulté de l'exercice tient au fait qu'on veut un commutateur de $\mathrm{SL}_2(\mathrm{K})$: il faut non seulement que A soit de déterminant 1 (ce qui n'est pas le cas ci-dessus) mais aussi que la matrice de passage B telle que $\mathrm{C} = \mathrm{BA}^{-1}\mathrm{B}^{-1}$ soit de déterminant 1.

• On peut noter que les corps de caractéristique 2 conviennent. En effet, dans ce cas $-\mathrm{I}_2 = \mathrm{I}_2$ et il suffit de prendre $\mathrm{A} = \mathrm{B} = \mathrm{I}_2$. On supposera dans la suite que la caractéristique de K n'est pas 2. On va chercher une condition nécessaire en supposant qu'il existe un couple $(\mathrm{A}, \mathrm{B}) \in \mathrm{SL}_2(\mathrm{K})^2$ tel que $-\mathrm{I}_2 = \mathrm{ABA}^{-1}\mathrm{B}^{-1}$. Cette relation est équivalente à $\mathrm{AB} + \mathrm{BA} = 0$, forme qui a le mérite de nous éviter de calculer les inverses. On peut aussi noter que $-\mathrm{A} = \mathrm{B}^{-1}\mathrm{AB}$, c'est-à-dire que A et $-\mathrm{A}$ sont semblables (et de même pour B). En particulier on a $\mathrm{Tr}\,\mathrm{A} = -\mathrm{Tr}\,\mathrm{A}$ et donc $\mathrm{Tr}\,\mathrm{A} = 0$ puisque $\mathrm{caract}(\mathrm{K}) \neq 2$. De même $\mathrm{Tr}\,\mathrm{B} = 0$. Posons alors $\mathrm{A} = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ et $\mathrm{B} = \begin{pmatrix} x & y \\ z & -x \end{pmatrix}$. La relation $\mathrm{AB} + \mathrm{BA} = 0$ est équivalente à l'égalité $2ax + bz + cy = 0$ (1). On a par ailleurs $a^2 + bc = -1$ (2) et $x^2 + yz = -1$ (3) car $\det \mathrm{A} = \det \mathrm{B} = 1$. Supposons $b \neq 0$. On extrait z de la relation (1) en fonction de x et y et on le remplace dans l'égalité (3). Il vient $x^2 - \dfrac{2a}{b}xy - \dfrac{c}{b}y^2 = -1$ et comme $\dfrac{c}{b} = -\dfrac{1}{b^2} - \dfrac{a^2}{b^2}$ par (2) on obtient finalement

$$x^2 - \frac{2a}{b}xy + \frac{a^2}{b^2}y^2 + \frac{y^2}{b^2} = \left(x - \frac{a}{b}y\right)^2 + \frac{y^2}{b^2} = -1.$$

Autrement dit, pour que A et B existent, il est nécessaire que -1 soit

une somme de deux carrés d'éléments de K. Dans le cas où $b = 0$ on a $-1 = a^2 = a^2 + 0^2$ qui est encore somme de deux carrés.

• Montrons maintenant que cette condition est suffisante. Supposons que $-1 = \alpha^2 + \beta^2$ où α, β sont deux éléments de K. On cherche une solution aux équations (1), (2) et (3) avec $b = c$ et $y = z$. Prenons $a = \alpha$ et $b = c = \beta$, c'est-à-dire $A = \begin{pmatrix} \alpha & \beta \\ \beta & -\alpha \end{pmatrix}$. La condition (1) s'écrit alors $2(\alpha x + \beta y) = 0$ et il suffit de prendre $x = \beta$ et $y = -\alpha$ pour la remplir. On a alors $B = \begin{pmatrix} \beta & -\alpha \\ -\alpha & -\beta \end{pmatrix}$. Ces deux matrices conviennent.

On peut noter qu'un corps de caractéristique 2 vérifie la condition trouvée puisqu'on a alors $-1 = 1 = 1^2 + 0^2$. On peut donc conclure que, dans tous les cas, $-I_2$ est un commutateur de $SL_2(K)$ si et seulement si -1 est somme de deux carrés. ◁

On peut montrer que si K *est un corps infini, toute matrice de* $SL_n(K)$ *est un commutateur de* $GL_n(K)$.

Nous abordons maintenant une série d'exercices consacrés à la recherche de morphismes de groupes ayant pour but ou pour source un groupe linéaire.

3.12. Morphismes de $(\mathbb{R}^*, \times)$ dans $(GL_n(\mathbb{C}), \times)$

Déterminer les morphismes $f : t \longmapsto (f_{ij}(t))_{1 \leqslant i,j \leqslant n}$ de $(\mathbb{R}^*, \times)$ dans $(GL_n(\mathbb{C}), \times)$ tels que chaque f_{ij} soit une fraction rationnelle.
(École normale supérieure)

▷ **Solution.**

• On va commencer par se ramener au cas où les fonctions f_{ij} sont polynomiales. Soit p un polynôme unitaire de degré minimal tel que, pour tout $(i, j) \in [\![1, n]\!]^2$, $g_{ij} = pf_{ij} \in \mathbb{C}[X]$. Soit $(i, j) \in [\![1, n]\!]^2$. On a, pour tout t et u réels non nuls,

$$f_{ij}(tu) = \sum_{k=1}^{n} f_{ik}(t) f_{kj}(u).$$

En dérivant par rapport à u, on obtient

$$t f'_{ij}(tu) = \sum_{k=1}^{n} f_{ik}(t) f'_{kj}(u)$$

puis en prenant $u = 1$,

$$t f'_{ij}(t) = \sum_{k=1}^{n} f_{ik}(t) f'_{kj}(1).$$

En remplaçant f_{ij} par $\dfrac{g_{ij}}{p}$, il vient

$$t\frac{g'_{ij}(t)}{p(t)} - t\frac{p'(t)g_{ij}(t)}{(p(t))^2} = \sum_{k=1}^{n} \frac{g_{ik}(t)}{p(t)} f'_{kj}(1)$$

et donc

$$t\frac{p'(t)g_{ij}(t)}{p(t)} = tg'_{ij}(t) - \sum_{k=1}^{n} g_{ik}(t)f'_{kj}(1).$$

On en déduit que p divise $\mathrm{X}p'g_{ij}$. Si α est une racine de p d'ordre de multiplicité $m \geqslant 1$, alors α est racine d'ordre $m-1$ de p'. Donc si α est non nul, il est racine de g_{ij} et cela pour tout couple (i,j). C'est impossible car $\mathrm{X} - \alpha$ diviserait tous les polynômes g_{ij} et pour tout $(i,j) \in [\![1,n]\!]^2$, $\dfrac{p}{\mathrm{X}-\alpha} f_{ij} \in \mathbb{C}[\mathrm{X}]$, ce qui est contraire à la définition de p. Ainsi 0 est la seule racine (éventuelle) de p. Il existe donc $m \in \mathbb{N}$ tel que $p(t) = t^m$. On a alors, pour $(i,j) \in [\![1,n]\!]^2$, t et u dans $\mathbb{R}^*$,

$$g_{ij}(tu) = p(tu)f_{ij}(tu) = t^m u^m \sum_{k=1}^{n} f_{ik}(t)f_{ki}(u) = \sum_{k=1}^{n} g_{ik}(t)g_{ki}(u).$$

Autrement dit, l'application $g : t \mapsto (g_{ij}(t))_{1 \leqslant i,j \leqslant n}$ est un morphisme de $\mathbb{R}^*$ dans $\mathrm{GL}_n(\mathbb{C})$ dont les coefficients sont tous polynomiaux en t.

• Notons d le maximum des degrés des polynômes g_{ij}. Il existe des matrices $\mathrm{A}_0, \ldots, \mathrm{A}_d$ dans $\mathcal{M}_n(\mathbb{C})$ telles que, pour tout $t \in \mathbb{R}^*$,

$$g(t) = \sum_{k=0}^{d} t^k \mathrm{A}_k.$$

Une telle écriture est unique car, en posant $\mathrm{A}_k = (a_{ij}^k)$, elle équivaut pour tout $(i,j) \in [\![1,n]\!]^2$ à

$$g_{ij}(t) = \sum_{k=0}^{d} a_{ij}^k t^k,$$

pour tout $t \neq 0$ et donc à l'égalité formelle $g_{ij} = \sum_{k=0}^{d} a_{ij}^k \mathrm{X}^k$.

On a, pour t et u dans $\mathbb{R}^*$,

$$g(tu) = \sum_{k=0}^{d} t^k u^k \mathrm{A}_k = g(t)g(u) = \sum_{k=0}^{d} t^k \mathrm{A}_k g(u)$$

et donc, par unicité de l'écriture, pour $u \neq 0$ et $k \in [\![1,n]\!]$

$$u^k \mathrm{A}_k = \mathrm{A}_k g(u) = \sum_{\ell=0}^{d} u^\ell \mathrm{A}_k \mathrm{A}_\ell.$$

On en déduit que $\mathrm{A}_k^2 = \mathrm{A}_k$ et que $\mathrm{A}_k\mathrm{A}_\ell = 0$ si $\ell \neq k$. On a de plus $g(1) = \sum_{k=0}^{d} \mathrm{A}_k = \mathrm{I}_n$.

• Réciproquement, soit m et d des entiers naturels, $A_0, \dots, A_d$ des éléments de $\mathcal{M}_n(\mathbb{C})$ tels que, pour tout $(k,\ell) \in [\![1,n]\!]^2$, $A_k^2 = A_k$ et $A_k A_\ell = 0$ si $k \neq \ell$, et $\sum_{k=0}^{d} A_k = I_n$. Considérons la fonction f de $\mathbb{R}^*$ dans $\mathcal{M}_n(\mathbb{C})$ définie par

$$f(t) = \frac{1}{t^m} \sum_{k=0}^{d} t^k A_k.$$

On vérifie facilement que $f(tu) = f(t)f(u)$ pour tout t et u non nuls. Comme $f(1) = I_n$ par hypothèse, on a $f(t)f\left(\frac{1}{t}\right) = f(1) = I_n$ pour t non nul, donc f est à valeurs dans $GL_n(\mathbb{C})$ et elle a toutes les propriétés voulues.

• Il reste juste à expliquer comment caractériser les familles de matrices A_k ayant les propriétés ci-dessus. Si elles existent, les A_k sont des matrices de projection. Pour tout $X \in \mathbb{C}^n$, on a $X = \sum_{k=0}^{d} A_k X$, donc $\mathbb{C}^n = \sum_{k=0}^{d} \operatorname{Im} A_k$. Si, de plus, il existe des éléments $X_0, \dots, X_d$ de $\mathbb{C}^n$ tels que $X = \sum_{k=0}^{d} A_k X_k$, on a alors $A_\ell X = A_\ell X_\ell$, ce qui montre l'unicité de la décomposition. On a donc $\mathbb{C}^n = \bigoplus_{k=0}^{d} \operatorname{Im} A_k$.

Réciproquement, toute décomposition de $\mathbb{C}^n$ en somme directe de sous-espaces fournit des matrices A_k qui conviennent, chacune de ces matrices étant la matrice d'une projection sur un des sous-espaces, parallèlement à la somme des autres. ◁

Dans l'exercice suivant on s'intéresse à des morphismes dont le groupe linéaire est la source. Un morphisme étant caractérisé par sa restriction à une partie génératrice, on va à nouveau utiliser le résultat de l'exercice 3.1.

3.13. Morphismes de $GL_n(K)$ dans un groupe abélien fini

Soit $n \geqslant 2$, G un groupe abélien fini et K un corps commutatif. Étudier les morphismes de $GL_n(K)$ dans G. On étudiera le cas $K = \mathbb{C}$ puis $K = \mathbb{R}$ et enfin $K = \mathbb{Z}/q\mathbb{Z}$ avec q premier.

(École normale supérieure)

▷ **Solution.**

Le groupe G sera noté multiplicativement et son neutre sera noté 1.

On rappelle que le groupe linéaire $\mathrm{GL}_n(\mathrm{K})$ est engendré par l'ensemble des matrices de dilatation et des matrices de transvection. Plus précisément, toute matrice M de $\mathrm{GL}_n(\mathrm{K})$ s'écrit $\mathrm{M} = \mathrm{B}_1 \dots \mathrm{B}_k \mathrm{A} \mathrm{B}'_1 \dots \mathrm{B}'_l$, où A est la matrice de dilation $\mathrm{Diag}(1, 1, \dots, 1, \det \mathrm{M})$ et $\mathrm{B}_1, \dots, \mathrm{B}_k$, $\mathrm{B}'_1, \dots, \mathrm{B}'_l$ des matrices de transvection (voir l'exercice 3.1).

Soit f un morphisme de $\mathrm{GL}_n(\mathrm{K})$ dans G. On va s'intéresser à l'image par f des matrices de transvection et de dilatation. Plusieurs idées sont possibles.

• Une première idée est d'utiliser le caractère *fini* de G. Si le cardinal de G est p, on a, pour tout $g \in \mathrm{G}$, $g^p = 1$. On en déduit que pour tout $\mathrm{M} \in \mathrm{GL}_n(\mathrm{K})$, $f(\mathrm{M}^p) = (f(\mathrm{M}))^p = 1$. Autrement dit, toutes les puissances p-ièmes sont dans $\mathrm{Ker}\, f$. On peut alors remarquer que si $i \neq j$, on a

$$(\mathrm{I}_n + \lambda \mathrm{E}_{ij})(\mathrm{I}_n + \lambda' \mathrm{E}_{ij}) = \mathrm{I}_n + (\lambda + \lambda')\mathrm{E}_{ij}.$$

On en déduit que $(\mathrm{I}_n + \lambda \mathrm{E}_{ij})^p = \mathrm{I}_n + p\lambda \mathrm{E}_{ij}$, puis que, si le corps K est de caractéristique nulle, toute matrice de transvection $\mathrm{B} = \mathrm{I}_n + \lambda \mathrm{E}_{ij}$ est une puissance p-ième puisqu'on peut écrire $\mathrm{B} = \left(\mathrm{I}_n + \dfrac{\lambda}{p} \mathrm{E}_{ij}\right)^p$. On a donc $f(\mathrm{B}) = 1$.

Ainsi, si K est de caractéristique nulle, la restriction de f à $\mathrm{SL}_n(\mathrm{K})$ est triviale.

• En fait, cela est vrai plus généralement. Une seconde idée, est d'exploiter le caractère *abélien* de G. Si $\mathrm{M} = \mathrm{ABA}^{-1}\mathrm{B}^{-1}$ est un commutateur de $\mathrm{GL}_n(\mathrm{K})$ on a $f(\mathrm{M}) = f(\mathrm{A})f(\mathrm{B})f(\mathrm{A})^{-1}f(\mathrm{B})^{-1} = 1$ car G est commutatif. On a vu dans l'exercice 3.10 que si K n'est pas le corps réduit à deux éléments, toute transvection est un commutateur. On retrouve donc de cette manière le fait que la restriction de f à $\mathrm{SL}_n(\mathrm{K})$ est triviale.

• On va maintenant regarder l'image d'une matrice de dilatation. Comme le demande l'énoncé, on regarde d'abord le cas où $\mathrm{K} = \mathbb{C}$. Si $\mathrm{A} = \mathrm{Diag}(1, \dots, 1, \alpha)$ est une matrice de dilatation et β une racine p-ième de α, on peut écrire $\mathrm{A} = (\mathrm{Diag}(1, \dots, 1, \beta))^p$. On a donc $f(\mathrm{A}) = 1$ d'après ce qui a été dit plus haut. Comme les matrices de transvection et de dilatation engendrent $\mathrm{GL}_n(\mathbb{C})$, f est le morphisme trivial $f : \mathrm{M} \longmapsto 1$.

• Supposons maintenant que $\mathrm{K} = \mathbb{R}$.

Si p est impair, la démonstration précédente s'applique, puisqu'alors tout nombre réel possède une racine p-ième. On trouve pour seul morphisme le morphisme trivial.

Supposons que p est pair. Si $\alpha > 0$, la matrice $\mathrm{A} = \mathrm{Diag}(1, \dots, 1, \alpha)$ possède une racine p-ième et $f(\mathrm{A}) = 1$. Toute matrice M de $\mathrm{GL}_n(\mathbb{R})$ s'écrit $\mathrm{M} = \mathrm{B}_1 \dots \mathrm{B}_k \mathrm{A} \mathrm{B}'_1 \dots \mathrm{B}'_l$, où A est une matrice de dilation et $\mathrm{B}_1, \dots, \mathrm{B}_k$, $\mathrm{B}'_1, \dots, \mathrm{B}'_l$ des matrices de transvection. D'après ce qui précède, on a $f(\mathrm{M}) = f(\mathrm{A})$. Comme $\det \mathrm{M} = \det \mathrm{A} = \alpha$, si $\det \mathrm{M} > 0$, on a $f(\mathrm{M}) = 1$.

Si M et N sont deux matrices de $\mathrm{GL}_n(\mathbb{R})$ de déterminant négatif, on a, puisque M^2 et NM ont un déterminant positif,

$$f(\mathrm{N}) = f(\mathrm{N})f(\mathrm{M}^2) = f(\mathrm{NM}^2) = f(\mathrm{NM})f(\mathrm{M}) = f(\mathrm{M}).$$

Il existe donc $a \in \mathrm{G}$ tel que

$$f(\mathrm{M}) = \begin{cases} 1 & \text{si } \det \mathrm{M} > 0 \\ a & \text{si } \det \mathrm{M} < 0. \end{cases}$$

L'élément a de G vérifie $a^2 = 1$, car si $\det \mathrm{M} < 0$, $1 = f(\mathrm{M}^2) = a^2$.

Réciproquement si a est un élément de G qui vérifie $a^2 = 1$, toute application définie ainsi est un morphisme de $\mathrm{GL}_n(\mathbb{R})$ dans G.

Notons que puisque p est pair, il résulte du lemme de Cauchy (cf. exercice 2.10 du tome 1 d'algèbre) que G possède des éléments a d'ordre 2. Ainsi, il existe des morphismes non triviaux de $\mathrm{GL}_n(\mathbb{R})$ dans G.

• Si K est un corps quelconque différent de $\mathbb{Z}/2\mathbb{Z}$, on a toujours $f(\mathrm{A}) = 1$ pour toute matrice de transvection A. On obtient donc

$$f(\mathrm{M}) = f(\mathrm{Diag}(1, \ldots, 1, \det \mathrm{M}))$$

de sorte qu'il existe un morphisme φ de $(\mathrm{K}^*, \times)$ dans $(\mathrm{G}, \times)$ tel que $f = \varphi \circ \det$. L'existence d'un morphisme non trivial φ dépend de K et du groupe G.

• Traitons encore l'exemple de $\mathrm{K} = \mathbb{Z}/q\mathbb{Z}$ où q est un nombre premier impair. On est amené à chercher les morphismes de $(\mathbb{Z}/q\mathbb{Z})^*$ dans G. Or on sait que le groupe $(\mathbb{Z}/q\mathbb{Z})^*$ est cyclique (voir l'exercice 4.13 du tome 1 d'algèbre). Si u est un générateur de ce groupe, un morphisme φ est parfaitement déterminé par l'élément $a = \varphi(u)$. En effet, pour tout $x \in (\mathbb{Z}/q\mathbb{Z})^*$, il existe $k \in \mathrm{N}$ tel que $x = u^k$ et on a alors $\varphi(x) = a^k$.

Voyons quels sont les $a \in \mathrm{G}$ qui peuvent convenir. Comme $u^{q-1} = 1$ on doit avoir $a^{q-1} = \varphi(u^{q-1}) = \varphi(1) = 1$.

Si p est premier avec $q - 1$, on a forcément $a = 1$. En effet, l'ordre de a dans G doit diviser $q - 1$ et p. Dans ce cas, le seul morphisme entre $(\mathbb{Z}/q\mathbb{Z})^*$ et G est donc le morphisme trivial.

Si $q - 1$ et p ne sont pas premiers entre eux, a doit avoir pour ordre un diviseur du pgcd de p et $q - 1$. On peut trouver de tels éléments non triviaux dans G : si m un nombre premier qui divise p et $q - 1$, le lemme de Cauchy assure l'existence d'un élément $a \in \mathrm{G}$ d'ordre m. Il vérifie $a^{q-1} = 1$ et l'application φ définie par $\varphi(u^k) = a^k$ est un morphisme non trivial de $(\mathbb{Z}/q\mathbb{Z})^*$ dans G et $f = \varphi \circ \det$, un morphisme non trivial de $\mathrm{GL}_n(\mathrm{K})$ dans G.

• Regardons enfin le cas particulier $\mathrm{K} = \mathbb{Z}/2\mathbb{Z}$. Dans ce cas il n'y a pas de matrice de dilatation (hormis l'identité) et $\mathrm{GL}_n(\mathbb{Z}/2\mathbb{Z}) = \mathrm{SL}_n(\mathbb{Z}/2\mathbb{Z})$. Nous avons indiqué dans la remarque qui suit l'exercice 3.10 que pour $n \geqslant 3$, $\mathrm{GL}_n(\mathbb{Z}/2\mathbb{Z})$ est égal à son groupe dérivé. Il en découle que le

seul morphisme $f : \mathrm{GL}_n(\mathbb{Z}/2\mathbb{Z}) \to \mathrm{G}$ est le morphisme trivial. Il reste encore le cas $n = 2$. On a, outre I_n, deux matrices de transvection dans $\mathrm{GL}_2(\mathbb{Z}/2\mathbb{Z})$ à savoir $\mathrm{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ et $\mathrm{U} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Les autres éléments de $\mathrm{GL}_2(\mathbb{Z}/2\mathbb{Z})$ sont I_2, $\mathrm{TU} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, $\mathrm{UT} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$ et $\mathrm{UTU} = \mathrm{TUT} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. On pose $f(\mathrm{T}) = a$ et $f(\mathrm{U}) = b$. Comme $\mathrm{T}^2 = \mathrm{U}^2 = \mathrm{I}_2$, on a $a^2 = b^2 = 1$. Si p est impair, cela impose $a = b = 1$ et f est trivial. Supposons p pair dans la suite. De $f(\mathrm{TUT}) = f(\mathrm{UTU})$, on tire $aba = bab$ et donc $a = b$ (a et b commutent). Si $a \in \mathrm{G}$ et $a^2 = 1$, on vérifie que l'application de $\mathrm{GL}_2(\mathrm{K})$ définie par $f(\mathrm{T}) = f(\mathrm{U}) = f(\mathrm{TUT}) = a$ et $f(\mathrm{UT}) = f(\mathrm{TU}) = f(1) = 1$ est un morphisme de $\mathrm{GL}_2(\mathrm{K})$ dans G. Comme p est pair, G possède des éléments d'ordre 2, donc on peut trouver des morphismes non triviaux.

En fait $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ *est isomorphe au groupe symétrique* $\mathcal{S}_3$ *et si on identifie le groupe* $\{1, a\}$ *engendré par l'élément d'ordre* 2 *choisi au groupe multiplicatif* $\{\pm 1\}$, *le morphisme construit n'est autre que la signature.* ◁

Un cas particulier qui a aussi été posé directement à l'oral de l'École normale supérieure consiste à déterminer les morphismes de $\mathrm{GL}_n(\mathrm{K})$ *dans* K^* *lorsque* K *est un corps fini. Comme dans l'exercice on voit qu'un tel morphisme est de la forme* $\varphi(\det)$ *où* φ *est un morphisme de* K^* *dans lui-même. Comme* K^* *est cyclique,* φ *est de la forme* $x \mapsto x^k$ *où* k *est un entier qu'on peut choisir entre* 0 *et* $\operatorname{Card}\mathrm{K} - 1$.

Les exercices qui suivent concernent les matrices inversibles à coefficients entiers. Rappelons que $\mathrm{GL}_n(\mathbb{Z})$ *désigne le groupe des inversibles de l'anneau* $\mathcal{M}_n(\mathbb{Z})$. *Il est facile de voir que* $\mathrm{A} \in \mathcal{M}_n(\mathbb{Z})$ *est dans* $\mathrm{GL}_n(\mathbb{Z})$ *si et seulement si* $\det \mathrm{A} = \pm 1$. *On note enfin* $\mathrm{SL}_n(\mathbb{Z})$ *le sous-groupe de* $\mathrm{GL}_n(\mathbb{Z})$ *formé des matrices de déterminant* 1.

On sait que sur un corps K, *le groupe linéaire* $\mathrm{GL}_n(\mathrm{K})$ *est engendré par les matrices de transvection et les matrices de dilatation, et que le groupe* $\mathrm{SL}_n(\mathrm{K})$ *est engendré par les matrices de transvection. L'exercice suivant étudie la situation sur l'anneau* $\mathbb{Z}$ *dans le cas* $n = 2$.

3.14. Génération de $\mathrm{SL}_2(\mathbb{Z})$

1. Montrer que le groupe $\mathrm{SL}_2(\mathbb{Z})$ est engendré par les deux matrices $\mathrm{U} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ et $\mathrm{V} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$.

2. En déduire qu'il est aussi engendré par U et $T = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

(École normale supérieure)

▷ **Solution.**

1. Notons G le sous-groupe engendré par U et V. On vérifie par récurrence, que $U^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ pour tout $k \in \mathbb{Z}$. Par suite, en prenant la transposée, on a $V^k = \begin{pmatrix} 1 & 0 \\ k & 1 \end{pmatrix}$. Le groupe G contient donc toutes les matrices de transvection (à coefficients dans $\mathbb{Z}$). En particulier, si A est une matrice de G, toute matrice obtenue à partir de A en faisant des opérations élémentaires de la forme $L_i \leftarrow L_i + kL_j$ ou $C_i \leftarrow C_i + kC_j$, avec $\{i, j\} = \{1, 2\}$ et $k \in \mathbb{Z}$, est encore dans G. L'idée de la preuve est alors la suivante. Soit $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$. On va effectuer des opérations du type précédent sur A jusqu'à tomber sur une matrice A′ de transvection. Cela prouve que A est dans G puisqu'elle s'écrit alors comme un produit de matrices de transvection.

• Supposons d'abord que $ac \neq 0$. Comme $ad - bc = 1$, a et c sont premiers entre eux. Effectuons la division euclidienne de a par c : $a = q_1 c + r_1$, avec $0 \leqslant r_1 < |c|$. En effectuant l'opération $L_1 \leftarrow L_1 - q_1 L_2$ on obtient une matrice de la forme : $\begin{pmatrix} r_1 & \times \\ c & \times \end{pmatrix}$. Si r_1 n'est pas nul, on continue en effectuant la division euclidienne de c par r_1 : $c = q_2 r_1 + r_2$ avec $0 \leqslant r_2 < r_1$. Par une manipulation on arrive alors à la matrice $\begin{pmatrix} r_1 & \times \\ r_2 & \times \end{pmatrix}$. On continue l'algorithme d'Euclide jusqu'au premier reste nul. L'avant dernier reste est égal à $1 = \mathrm{pgcd}(a, c)$ et on tombe sur une matrice de la forme $\begin{pmatrix} 1 & b' \\ 0 & d' \end{pmatrix}$ ou de la forme $\begin{pmatrix} 0 & b' \\ 1 & d' \end{pmatrix}$. Dans le premier cas $d' = 1$, puisque le déterminant ne change pas au cours des manipulations, de sorte qu'il s'agit d'une matrice de G. Dans le second cas, on rajoute la deuxième ligne à la première, puis on retranche la première ligne à la deuxième pour se ramener au cas précédent.

• Si l'un des coefficients a ou c est nul, l'autre ne l'est pas (car $ad - bc = 1$) et il suffit d'ajouter la ligne qui le contient à l'autre ligne pour se ramener au cas précédent.

2. Il suffit de constater que UTU = V. ◁

La résultat de cet exercice a aussi été obtenu dans le tome 1 comme corollaire de l'étude du groupe modulaire (exercice 2.19).

Dans l'exercice 7.21 de ce même tome il est montré que les matrices de transvection $I_n + \lambda E_{ij}$ *avec* $\lambda \in \mathbb{Z}$ *et* $i \neq j$ *et les matrices de dilatation*

$\mathrm{I}_n - 2\mathrm{E}_{ii}$ *engendrent le groupe* $\mathrm{GL}_n(\mathbb{Z})$. *En adaptant un peu la solution de cet exercice on peut montrer que les matrices de transvection engendrent tout le groupe* $\mathrm{SL}_n(\mathbb{Z})$. *En fait, comme* $(\mathrm{I}_n + \mathrm{E}_{ij})^k = \mathrm{I}_n + k\mathrm{E}_{ij}$ *pour tout* $k \in \mathbb{Z}$, *les matrices* $\mathrm{I}_n + \mathrm{E}_{ij}$ *avec* $i \neq j$ *suffisent. Cela généralise donc le résultat de cet exercice.*

Le fait que $\mathrm{SL}_n(\mathbb{Z})$ *est de type fini, c'est-à-dire engendré par un nombre fini d'éléments, est utilisé dans l'exercice suivant.*

3.15. Endomorphismes surjectifs de $\mathrm{SL}_n(\mathbb{Z})$

Soit G un groupe de type fini (*i.e.* possédant un système fini de générateurs) et $f : \mathrm{G} \to \mathrm{G}$ un morphisme surjectif. Soit H un groupe fini et g un morphisme de G dans H. On veut montrer que $\mathrm{Ker}\, f \subset \mathrm{Ker}\, g$.

1. Montrer que l'ensemble des morphismes de G dans H est fini.

2. Soit $a \in \mathrm{Ker}\, f$. Montrer l'existence d'une suite (b_n) de G telle que $f^n(b_n) = a$.

3. On pose $g_n = g \circ f^n$. Montrer que pour $m > n$, $g_m(b_n) = e$.

4. Montrer que si $a \notin \mathrm{Ker}\, g$, alors $g_m \neq g_n$ pour $m > n$ et conclure.

5. Application. Montrer que $\mathrm{SL}_2(\mathbb{Z})$ est engendré par $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ et $\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$. Montrer que pour tout $\mathrm{A} \in \mathrm{SL}_2(\mathbb{Z})$, $\mathrm{A} \neq \mathrm{I}_2$, on peut trouver un groupe fini H et un morphisme g de $\mathrm{SL}_2(\mathbb{Z})$ dans H tels que $g(\mathrm{A})$ n'est pas l'élément neutre de H. Montrer que tout endomorphisme surjectif de $\mathrm{SL}_2(\mathbb{Z})$ est bijectif. Généraliser à $\mathrm{SL}_n(\mathbb{Z})$.

(École normale supérieure)

▷ **Solution.**

1. Soit S une partie génératrice finie de G. L'application qui à un morphisme $f : \mathrm{G} \to \mathrm{H}$ associe sa restriction à S est injective car deux morphismes qui coïncident sur une partie génératrice sont forcément égaux. Comme $\mathcal{F}(\mathrm{S}, \mathrm{H})$ est fini car S et H sont finis, il n'y a donc qu'un nombre fini de morphismes de G dans H.

2. Pour tout $n \in \mathbb{N}$, f^n est surjectif, donc il existe $b_n \in \mathrm{G}$ tel que $f^n(b_n) = a$.

3. Pour $m > n$, on a

$$g_m(b_n) = g(f^m(b_n)) = g(f^{m-n}(f^n(b_n))) = g(f^{m-n}(a)) = g(e) = e$$

car $m - n > 0$.

4. Par ailleurs, $g_n(b_n) = g(a)$. Donc si $g(a) \neq e$, on a $g_m \neq g_n$ pour $m > n$. Les morphismes g_m sont deux à deux distincts. Cela est

impossible à cause du résultat de la question 1. Il en résulte que $g(a) = e$ et donc que $\operatorname{Ker} f \subset \operatorname{Ker} g$.

5. Le fait que $\mathrm{SL}_2(\mathbb{Z})$ est engendré par $\mathrm{U} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ et $\mathrm{V} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$ est l'objet de l'exercice 3.14. Il s'agit donc d'un groupe de type fini. Soit A une matrice de $\mathrm{SL}_2(\mathbb{Z})$ distincte de l'identité. Si A n'est diagonale, on choisit p un nombre premier qui ne divise pas l'un des coefficients non diagonal de A et H le groupe fini $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. La réduction modulo p est un morphisme de groupes qui envoie A sur une matrice qui n'est pas l'identité. Si A est diagonale, forcément $\mathrm{A} = -\mathrm{I}$ et on prend par exemple la réduction modulo 3. Il découle alors des question précédente que si f est un endomorphisme surjectif de $\mathrm{SL}_2(\mathbb{Z})$, le noyau de f est inclus dans celui de g et donc ne contient pas A. Cela est vrai pour tout $\mathrm{A} \neq \mathrm{I}_2$ donc f a un noyau trivial et est bijectif.

La même argumentation s'applique au groupe $\mathrm{SL}_n(\mathbb{Z})$ qui est aussi de type fini d'après les remarques qui précèdent l'exercice. ◁

Dans l'énoncé suivant on montre que les sous-groupes finis de $\mathrm{SL}_2(\mathbb{Z})$ *sont tous cycliques.*

3.16. Sous-groupes finis de $\mathrm{SL}_2(\mathbb{Z})$

1. Soit G un sous-groupe fini de $\mathrm{Gl}_2(\mathbb{Z})$. Montrer qu'il existe un produit scalaire sur $\mathbb{R}^2$, stable par tous les éléments de G, ou si l'on préfère, tel que les éléments de G soient des endomorphismes orthogonaux pour ce produit scalaire.

2. En déduire que les sous-groupes finis de $\mathrm{SL}_2(\mathbb{Z})$ sont cycliques.

(École normale supérieure)

▷ **Solution.**

1. Notons $\langle\ ,\ \rangle$ le produit scalaire canonique sur $\mathbb{R}^2$. On utilise un argument classique consistant à moyenner les formes bilinéaires $(x, y) \mapsto \langle g(x), g(y)\rangle$ pour g parcourant G. Posons donc

$$\langle x, y\rangle_{\mathrm{G}} = \frac{1}{|\mathrm{G}|} \sum_{g \in \mathrm{G}} \langle g(x), g(y)\rangle$$

pour x, y dans $\mathbb{R}^2$. L'application $\langle\ ,\ \rangle_{\mathrm{G}}$ est clairement une forme bilinéaire symétrique. On a $\langle x, x\rangle_{\mathrm{G}} = \dfrac{1}{|\mathrm{G}|} \displaystyle\sum_{g \in \mathrm{G}} \|g(x)\|^2$ qui est positif, et nul seulement lorsque $x = 0$ car $\mathrm{Id} \in \mathrm{G}$. Il en résulte que $\langle\ ,\ \rangle_{\mathrm{G}}$ est un produit scalaire sur $\mathbb{R}^2$. Montrons que tous les éléments de G sont orthogonaux pour ce produit scalaire. Soit $g_0 \in \mathrm{G}$. On a, pour x, y dans $\mathbb{R}^2$,

$$\langle g_0(x), g_0(y)\rangle_G = \frac{1}{|G|}\sum_{g\in G}\langle (g\circ g_0)(x), (g\circ g_0)(y)\rangle$$
$$= \frac{1}{|G|}\sum_{h\in G}\langle h(x), h(y)\rangle = \langle x, y\rangle_G$$

car l'application $g \mapsto g \circ g_0$ est une permutation des éléments de G.

2. Le groupe orthogonal de $\mathbb{R}^2$ pour le produit scalaire $\langle\ ,\ \rangle_G$ est isomorphe à $\mathrm{O}_2(\mathbb{R})$ et comme les éléments de G sont de déterminant 1, G est isomorphe à un sous-groupe fini de $\mathrm{SO}_2(\mathbb{R})$ lui-même isomorphe au groupe multiplicatif S^1 des nombres complexes de module 1. Pour conclure, il nous suffit donc de prouver que tous les sous-groupes finis de S^1 sont cycliques. Soit H un tel sous-groupe et d son cardinal. Par le théorème de Lagrange, on a $z^d = 1$ pour tout $z \in \mathrm{H}$. Donc H est inclus dans le groupe U_d des racines d-ièmes de l'unité. Comme H possède d éléments, on a égalité $\mathrm{H} = \mathrm{U}_d$. Et il est bien connu que U_d est un groupe cyclique. ◁

L'exercice suivant montre qu'il n'y a qu'un nombre fini d'ordres possibles pour les éléments de $\mathrm{GL}_2(\mathbb{Z})$ *et donc un nombre fini de cardinaux possibles pour les sous-groupes cycliques de* $\mathrm{SL}_2(\mathbb{Z})$.

3.17. Ordre des éléments de $\mathrm{GL}_2(\mathbb{Z})$

Montrer que l'ordre de toute matrice A de $\mathrm{GL}_2(\mathbb{Z})$ est infini ou égal à 1, 2, 3, 4, ou 6. Montrer que dans chaque cas hormis 1, il existe une infinité de matrices de cet ordre.

(**École normale supérieure**)

▷ **Solution.**

Soit $\mathrm{A} \in \mathrm{GL}_2(\mathbb{Z})$ que l'on suppose distincte de l'identité. Dire que A est d'ordre fini revient à dire qu'il existe un entier $p \geqslant 2$ tel que $\mathrm{A}^p = \mathrm{I}_2$, ou si l'on préfère que A est annulée par le polynôme $\mathrm{X}^p - 1$. Cela est le cas si et seulement si $\mathrm{X}^p - 1$ est divisible par le polynôme minimal de A (en tant que matrice de $\mathcal{M}_2(\mathbb{Q})$). On sait que A est annulée par son polynôme caractéristique $\chi = \mathrm{X}^2 - t\mathrm{X} + d$ où $t = \mathrm{Tr}(\mathrm{A}) \in \mathbb{Z}$ et $d = \det(\mathrm{A}) = \pm 1$. Le polynôme χ est le polynôme minimal de A sauf dans le cas où A est annulée par un polynôme de degré 1, c'est-à-dire est une matrice scalaire. Comme on doit avoir $\det \mathrm{A} = \pm 1$, cela ne se produit que pour $\mathrm{A} = \mathrm{I}_2$, cas qu'on a exclu, et $\mathrm{A} = -\mathrm{I}_2$ qui est une matrice d'ordre 2. Supposons à partir de maintenant que $\mathrm{A} \neq \pm \mathrm{I}_2$. Notons que le polynôme χ divise $\mathrm{X}^p - 1$ si et seulement si il a deux racines distinctes qui sont des racines p-ièmes de l'unité. On va donc regarder les racines de χ en discutant selon les valeurs de d et t. Le discriminant de χ est $\Delta = t^2 - 4d$.

• Supposons pour commencer $d = -1$. Le discriminant Δ est alors strictement positif. Les racines de χ sont donc réelles. Or les seules racines réelles de l'unité sont -1 et 1. On a $\chi(-1) = t$. Donc si la trace t de A est non nulle, A est d'ordre infini. Si $t = 0$ alors $\chi = X^2 - 1$ et A est d'ordre 2. Montrons qu'il existe une infinité de matrices d'ordre 2 de ce type. On choisit le coefficient $(1, 1)$ égal à un entier n quelconque. Le coefficient $(2, 2)$ doit valoir $-n$ pour que la trace soit nulle. On complète pour avoir un déterminant égal à -1 en prenant par exemple la matrice $\begin{pmatrix} n & 1+n \\ 1-n & -n \end{pmatrix}$.

• Supposons maintenant $d = 1$ et $|t| \geqslant 3$. On a encore $\Delta > 0$ et les racines de χ sont réelles. Comme $\chi(-1) \neq 0$, la matrice A est d'ordre infini.

• Supposons $d = 1$ et $|t| = 2$. On a $\Delta = 0$ et χ admet une racine double. La matrice A est encore d'ordre infini. C'est le cas par exemple de toutes les matrices triangulaires $\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$ pour $n \geqslant 1$.

• Supposons $d = 1$ et $t = -1$. On a $\chi = X^2+X+1$. Ce polynôme divise $X^3 - 1$ et A est d'ordre 3. C'est le cas des matrices $\begin{pmatrix} n & n^2+n+1 \\ -1 & -n-1 \end{pmatrix}$ pour tout $n \in \mathbb{Z}$.

• Supposons $d = 1$ et $t = 0$. On a $\chi = X^2 + 1$ qui divise $X^4 - 1$ et A est d'ordre 4 (car $A^2 = -I_2 \neq I_2$). C'est le cas des matrices $\begin{pmatrix} n & n^2+1 \\ -1 & -n \end{pmatrix}$ pour tout $n \in \mathbb{Z}$.

• Supposons enfin $d = 1$ et $t = 1$. On a $\chi = X^2 - X + 1$ qui divise $X^6 - 1$ et A est d'ordre 6 (car χ ne divise ni $X^3 - 1$ ni $X^2 - 1$). C'est le cas des matrices $\begin{pmatrix} n & n^2+n-1 \\ -1 & -n+1 \end{pmatrix}$ pour tout $n \in \mathbb{Z}$.

Ainsi, l'ordre d'une matrice de $GL_2(\mathbb{Z})$ appartient à l'ensemble $\{1, 2, 3, 4, 6, \infty\}$ et dans chacun des cas (hormis 1) on a bien une infinité de matrices ayant cet ordre. ◁

L'énoncé suivant montre que le même phénomène se produit en toute dimension : les ordres finis possibles des éléments de $GL_n(\mathbb{Z})$ sont majorés.

3.18. Ordre des éléments de $GL_n(\mathbb{Z})$

Établir l'existence de $r \in \mathbb{N}^*$ tel que pour toute matrice M de $\mathcal{M}_n(\mathbb{Z})$ pour laquelle il existe $m \in \mathbb{N}^*$ vérifiant $M^m = I_n$, on ait $M^r = I_n$.

(École normale supérieure)

▷ **Solution.**

Si M est une matrice de $\mathcal{M}_n(\mathbb{Z})$ vérifiant $\mathrm{M}^m = \mathrm{I}_n$ pour un certain entier $m \in \mathbb{N}^*$ on a $(\det \mathrm{M})^m = 1$, donc $\det \mathrm{M} = \pm 1$, et ainsi M est un élément du groupe $\mathrm{GL}_n(\mathbb{Z})$, qui plus est, un élément d'ordre fini. On cherche donc à montrer qu'il existe un entier $r \in \mathbb{N}^*$ tel que toute matrice d'ordre fini de $\mathrm{GL}_n(\mathbb{Z})$ ait un ordre qui divise r. L'exercice précédent a par exemple montré que pour $n = 2$ on peut prendre $r = 12$.

On va montrer l'existence d'un tel entier r dans le cas général, mais sans chercher à le préciser. Soit $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$ d'ordre fini et $m \in \mathbb{N}^*$ tel que $\mathrm{M}^m = \mathrm{I}_n$. Le polynôme $\mathrm{X}^m - 1$ annule M donc les valeurs propres complexes de M sont des racines de l'unité. Par ailleurs le polynôme caractéristique de M est à coefficients dans $\mathbb{Z}$. Or l'ensemble E des polynômes unitaires de degré n, à coefficients dans $\mathbb{Z}$, et dont les racines sont des racines de l'unité est fini (car les relations coefficients-racines montrent que les coefficients d'un élément $\mathrm{P} \in \mathrm{E}$ ne peuvent prendre qu'un nombre fini de valeurs). Il nous suffit de prendre pour r le ppcm des ordres (dans $\mathbb{C}^*$) de toutes les racines de tous les polynômes de E. Si $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$ est d'ordre fini quelconque on a $\lambda^r = 1$ pour toute valeur propre λ de M et, comme M est diagonalisable dans $\mathcal{M}_n(\mathbb{C})$, on a $\mathrm{M}^r = \mathrm{I}_n$. D'où le résultat. ◁

Avec quelques connaissances sur les polynômes cyclotomiques (leur irréductibilité sur $\mathbb{Q}$ et la factorisation sur $\mathbb{Q}$ des polynômes $\mathrm{X}^n - 1$) il est possible de caractériser plus précisément les ordres possibles des éléments de $\mathrm{GL}_n(\mathbb{Z})$.

L'exercice suivant étudie les éléments d'ordre 4 ou 3 de $\mathrm{SL}_2(\mathbb{Z})$. La première question reprend des choses déjà vues mais les questions suivantes sont difficiles.

3.19. Éléments d'ordre 4 ou 3 dans $\mathrm{SL}_2(\mathbb{Z})$

Soit $\mathrm{M} \in \mathrm{SL}_2(\mathbb{Z})$.

1. On suppose qu'il existe $n \in \mathbb{N}^*$ tel que $\mathrm{M}^n = \mathrm{I}_2$. Montrer que l'ordre de M dans $\mathrm{SL}_2(\mathbb{Z})$ est au plus 6.

2. On suppose M d'ordre 4. Montrer que M est conjuguée dans $\mathrm{GL}_2(\mathbb{Z})$ à la matrice $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

3. On suppose M d'ordre 3. Montrer que M est conjuguée dans $\mathrm{GL}_2(\mathbb{Z})$ à la matrice $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$.

(École normale supérieure)

▷ **Solution.**

Si M est une matrice de $\mathcal{M}_n(\mathbb{Z})$ vérifiant $\mathrm{M}^m = \mathrm{I}_n$ pour un certain entier $m \in \mathbb{N}^*$ on a $(\det \mathrm{M})^m = 1$, donc $\det \mathrm{M} = \pm 1$, et ainsi M est un élément du groupe $\mathrm{GL}_n(\mathbb{Z})$, qui plus est, un élément d'ordre fini. On cherche donc à montrer qu'il existe un entier $r \in \mathbb{N}^*$ tel que toute matrice d'ordre fini de $\mathrm{GL}_n(\mathbb{Z})$ ait un ordre qui divise r. L'exercice précédent a par exemple montré que pour $n = 2$ on peut prendre $r = 12$.

On va montrer l'existence d'un tel entier r dans le cas général, mais sans chercher à le préciser. Soit $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$ d'ordre fini et $m \in \mathbb{N}^*$ tel que $\mathrm{M}^m = \mathrm{I}_n$. Le polynôme $\mathrm{X}^m - 1$ annule M donc les valeurs propres complexes de M sont des racines de l'unité. Par ailleurs le polynôme caractéristique de M est à coefficients dans $\mathbb{Z}$. Or l'ensemble E des polynômes unitaires de degré n, à coefficients dans $\mathbb{Z}$, et dont les racines sont des racines de l'unité est fini (car les relations coefficients-racines montrent que les coefficients d'un élément $\mathrm{P} \in \mathrm{E}$ ne peuvent prendre qu'un nombre fini de valeurs). Il nous suffit de prendre pour r le ppcm des ordres (dans $\mathbb{C}^*$) de toutes les racines de tous les polynômes de E. Si $\mathrm{M} \in \mathrm{GL}_n(\mathbb{Z})$ est d'ordre fini quelconque on a $\lambda^r = 1$ pour toute valeur propre λ de M et, comme M est diagonalisable dans $\mathcal{M}_n(\mathbb{C})$, on a $\mathrm{M}^r = \mathrm{I}_n$. D'où le résultat. ◁

Avec quelques connaissances sur les polynômes cyclotomiques (leur irréductibilité sur $\mathbb{Q}$ et la factorisation sur $\mathbb{Q}$ des polynômes $\mathrm{X}^n - 1$) il est possible de caractériser plus précisément les ordres possibles des éléments de $\mathrm{GL}_n(\mathbb{Z})$.

L'exercice suivant étudie les éléments d'ordre 4 ou 3 de $\mathrm{SL}_2(\mathbb{Z})$. La première question reprend des choses déjà vues mais les questions suivantes sont difficiles.

3.19. Éléments d'ordre 4 ou 3 dans $\mathrm{SL}_2(\mathbb{Z})$

Soit $\mathrm{M} \in \mathrm{SL}_2(\mathbb{Z})$.

1. On suppose qu'il existe $n \in \mathbb{N}^*$ tel que $\mathrm{M}^n = \mathrm{I}_2$. Montrer que l'ordre de M dans $\mathrm{SL}_2(\mathbb{Z})$ est au plus 6.

2. On suppose M d'ordre 4. Montrer que M est conjuguée dans $\mathrm{GL}_2(\mathbb{Z})$ à la matrice $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$.

3. On suppose M d'ordre 3. Montrer que M est conjuguée dans $\mathrm{GL}_2(\mathbb{Z})$ à la matrice $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$.

(École normale supérieure)

▷ **Solution.**

1. Le polynôme $X^n - 1$ est un polynôme annulateur à racines simples dans $\mathbb{C}$ donc M est diagonalisable dans $\mathcal{M}_2(\mathbb{C})$ et ses valeurs propres complexes sont des racines n-ièmes de l'unité, λ et λ'. La matrice M est semblable à $D = \begin{pmatrix} \lambda & 0 \\ 0 & \lambda' \end{pmatrix}$ et l'ordre de M est le plus petit entier $k \geqslant 1$ tel que $\lambda^k = \lambda'^k = 1$.

Si les valeurs propres sont réelles alors $\mathrm{Sp}(M) \subset \{-1, 1\}$ et $M^2 = I_2$: M est d'ordre au plus 2.

Sinon, comme M est réelle, les deux valeurs propres sont conjuguées distinctes $\lambda = e^{\frac{2ik\pi}{n}}$ $(1 < k \leqslant n-1)$ et $\lambda' = \overline{\lambda}$. Mais $\mathrm{Tr}\, M = 2\cos\dfrac{2k\pi}{n}$ est un entier compris entre -2 et 2, donc $\cos\dfrac{2k\pi}{n} \in \left\{0, \pm\dfrac{1}{2}\right\}$ (on exclut le cas où ce cosinus vaut ± 1 qui équivaut à $\lambda = \pm 1$, déjà traité). On obtient $\lambda \in \left\{\pm i, \pm\dfrac{1}{2} \pm i\dfrac{\sqrt{3}}{2}\right\}$. Si les valeurs propres sont $\pm i$, l'ordre de M est 4. Si les valeurs propres de M sont $e^{\frac{i\pi}{3}}$ et $e^{-\frac{i\pi}{3}}$, l'ordre de M est 6. Si les valeurs propres de M sont $e^{\frac{2i\pi}{3}}$ et $e^{-\frac{2i\pi}{3}}$, l'ordre de M est 3.

Dans tous les cas, l'ordre de M est inférieur ou égal à 6.

2. Si l'ordre de M est 4, ses valeurs propres sont $\pm i$ d'après la question précédente. Donc $\mathrm{Tr}\, M = 0$ et $\chi_M = X^2 + 1$. D'après le théorème de Cayley-Hamilton, $M^2 = -I_2$. Soit f l'endomorphisme de $\mathbb{R}^2$ canoniquement associé à M. Si u est un vecteur non nul quelconque de $\mathbb{Z}^2$, la famille $(u, f(u))$ est libre car M n'a pas de valeur propre réelle et la matrice de f dans la base $(u, f(u))$ est clairement $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Si on note P la matrice de passage de la base canonique à la base $(u, f(u))$, on obtient $M = P\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}P^{-1}$. La matrice P est à coefficients dans $\mathbb{Z}$ mais on veut que P appartienne à $\mathrm{GL}_2(\mathbb{Z})$, c'est-à-dire que $\det P = \pm 1$. Cela équivaut à $\det(u, f(u)) = \pm 1$. Posons $M = \begin{pmatrix} a & b \\ c & -a \end{pmatrix}$ avec $-a^2 - bc = 1$, car $\mathrm{Tr}(M) = 0$ et $\det M = 1$. Si $u = (x, y) \in \mathbb{Z}^2$, on a

$$\det(u, f(u)) = \begin{vmatrix} x & ax + by \\ y & cx - ay \end{vmatrix} = cx^2 - 2axy - by^2.$$

Il s'agit donc de démontrer que la forme quadratique définie par $q(x, y) = cx^2 - 2axy - by^2$ prend la valeur 1 ou -1 sur $\mathbb{Z}^2$. La matrice de q dans la base canonique de $\mathbb{R}^2$ est $\begin{pmatrix} c & -a \\ -a & -b \end{pmatrix}$. Comme le déterminant $-bc - a^2 = 1$ est strictement positif, la forme quadratique est définie positive ou définie négative (en fonction du signe des deux valeurs propres).

Démontrons alors le lemme suivant (voir aussi l'exercice 3.43 du tome algèbre 3) :

Lemme. *Soit $q(x,y) = ax^2 + 2bxy + cy^2$ une forme quadratique sur $\mathbb{R}^2$. On suppose que q est définie positive ou définie négative et on considère $e = \inf_{(x,y)\in\mathbb{Z}^2\setminus\{(0,0)\}} |q(x,y)|$. On a alors $3e^2 \leqslant 4(ac-b^2)$.*

Démonstration.

On a, pour tout $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $|q(x,y)| > 0$. On en déduit que $\inf_{\|(x,y)\|=1} |q(x,y)| = m > 0$, car cette borne inférieure est atteinte, puisque le cercle unité de $\mathbb{R}^2$ est compact. Par homogénéité, on obtient, pour tout $(x,y) \in \mathbb{R}^2$, $|q(x,y)| \geqslant m(x^2+y^2)$. Soit N un entier naturel tel que $\mathrm{N} \geqslant \sqrt{\dfrac{e+1}{m}}$. D'après ce qui précède, on a, si $(x,y) \in \mathbb{Z}^2$ et $|x| > \mathrm{N}$ ou $|y| > \mathrm{N}$, $|q(x,y)| \geqslant e+1$. On a donc $e = \inf_{\substack{(x,y)\in\mathbb{Z}^2\setminus\{(0,0)\} \\ |x|\leqslant \mathrm{N}, |y|\leqslant \mathrm{N}}} |q(x,y)|$. Cette borne inférieure est atteinte car l'ensemble est fini. On en déduit que $e > 0$.

Soit $u = (\alpha,\beta) \in \mathbb{Z}^2$ tel que $|q(u)| = e$. Notons $d = \mathrm{pgcd}(\alpha,\beta)$ et α', β' les entiers tels que $\alpha = d\alpha'$ et $\beta = d\beta'$. On a

$$e = |q(u)| = d^2|q(\alpha',\beta')| \geqslant d^2 e$$

et donc $d^2 \leqslant 1$, c'est-à-dire $d = 1$.

Ainsi α et β sont premiers entre eux et d'après le théorème de Bezout, il existe $(\gamma,\delta) \in \mathbb{Z}^2$ tel que $\alpha\gamma + \beta\delta = 1$. Posons $v = (-\gamma,\delta)$ et montrons que $\mathbb{Z}^2 = \mathbb{Z}u + \mathbb{Z}v$. On a évidemment $\mathbb{Z}u + \mathbb{Z}v \subset \mathbb{Z}^2$. Soit $\mathrm{P} = \begin{pmatrix} \alpha & -\delta \\ \beta & \gamma \end{pmatrix}$. Cette matrice est dans $\mathcal{M}_2(\mathbb{Z})$ et on a $\det \mathrm{P} = 1$, donc P est inversible et $\mathrm{P}^{-1} \in \mathcal{M}_2(\mathbb{Z})$. La matrice P^{-1} est la matrice de passage de (u,v) à la base canonique (e_1,e_2) de $\mathbb{R}^2$. Comme elle est à coefficients entiers, e_1 et e_2 appartiennent à $\mathbb{Z}u + \mathbb{Z}v$ et finalement $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}^2$.

Soit $\varepsilon = \pm 1$ tel que $q(u) = \varepsilon e$. Il existe $(b',c') \in \mathbb{R}^2$ tel que l'expression de q dans la base (u,v) est $\varepsilon e x^2 + 2b'xy + c'y^2$. Par les formules de changement de base on a alors $\begin{pmatrix} \varepsilon e & b' \\ b' & c' \end{pmatrix} = {}^t\mathrm{P}\begin{pmatrix} a & b \\ b & c \end{pmatrix}\mathrm{P}$ et donc $\varepsilon e c' - b'^2 = (\det \mathrm{P})^2(ac-b^2) = ac - b^2 > 0$ car la forme est définie.

Pour tout $(x,y) \in \mathbb{Z}^2$, on obtient

$$q(xu+yv) = \varepsilon e x^2 + 2b'xy + c'y^2 = \varepsilon e\left(x + \frac{\varepsilon b'}{e}y\right)^2 + \left(c' - \frac{b'^2}{\varepsilon e}\right)y^2$$

$$= \varepsilon e\left(x + \frac{\varepsilon b'}{e}y\right)^2 + \frac{ac-b^2}{\varepsilon e}y^2.$$

En prenant $y = 1$ et en choisissant $x \in \mathbb{Z}$ tel que $\left|x + \frac{\varepsilon b'}{c}\right| \leqslant \frac{1}{2}$, on a

$$e \leqslant |q(xu + v)| \leqslant \frac{1}{4}c + \frac{ac - b^2}{e}$$

et donc $3e^2 \leqslant 4(ac - b^2)$. $\diamondsuit$

Avec les notations de l'exercice on obtient $3c^2 \leqslant 4(-bc - a^2) \leqslant 4$. Ici, e est un entier non nul, puisque le minimum est atteint et $(a, b, c) \in \mathbb{Z}^3$, donc $e^2 = 1$ et $e = 1$. Il existe $(x, y) \in \mathbb{Z}^2$ tel que $e = 1 = |q(x, y)|$.

3. On a $M^3 = I_2$ et $M \neq I_2$. Les seules valeurs propres possibles pour M sont 1, j et j^2 et M est diagonalisable dans $\mathcal{M}_2(\mathbb{C})$ car $X^3 - 1$ est un polynôme annulateur à racines simples. En fait, $1 \notin \operatorname{Sp} M$, car sinon les deux valeurs propres de M seraient réelles donc égales à 1 et comme M est diagonalisable, elle serait égale à I_2. La matrice $M - I_2$ est donc inversible ; de $M^3 - I_2 = (M - I_2)(M^2 + M^2 + I_2)$, on tire donc $M^2 + M + I_2 = 0$. Le polynôme $X^2 + X + 1$ est un polynôme annulateur. C'est le polynôme minimal car M n'est pas scalaire et aussi le polynôme caractéristique car il est de degré 2.

Si u est un vecteur de $\mathbb{Z}^2$, la famille $(u, f(u))$ est libre car M n'a pas de valeur propre réelle et la matrice de f dans la base $(u, f(u))$ est $\begin{pmatrix} 0 & -1 \\ 1 & -1 \end{pmatrix}$, car $f^2(u) = -f(u) - u$. On raisonne comme dans la question précédente. On note P la matrice de passage de la base canonique à la base $(u, f(u))$, et on cherche u tel que $\det P = \pm 1$, c'est-à-dire $\det(u, f(u)) = \pm 1$. On a, puisque $\chi_M = X^2 + X + 1$, $\operatorname{Tr}(M) = -1$ et $\det M = 1$. On pose $M = \begin{pmatrix} a & b \\ c & -1 - a \end{pmatrix}$ avec $-a^2 - a - bc = 1$. Si $u = (x, y) \in \mathbb{Z}^2$, alors

$$q(u) = \det(u, f(u)) = \begin{vmatrix} x & ax + by \\ y & cx - (a + 1)y \end{vmatrix} = cx^2 - (2a + 1)xy - by^2$$

et q a pour matrice $\begin{pmatrix} c & -a - \frac{1}{2} \\ -a - \frac{1}{2} & -b \end{pmatrix}$ dans la base canonique. Le déterminant $-bc - \left(a + \frac{1}{2}\right)^2 = \frac{3}{4}$ est strictement positif donc la forme quadratique est définie positive ou négative. Le minimum e de $|q|$ vérifie d'après le lemme $3e^2 \leqslant 4(-bc - (a + \frac{1}{2})^2) \leqslant 4(-bc - a^2 - a) - 1 \leqslant 3$ et donc $e = 1$. On conclut comme dans la question précédente. $\triangleleft$

L'exercice suivant va généraliser l'exercice 3.18. En effet, il démontre que l'ensemble des cardinaux des sous-groupes finis de $\mathrm{GL}_n(\mathbb{Z})$ est majoré (on parle de restriction cristallographique). En particulier, il existe un

entier N_n *tel que l'ordre de toute matrice d'ordre fini de* $\mathrm{GL}_n(\mathbb{Z})$ *soit inférieur à* N_n. *On a vu en 3.17 que pour* $n = 2$ *l'entier* $N_2 = 6$ *convient.*

3.20. Sous-groupes finis de $\mathrm{GL}_n(\mathbb{Z})$

1. Soit $A \in \mathcal{M}_n(\mathbb{Z})$. On suppose que A est annulée par un polynôme complexe non nul P dont les racines sont simples et de module < 1. Montrer que $A = 0$.

2. Soit G un sous-groupe fini de $\mathrm{GL}_n(\mathbb{Z})$. Montrer que si p est un nombre premier impair, la réduction modulo p de $\mathrm{GL}_n(\mathbb{Z})$ dans $\mathrm{GL}_n(\mathbb{Z}/p\mathbb{Z})$ restreinte à G est injective. Que peut-on en déduire ?

(**École normale supérieure**)

▷ **Solution.**

1. Comme le polynôme P est à racines simples, la matrice A est diagonalisable dans $\mathcal{M}_n(\mathbb{C})$. Soit D diagonale et Q inversible telles que $A = Q^{-1}DQ$. Comme toutes les valeurs propres de A sont de module < 1, la suite $(A^p)_{p\geqslant 0}$ tend vers 0 lorsque p tend vers l'infini. En particulier, il existe un entier p tel que $\|A^p\| < 1$ où $\|(m_{ij})\| = \max_{1\leqslant i,j\leqslant n} |m_{ij}|$. Mais la matrice A^p étant à coefficients dans $\mathbb{Z}$, cela impose que $A^p = 0$. On a alors $D^p = 0$ et donc $D = 0$ car D est diagonale. Ainsi, A est nulle.

2. L'application $f : G \to \mathrm{GL}_n(\mathbb{Z}/p\mathbb{Z})$ qui à une matrice $M = (m_{ij})$ de G associe $\overline{M} = (\overline{m_{ij}})$, sa réduction modulo p, est un morphisme de groupes. Pour montrer qu'il est injectif, on montre que son noyau est réduit à l'identité. Soit $M \in \operatorname{Ker} f$. On a donc $M = I_n$ modulo p de sorte qu'on peut écrire $M = I_n + pA$ avec $A \in \mathcal{M}_n(\mathbb{Z})$. Comme G est fini, M est d'ordre fini $q \geqslant 1$. L'égalité $M^q = I_n$ signifie que le polynôme $P(X) = (1 + pX)^q - 1$ annule la matrice A. Or les racines de P sont les nombres complexes $z_k = \dfrac{w_k - 1}{p}$ où $w_1, \ldots, w_q$ sont les racines q-ièmes de l'unité. Ces racines sont deux à deux distinctes et comme $p \geqslant 3$ on a $|z_k| < 1$ pour tout k. La question précédente permet d'affirmer que A est nulle et donc que M est la matrice identité.

Comme $\mathrm{GL}_n(\mathbb{Z}/p\mathbb{Z})$ est fini, on en déduit que les sous-groupes finis de $\mathrm{GL}_n(\mathbb{Z})$ ont un cardinal majoré par le cardinal de $\mathrm{GL}_n(\mathbb{Z}/p\mathbb{Z})$. En particulier, il n'y a, à isomorphisme près, qu'un nombre fini de sous-groupes finis dans $\mathrm{GL}_n(\mathbb{Z})$ (car il n'y a qu'un nombre fini de groupes de cardinal fini donné, à isomorphisme près bien entendu). ◁

Voici maintenant quelques exercices où le corps de base est un corps fini.

3.21. Un calcul de signature

Soit $p \geqslant 3$ un nombre premier et M une matrice de $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$. L'application $X \mapsto MX$ est une permutation de $(\mathbb{Z}/p\mathbb{Z})^2$. Quelle est sa signature ? *On pourra utiliser sans le redémontrer le fait que* $(\mathbb{Z}/p\mathbb{Z})^*$ *est un groupe cyclique.*

(École normale supérieure)

▷ **Solution.**

• La signature induit bien entendu un morphisme de groupes de $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ dans $\{\pm 1\}$, que l'on notera ε. Pour connaître un morphisme de groupes, il suffit de le connaître sur une partie génératrice. Or, on sait que le groupe linéaire est engendré par les matrices de transvection $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$ avec $\lambda \in \mathbb{Z}/p\mathbb{Z}$ et les matrices de dilatation $\begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}, \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}$ avec $\alpha \in (\mathbb{Z}/p\mathbb{Z})^*$. Le lecteur se reportera à l'exercice 3.1 en notant que le résultat est rapide à prouver dans le cas des matrices 2×2 (ce qu'il faudrait peut être faire le jour de l'oral).

• Montrons que toutes les matrices de transvection ont une signature égale à 1. En effet, c'est clair pour la matrice identité, et si λ n'est pas nul, on observe que $\mathrm{T}_\lambda = \begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ est d'ordre p dans le groupe $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$. Par conséquent, on a $\varepsilon(\mathrm{T}_\lambda)^p = \varepsilon(\mathrm{T}_\lambda^p) = \varepsilon(Id) = 1$. Comme p est impair, on a forcément $\varepsilon(\mathrm{T}_\lambda) = 1$. Le raisonnement est identique pour la transposée de T_λ.

On peut aussi utiliser le fait que les transvections sont des carrés du groupe $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$. *Comme les matrices de transvection engendrent le groupe spécial linéaire* $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$, *la signature restreinte à ce sous-groupe est donc triviale. Cette remarque n'est toutefois pas indispensable pour poursuivre.*

• Les matrices de dilatation $\mathrm{D}_\alpha = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}$ avec $\alpha \in (\mathbb{Z}/p\mathbb{Z})^*$ forment un sous-groupe isomorphe à $(\mathbb{Z}/p\mathbb{Z})^*$. Comme $(\mathbb{Z}/p\mathbb{Z})^*$ est cyclique (l'énoncé le rappelle et et le lecteur en trouvera une preuve dans l'exercice 4.13 du tome 1 d'algèbre) il suffit ici encore de s'intéresser à la signature d'un générateur.

Soit donc α qui engendre $(\mathbb{Z}/p\mathbb{Z})^*$. On va compter les orbites de $(\mathbb{Z}/p\mathbb{Z})^2$ sous l'action de D_α. Il y a tout d'abord les vecteurs $(0, x)$ qui donnent des orbites réduites à des singletons (et il y en a p) puis les orbites des vecteurs $(1, x)$ qui contiennent les $p-1$ vecteurs (y, x) avec $y \neq 0$ (et on en a encore p). Au total il y a donc $2p$ orbites et la signature est donc $(-1)^{p^2-2p} = -1$ car p est impair.

• On aimerait tout de même avoir une vision globale plus claire de

cette signature. On connaît un morphisme de groupes important sur $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ à savoir le déterminant. Il est assez naturel d'essayer de relier notre signature au déterminant.

Si x est dans $\mathbb{Z}/p\mathbb{Z}$ on sait que $x^{p-1} = 1$ (c'est le petit théorème de Fermat) et donc $x^{\frac{p-1}{2}}$ vaut 1 ou -1 puisque son carré fait 1 (et le polynôme $\mathrm{X}^2 - 1$ n'a que 1 et -1 pour racines dans $\mathbb{Z}/p\mathbb{Z}$, qui est un corps). Pour notre générateur α de tout à l'heure on a forcément $\alpha^{\frac{p-1}{2}} = -1$, car sinon on aurait $x^{\frac{p-1}{2}} = 1$ pour tout x non nul et le polynôme $\mathrm{X}^{\frac{p-1}{2}} - 1$ aurait $p-1$ racines, ce qu'interdit son degré.

Ainsi, on a $\varepsilon(\mathrm{D}_\alpha) = -1 = \alpha^{\frac{p-1}{2}} = \det(\mathrm{D}_\alpha)^{\frac{p-1}{2}}$. Cette relation est alors vraie pour toute les matrices de dilatation et aussi pour les matrices de transvection (le déterminant vaut 1). Deux morphismes de groupes qui coïncident sur une partie génératrice étant égaux, on a donc démontré la jolie formule

$$\boxed{\forall \mathrm{M} \in \mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z}), \quad \varepsilon(\mathrm{M}) = \det \mathrm{M}^{\frac{p-1}{2}}}. \lhd$$

Les groupes linéaires sur les corps finis fournissent des exemples importants de groupes finis. L'exercice suivant propose l'étude du groupe $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$.

3.22. Étude de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$

1. Quel est le cardinal de $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$? de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$?

2. Montrer qu'il n'existe aucun morphisme surjectif de groupes de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ dans $\mathbb{Z}/2\mathbb{Z}$.

3. Montrer qu'il n'existe aucun sous-groupe de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ de cardinal 12.

4. $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ est-il isomorphe à $\mathcal{S}_4$?

5. Montrer qu'il existe un morphisme de groupes surjectif de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ sur $\mathcal{A}_4$.

(**École normale supérieure**)

▷ **Solution.**

1. La donnée d'une matrice de $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ équivaut à la donnée d'une famille libre de deux vecteurs du plan $(\mathbb{Z}/3\mathbb{Z})^2$. Il y a $3^2 - 1 = 8$ choix pour le premier vecteur, qui doit simplement être non nul, et pour chacun de ces choix, il y a $9 - 3 = 6$ manières de le compléter en une famille libre puisqu'il faut choisir le second vecteur en dehors de la droite vectorielle engendrée par le premier. Ainsi, $|\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})| = 8 \times 6 = 48$.

Comme le déterminant est un morphisme surjectif de groupes de $\mathrm{GL}_2(\mathbb{Z}/3\mathbb{Z})$ dans $(\mathbb{Z}/3\mathbb{Z})^*$, dont le noyau est $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$, on a

$$|\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})| = 24.$$

2. Soit ψ un morphisme de groupes de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ dans $\mathbb{Z}/2\mathbb{Z}$. On se propose de prouver que ψ est trivial. Nous savons qu'un morphisme de groupes est uniquement déterminé par l'image d'une partie génératrice. Or, nous savons que les matrices de transvection $\begin{pmatrix} 1 & \lambda \\ 0 & 1 \end{pmatrix}$ et $\begin{pmatrix} 1 & 0 \\ \lambda & 1 \end{pmatrix}$, où $\lambda \in \{1, 2\}$ engendrent $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ (voir l'exercice 3.1).

On a pour tout $A \in \mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$, $\psi(A^2) = 2\psi(A) = 0$, c'est-à-dire que ψ est nul sur les carrés. Or, $\begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^2$ et $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & 1 \end{pmatrix}^2$. Donc ψ est nul sur toutes les matrices de transvection et par suite identiquement nul.

3. Supposons par l'absurde qu'il existe un sous-groupe H de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ de cardinal 12 et soit $a \notin \mathrm{H}$. On a alors $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z}) = \mathrm{H} \cup a\mathrm{H}$, la réunion étant disjointe. Le produit de deux éléments de H est dans H, le produit d'un élément de H et d'un élément de aH est dans aH et le produit de deux éléments de aH est dans H (car l'égalité $ahah' = ah''$ avec h, h', h'' dans H implique que $a \in \mathrm{H}$). Il en découle directement que l'application qui envoie les éléments de H sur 0 et ceux de aH sur 1 est un morphisme de groupes surjectif de $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ sur $\mathbb{Z}/2\mathbb{Z}$, ce qui contredit le résultat de la question précédente.

4. Comme $\mathcal{S}_4$ (qui est aussi de cardinal 24) admet un sous-groupe d'ordre 12, à savoir le groupe alterné $\mathcal{A}_4$, la question précédente permet d'affirmer que $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ n'est pas isomorphe à $\mathcal{S}_4$.

5. On va utiliser une opération du groupe $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ sur un ensemble à 4 éléments. On prend l'ensemble E des droites vectorielles du plan $(\mathbb{Z}/3\mathbb{Z})^2$. Il y a 4 droites qui sont $\mathrm{D}_1 = \mathrm{Vect}(1, 0)$, $\mathrm{D}_2 = \mathrm{Vect}(0, 1)$, $\mathrm{D}_3 = \mathrm{Vect}(1, 1)$ et $\mathrm{D}_4 = \mathrm{Vect}(1, 2)$. Une matrice $A \in \mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ envoie une droite sur une droite et cela de manière injective (deux vecteurs libres sont envoyés sur deux vecteurs libres). Il en résulte que $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ opère sur E. Notons $\varphi : \mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z}) \to \mathcal{S}_\mathrm{E}$ le morphisme de groupes correspondant à cette opération. Soit $A \in \operatorname{Ker}\varphi$. On a $A(\mathrm{D}_1) = \mathrm{D}_1$ et $A(\mathrm{D}_2) = \mathrm{D}_2$, donc A est une homothétie, *i.e.* $A = \pm \mathrm{Id}$ (car $A \in \mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$). Inversement, Id et $-\mathrm{Id}$ opèrent de manière triviale sur E. Donc $|\operatorname{Ker}\varphi| = 2$ et par suite $|\operatorname{Im}\varphi| = 12$. On observe que les transvections vues ci-dessus qui engendrent $\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})$ donnent des 3-cycles de E. Par exemple $\varphi\left(\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\right) = (\mathrm{D}_2, \mathrm{D}_3, \mathrm{D}_4)$. Il en résulte que $\varphi(\mathrm{SL}_2(\mathbb{Z}/3\mathbb{Z})) = \mathcal{A}_\mathrm{E} \simeq \mathcal{A}_4$. D'où le résultat. ◁

Les trois exercices suivants étudient encore des groupes linéaires finis mais sur des anneaux quotients $\mathbb{Z}/m\mathbb{Z}$. Les techniques sont les mêmes

avec des complications supplémentaires liées à la non-intégrité de l'anneau.

3.23. Étude de $\mathbf{SL_2(\mathbb{Z}/2^n\mathbb{Z})}$

Soit A un anneau commutatif et $\mathrm{G} = \mathrm{SL}_2(\mathrm{A})$ l'ensemble des matrices $(2,2)$ à coefficients dans A et de déterminant 1.

1. Montrer que G est un groupe pour le produit matriciel.

2. On prend $\mathrm{A} = \mathbb{Z}/p\mathbb{Z}$ avec p premier. Calculer $\mathrm{Card}\,\mathrm{G}$. Dans le cas où $p = 2$, étudier l'ordre des éléments de G ; reconnaître un groupe connu isomorphe à G.

3. On prend $\mathrm{A} = \mathbb{Z}/2^n\mathbb{Z}$. Quel est le cardinal de G ? Montrer que tout élément de G a un ordre inférieur ou égal à $3 \times 2^{n-1}$.

(École normale supérieure)

▷ **Solution.**

1. L'ensemble $\mathcal{M}_2(\mathrm{A})$ des matrices carrées de taille $(2,2)$ à coefficients dans A est un anneau pour les opérations usuelles. Le déterminant d'une matrice $\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathrm{A})$ étant défini par $\det \mathrm{M} = ad - bc$, il est facile de vérifier que $\det(\mathrm{MN}) = \det \mathrm{M} \det \mathrm{N}$ pour tout couple de matrices (M, N). Par suite, si M est inversible dans $\mathcal{M}_2(\mathrm{A})$, $\det \mathrm{M}$ est un inversible de l'anneau A. Réciproquement, soit $\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathrm{A})$ telle que $\det \mathrm{M}$ soit inversible dans A. Il est facile de vérifier que la matrice $(\det \mathrm{M})^{-1} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$ est l'inverse de M. On a donc montré que le groupe des inversibles de l'anneau $\mathcal{M}_2(\mathrm{A})$, notons-le $\mathrm{GL}_2(\mathrm{A})$, est exactement l'ensemble des matrices dont le déterminant appartient à $\mathrm{U}(\mathrm{A})$, le groupe des inversibles de A. De plus, le déterminant établit un morphisme de groupes entre $\mathrm{GL}_2(\mathrm{A})$ et $\mathrm{U}(\mathrm{A})$ dont $\mathrm{G} = \mathrm{SL}_2(\mathrm{A})$ n'est autre que le noyau. Par conséquent G est bien un groupe pour le produit matriciel.

2. Notons que le morphisme $\det : \mathrm{GL}_2(\mathrm{A}) \to \mathrm{U}(\mathrm{A})$ est surjectif car si a est un inversible de A la matrice diagonale $\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}$ est inversible et a pour déterminant a. Il en découle que $\mathrm{Card}\,\mathrm{G} \times \mathrm{Card}\,\mathrm{U}(\mathrm{A}) = \mathrm{Card}\,\mathrm{GL}_2(\mathrm{A})$: c'est un corollaire direct du lemme des bergers et le lecteur trouvera une preuve détaillée de ce fait dans la solution de l'exercice 2.4 du premier tome d'algèbre.

Lorsque $\mathrm{A} = \mathbb{Z}/p\mathbb{Z}$ avec p premier il s'agit d'un corps, et $\mathrm{U}(\mathrm{A})$ est de cardinal $p-1$. Le calcul du cardinal de $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ est facile : pour définir une matrice inversible, on choisit d'abord sa première colonne qui doit être non nulle (il y a donc $p^2 - 1$ possibilités) puis la seconde qui ne

doit pas être colinéaire à la première (et il y a donc p^2-p possibilités). Le cardinal de $\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})$ est donc $(p^2-1)(p^2-p)$ et celui de $\mathrm{G}=\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$ est $p(p^2-1)$. Le lecteur trouvera dans l'exercice 1.8 du tome 1 d'algèbre le calcul du cardinal de $\mathrm{GL}_n(\mathrm{K})$ pour n quelconque et K corps fini quelconque.

En particulier, avec $p=2$, on a $\mathrm{Card}\,\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})=6$. Les 6 éléments de ce groupe sont l'identité, les matrices $\begin{pmatrix}1&1\\0&1\end{pmatrix}$, $\begin{pmatrix}1&0\\1&1\end{pmatrix}$ et $\begin{pmatrix}0&1\\1&0\end{pmatrix}$ qui sont d'ordre 2 et les matrices $\begin{pmatrix}0&1\\1&1\end{pmatrix}$ et $\begin{pmatrix}1&1\\1&0\end{pmatrix}$ qui sont d'ordre 3. On a la même répartition des ordres que dans le groupe symétrique $\mathcal{S}_3$ qui, outre l'identité, contient 3 transpositions et deux 3-cycles. Et effectivement ces deux groupes sont isomorphes. Pour le voir, il suffit d'observer qu'une matrice M de $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ induit une permutation des 3 vecteurs non nuls du plan $(\mathbb{Z}/2\mathbb{Z})^2$ et que cette permutation n'est l'identité que si $\mathrm{M}=\mathrm{I}_2$.

À isomorphisme près il n'y a que deux groupes de cardinal 6 : $\mathbb{Z}/6\mathbb{Z}$ qui est cyclique et $\mathcal{S}_3$. Comme $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z})$ n'est pas commutatif, ce fait permet de dire directement qu'il est isomorphe à $\mathcal{S}_3$.

3. L'étude détaillée du cas $n=1$ menée dans la question précédente incite à procéder par récurrence sur l'entier n. Et effectivement, comme l'anneau $\mathbb{Z}/2^n\mathbb{Z}$ n'est pas intègre pour $n\geqslant 2$, il n'est pas évident de trouver directement le cardinal de $\mathrm{GL}_2(\mathbb{Z}/2^n\mathbb{Z})$. On dispose d'un morphisme d'anneau naturel π_n entre $\mathbb{Z}/2^{n+1}\mathbb{Z}$ et $\mathbb{Z}/2^n\mathbb{Z}$ qui à la classe d'un entier k modulo 2^{n+1} associe sa classe modulo 2^n (qui bien entendu ne dépend pas du représentant k choisi dans sa classe modulo 2^{n+1}). Ce morphisme est surjectif et son noyau est de cardinal 2 (il contient les classes de 0 et de 2^n). On a alors un morphisme de groupes entre $\mathrm{SL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z})$ et $\mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$ en considérant l'application φ_n qui à la matrice $\mathrm{M}=\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z})$ associe la matrice $\begin{pmatrix}\pi_n(a)&\pi_n(b)\\\pi_n(c)&\pi_n(d)\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$. Le noyau de φ_n est de cardinal 8. Vérifions que φ_n est surjectif. Soit M une matrice de $\mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$. Comme π_n est surjectif on peut trouver quatre entiers a,b,c,d tels que $\varphi_n\begin{pmatrix}\bar a&\bar b\\\bar c&\bar d\end{pmatrix}=\mathrm{M}$ où $\bar k$ désigne la classe de l'entier k dans $\mathbb{Z}/2^{n+1}\mathbb{Z}$. Mais cette matrice n'est pas forcément de déterminant 1 ! Par hypothèse on a $ad-bc\equiv 1\ [2^n]$ donc, modulo 2^{n+1}, on a soit $ad-bc\equiv 1$, soit $ad-bc\equiv 1+2^n$. Dans le premier cas c'est bon. Dans le second cas, l'un au moins des quatre entiers a,b,c,d est impair, disons par exemple a pour fixer les idées. On remplace alors d par $d+2^n$. On obtient alors un antécédent de M par φ_n qui est bien dans $\mathrm{SL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z})$.

Ce travail montre donc que $\operatorname{Card} \mathrm{SL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z}) = 8 \times \operatorname{Card} \mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$ et, compte tenu de la question précédente, on peut conclure que

$$\boxed{\operatorname{Card} \mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z}) = 3 \times 2^{3n-2}}.$$

Comme $\mathrm{U}(\mathbb{Z}/2^n\mathbb{Z})$ est de cardinal 2^{n-1} on en déduit que $\mathrm{GL}_2(\mathbb{Z}/2^n\mathbb{Z})$ est de cardinal $3 \times 2^{4n-3}$. Notons qu'on aurait aussi pu étudier le morphisme induit par π_n entre $\mathrm{GL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z})$ et $\mathrm{GL}_2(\mathbb{Z}/2^n\mathbb{Z})$. Le noyau est le même et la surjectivité est immédiate. On en aurait déduit le cardinal de $\mathrm{GL}_2(\mathbb{Z}/2^n\mathbb{Z})$ puis celui de $\mathrm{SL}_2(\mathbb{Z}/2^n\mathbb{Z})$.

Pour la question sur les ordres des éléments de G on procède encore par récurrence sur n, le cas $n = 1$ étant traité dans la question 2. Supposons le résultat vrai au rang n et soit $\mathrm{M} \in \mathrm{SL}_2(\mathbb{Z}/2^{n+1}\mathbb{Z})$. Notons p l'ordre de $\varphi_n(\mathrm{M})$. Par hypothèse de récurrence, $p \leqslant 3 \times 2^{n-1}$. On a $\varphi_n(\mathrm{M}^p) = \varphi_n(\mathrm{M})^p = \mathrm{I}_2$ donc $\mathrm{M}^p \in \operatorname{Ker} \varphi_n$. On constate que les 7 matrices du noyau de φ_n distinctes de l'identité sont toutes d'ordre 2. On a donc $\mathrm{M}^{2p} = \mathrm{I}_2$ et l'ordre de M est inférieur à $2p$ donc à 3×2^n. Cela termine la récurrence.

Le lecteur montrera sans mal que ce résultat est aussi vrai pour les éléments de $\mathrm{GL}_2(\mathbb{Z}/2^n\mathbb{Z})$. ◁

On trouvera dans la solution de l'exercice suivant le calcul du cardinal de $\mathrm{SL}_2(\mathbb{Z}/d\mathbb{Z})$ pour tout entier d. L'exercice est difficile, surtout la dernière question.

3.24. Réduction modulo n

Si A est un anneau commutatif, $\mathrm{SL}_2(\mathrm{A})$ désigne le groupe des matrices à coefficients dans A et de déterminant 1. Pour $n \in \mathbb{N}^*$ on considère $\psi_n : \mathrm{SL}_2(\mathbb{Z}) \to \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ l'application qui à toute matrice A associe la matrice $\bar{\mathrm{A}}$ dont les coefficients sont les classes modulo n des coefficients de A.

1. Montrer que ψ_n est un morphisme surjectif de groupes.

2. Cela est-il encore vrai avec $\varphi_n : \mathrm{GL}_2(\mathbb{Z}) \to \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$?

3. Si $d = \operatorname{pgcd}(n, m)$ montrer que $\operatorname{Ker} \psi_d = \operatorname{Ker} \psi_n \operatorname{Ker} \psi_m$ (ensemble des produits d'un élément de $\operatorname{Ker} \psi_n$ par un élément de $\operatorname{Ker} \psi_m$).

(École normale supérieure)

▷ **Solution.**

1. Notons que le fait que $\mathrm{SL}_2(\mathrm{A})$ est un groupe est prouvé dans l'exercice précédent. Si $\mathrm{A} \in \mathrm{SL}_2(\mathbb{Z})$, alors $\det(\psi_n(\mathrm{A})) = \overline{\det \mathrm{A}} = \bar{1}$, donc $\psi_n(\mathrm{A})$ appartient à $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$. L'application ψ_n est clairement un morphisme de groupes. Il reste à montrer qu'il est surjectif. Rappelons

que si K est un corps, $\mathrm{SL}_2(\mathrm{K})$ est engendré par les transvections (cf. exercice 3.1). Nous allons démontrer qu'il en est de même de $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$, pour $n \in \mathbb{N}^*$ quelconque. On notera $\overline{m}$ la classe modulo n d'un entier m.

Soit $\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$. On va essayer de décomposer M comme un produit de transvections en suivant la preuve déjà utilisée avec un corps. Il y a une complication supplémentaire liée au fait qu'un élément non nul de $\mathbb{Z}/n\mathbb{Z}$ n'est pas forcément inversible. On distingue deux cas.

• Supposons pour commencer que a est inversible dans $\mathbb{Z}/n\mathbb{Z}$. Comme $ad - bc = \overline{1}$, on obtient

$$\begin{pmatrix} \overline{1} & \overline{0} \\ -a^{-1}c & \overline{1} \end{pmatrix}\mathrm{M} = \begin{pmatrix} a & b \\ \overline{0} & d - a^{-1}bc \end{pmatrix} = \begin{pmatrix} a & b \\ \overline{0} & a^{-1} \end{pmatrix}$$

puis

$$\begin{pmatrix} \overline{1} & -ab \\ \overline{0} & \overline{1} \end{pmatrix}\begin{pmatrix} \overline{1} & \overline{0} \\ -a^{-1}c & \overline{1} \end{pmatrix}\mathrm{M} = \begin{pmatrix} a & \overline{0} \\ \overline{0} & a^{-1} \end{pmatrix}.$$

On montre que $\begin{pmatrix} a & \overline{0} \\ \overline{0} & a^{-1} \end{pmatrix}$ est un produit de matrices de transvection. On commence par faire apparaître un $\overline{1}$ en haut à gauche en écrivant

$$\begin{pmatrix} \overline{1} & -\overline{1} + a^{-1} \\ \overline{0} & \overline{1} \end{pmatrix}\begin{pmatrix} \overline{1} & \overline{0} \\ \overline{1} & \overline{1} \end{pmatrix}\begin{pmatrix} a & \overline{0} \\ \overline{0} & a^{-1} \end{pmatrix} = \begin{pmatrix} \overline{1} & a^{-1}(a^{-1} - \overline{1}) \\ a & a^{-1} \end{pmatrix}.$$

Ensuite on écrit

$$\begin{pmatrix} \overline{1} & \overline{0} \\ -a & \overline{1} \end{pmatrix}\begin{pmatrix} \overline{1} & a^{-1}(a^{-1} - \overline{1}) \\ a & a^{-1} \end{pmatrix} = \begin{pmatrix} \overline{1} & a^{-1}(a^{-1} - \overline{1}) \\ \overline{0} & \overline{1} \end{pmatrix}.$$

L'inverse d'une matrice de transvection étant une matrice de transvection, on obtient que M est un produit de matrices de transvection.

• Dans le cas où a n'est pas inversible, on considère

$$\begin{pmatrix} \overline{1} & \lambda \\ \overline{0} & \overline{1} \end{pmatrix}\mathrm{M} = \begin{pmatrix} a + c\lambda & b + d\lambda \\ c & d \end{pmatrix}.$$

On va essayer de choisir λ de sorte que $a + \lambda c$ soit inversible pour se ramener au cas précédent. Notons $p_1, \ldots, p_k$ les diviseurs premiers de n. On considère des entiers α, β, γ, δ tels que $a = \overline{\alpha}$, $b = \overline{\beta}$, $c = \overline{\gamma}$, $d = \overline{\gamma}$. Puisque a n'est pas inversible, α n'est pas premier avec n ; ils ont des facteurs premiers en commun. Supposons par exemple que α est divisible par $p_1, \ldots, p_r$ $(r \leqslant k)$ et n'est pas divisible par $p_{r+1}, \ldots, p_k$. De $ad - bc = 1$, on tire $\alpha\delta - \beta\gamma \equiv 1 \pmod{p_i}$, pour tout i, donc $-\beta\gamma \equiv 1 \pmod{p_i}$ pour $1 \leqslant i \leqslant r$. On en déduit que γ est premier avec p_i pour $1 \leqslant i \leqslant r$.

Posons $\mu = \prod_{i=r+1}^{k} p_i$ et $\lambda = \overline{\mu}$. On vérifie que $\alpha + \gamma\mu$ n'est divisible par

aucun des entiers p_i donc $a+\lambda c$ est inversible dans $\mathbb{Z}/n\mathbb{Z}$. Ce qui précède montre que $\begin{pmatrix} \overline{1} & \lambda \\ \overline{0} & \overline{1} \end{pmatrix}$M est produit de matrices de transvection. Il en est donc de même de M.

On peut alors conclure quant à la surjectivité de ψ_n. En effet, le groupe $\operatorname{Im}\psi_n$ contient toutes les matrices de transvection puisque si $\mu \in \mathbb{Z}$, $\begin{pmatrix} \overline{1} & \overline{\mu} \\ \overline{0} & \overline{1} \end{pmatrix}$ est l'image de $\begin{pmatrix} 1 & \mu \\ 0 & 1 \end{pmatrix}$, et de même pour les transposées. Ce qui précède montre donc que $\operatorname{Im}\psi_n = \mathrm{SL}(\mathbb{Z}/n\mathbb{Z})$.

2. En général, l'application $\varphi_n : \mathrm{GL}_2(\mathbb{Z}) \longrightarrow \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ n'est pas surjective. En effet les matrices inversibles de $\mathrm{GL}_2(\mathbb{Z})$ sont les matrices dont le déterminant est inversible dans $\mathbb{Z}$ c'est-à-dire égal à ± 1. On a donc pour tout $\mathrm{A} \in \mathrm{GL}_2(\mathbb{Z})$, $\det(\varphi_n(\mathrm{A})) = \pm\overline{1}$. Or l'image de $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ par le déterminant est l'ensemble des éléments inversibles de $\mathbb{Z}/n\mathbb{Z}$ et cet ensemble contient strictement $\{\pm\overline{1}\}$ pour $n \geqslant 5$.

Pour $n = 2$, on a $\mathrm{SL}_2(\mathbb{Z}/2\mathbb{Z}) = \mathrm{GL}_2(\mathbb{Z}/2\mathbb{Z})$ et l'application φ_2 est surjective.

Pour $n = 3$ ou 4, l'ensemble $\varphi_n(\mathrm{GL}_2(\mathbb{Z}))$ contient $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ et $\begin{pmatrix} -\overline{1} & \overline{0} \\ \overline{0} & \overline{1} \end{pmatrix} = \varphi_n\begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. Toute matrice de déterminant $-\overline{1}$ est produit de $\begin{pmatrix} -\overline{1} & \overline{0} \\ \overline{0} & \overline{1} \end{pmatrix}$ et d'un élément de $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ donc appartient à $\varphi_n(\mathrm{GL}_2(\mathbb{Z})$ de sorte que l'application est surjective.

Conclusion. L'application $\varphi_n : \mathrm{GL}_2(\mathbb{Z}) \longrightarrow \mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ est surjective si et seulement si $n \leqslant 4$.

3. Il est évident que $\operatorname{Ker}\psi_n \subset \operatorname{Ker}\psi_d$: si une matrice $\mathrm{A} \in \mathrm{SL}_2(\mathbb{Z})$ donne l'identité lorsqu'on la réduit modulo n, c'est *a fortiori* aussi le cas lorsqu'on la réduit modulo d puisque d divise n. On a de même $\operatorname{Ker}\psi_m \subset \operatorname{Ker}\psi_d$ et par suite $\operatorname{Ker}\psi_n \operatorname{Ker}\psi_m \subset \operatorname{Ker}\psi_d$. L'autre inclusion est beaucoup moins évidente (il n'est même pas évident que $\operatorname{Ker}\psi_n \operatorname{Ker}\psi_m$ soit un sous-groupe !). Si on travaillait avec des groupes finis on pourrait utiliser un argument de cardinalité. C'est ce qu'on va faire en introduisant [1] un groupe fini intermédiaire, à savoir $\mathbb{Z}/e\mathbb{Z}$ où e désigne le ppcm de n et m. Notons ρ_n (resp. ρ_m et ρ_d) le morphisme naturel de $\mathrm{SL}_2(\mathbb{Z}/e\mathbb{Z})$ dans $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ (resp. $\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})$ et $\mathrm{SL}_2(\mathbb{Z}/d\mathbb{Z})$). On a bien entendu $\psi_n = \rho_n \circ \psi_e$ et de même pour les autres. En particulier ρ_n est un morphisme surjectif puisque ψ_n l'est d'après la première question.

• On a $\operatorname{Ker}\psi_n = \psi_e^{-1}(\operatorname{Ker}\rho_n)$, et de même pour les autres noyaux. Il suffit alors de prouver que $\operatorname{Ker}\rho_d = \operatorname{Ker}\rho_n \operatorname{Ker}\rho_m$. En effet, supposons que cela soit établi et montrons que $\operatorname{Ker}\psi_n \operatorname{Ker}\psi_m = \operatorname{Ker}\psi_d$. On a déjà vu l'inclusion triviale ci-dessus. Soit $\mathrm{A} \in \operatorname{Ker}\psi_d$. Alors $\psi_e(\mathrm{A}) \in \operatorname{Ker}\rho_d$

1. Nous en profitons pour remercier M. Franck Taïeb qui nous a suggéré cette idée.

donc elle peut s'écrire $\psi_e(A) = B'C'$ avec $B' \in \operatorname{Ker}\rho_n$ et $C' \in \operatorname{Ker}\rho_m$. Comme ψ_e est surjectif, on peut écrire $B' = \psi_e(B)$ avec $B \in \operatorname{Ker}\psi_n$ et $C' = \psi_e(C)$ avec $C \in \operatorname{Ker}\psi_m$. On a alors $\psi_e(A) = \psi_e(BC)$ et donc $A = BCD$ avec $D \in \operatorname{Ker}\psi_e$. C'est une décomposition de la forme voulue car $CD \in \operatorname{Ker}\psi_m$ puisque m divise e.

• On va donc s'attacher à prouver que $\operatorname{Ker}\rho_d = \operatorname{Ker}\rho_n \operatorname{Ker}\rho_m$, égalité qui a lieu dans le groupe fini $\mathrm{SL}_2(\mathbb{Z}/e\mathbb{Z})$. Pour cela, on note que l'inclusion $\operatorname{Ker}\rho_n \operatorname{Ker}\rho_m \subset \operatorname{Ker}\rho_d$ est toujours triviale. Pour obtenir l'égalité on va montrer que les deux parties ont le même cardinal ! Notons tout d'abord que $\operatorname{Ker}\rho_n \cap \operatorname{Ker}\rho_m = \{\psi_e(\mathrm{I}_2)\}$ de sorte que le cardinal de $\operatorname{Ker}\rho_n \operatorname{Ker}\rho_m$ est égal au produit $|\operatorname{Ker}\rho_n| \times |\operatorname{Ker}\rho_m|$. En effet, considérons un élément de $\operatorname{Ker}\rho_n \cap \operatorname{Ker}\rho_m$ qu'on écrit $\psi_e(A)$ où $A \in \mathrm{SL}_2(\mathbb{Z})$. On a alors $A \equiv \mathrm{I}_2\ [n]$ et $A \equiv \mathrm{I}_2\ [m]$ de sorte que $A \equiv \mathrm{I}_2\ [e]$ et finalement $\psi_e(A) = \psi_e(\mathrm{I}_2)$. Il est alors facile d'en déduire que l'application qui à $(U, V) \in \operatorname{Ker}\rho_n \times \operatorname{Ker}\rho_m$ associe le produit UV est injective et établit donc une bijection entre $\operatorname{Ker}\rho_n \times \operatorname{Ker}\rho_m$ et $\operatorname{Ker}\rho_n \operatorname{Ker}\rho_m$.

• Les cardinaux des noyaux ci-dessus se calculent : par surjectivité de ρ_n on a $|\operatorname{Ker}\rho_n| \times |\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})| = |\mathrm{SL}_2(\mathbb{Z}/e\mathbb{Z})|$. Ainsi, on doit, pour conclure, démontrer la formule suivante :

$$|\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})| \times |\mathrm{SL}_2(\mathbb{Z}/m\mathbb{Z})| = |\mathrm{SL}_2(\mathbb{Z}/d\mathbb{Z})| \times |\mathrm{SL}_2(\mathbb{Z}/e\mathbb{Z})|.$$

Cela nous amène donc au problème combinatoire suivant : quel est le cardinal de $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$, $n \geqslant 1$ étant quelconque ? Le lecteur a déjà rencontré un cas particulier de cette question dans l'exercice 3.23 où n était une puissance de 2. On va ici aussi se ramener aux puissances de nombres premiers. Notons $n = p_1^{\alpha_1} \dots p_\zeta^{\alpha_\zeta}$ la factorisation de n en nombres premiers. Comme précédemment, on dispose d'un morphisme naturel f de $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ dans le groupe produit $\mathrm{SL}_2(\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z}) \times \cdots \times \mathrm{SL}_2(\mathbb{Z}/p_\zeta^{\alpha_\zeta}\mathbb{Z})$ qui à la classe d'une matrice A modulo n associe le r-uplet formé de ses classes modulo $p_1^{\alpha_1}, \dots, p_r^{\alpha_r}$. Ce morphisme est injectif, car si $A \equiv \mathrm{I}_2\ [p_i^{\alpha_i}]$ pour tout i alors $A \equiv \mathrm{I}_2\ [n]$. Il est aussi surjectif par le théorème chinois. En effet, si $A_1, \dots, A_r$ sont des matrices données dans $\mathrm{SL}_2(\mathbb{Z})$ il existe A telle que $A \equiv A_i\ [p_i^{\alpha_i}]$ pour tout i. On a alors $\det A \equiv \det A_i \equiv 1\ [p_i^{\alpha_i}]$ pour tout i et donc $\det A \equiv 1\ [n]$, de sorte que la réduction modulo n de A est dans $\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})$ et a pour image le r-uplet formé des réductions des matrices A_i. Il en découle que f est un isomorphisme et en particulier que

$$|\mathrm{SL}_2(\mathbb{Z}/n\mathbb{Z})| = |\mathrm{SL}_2(\mathbb{Z}/p_1^{\alpha_1}\mathbb{Z})| \times \cdots \times |\mathrm{SL}_2(\mathbb{Z}/p_s^{\alpha_s}\mathbb{Z})|.$$

Il nous reste donc à trouver le cardinal de $\mathrm{SL}_2(\mathbb{Z}/p^\alpha\mathbb{Z})$, où p est un nombre premier. On procède comme dans l'exercice 3.23 avec $p = 2$. Considérons le morphisme naturel de $\mathrm{SL}_2(\mathbb{Z}/p^\alpha\mathbb{Z})$ dans $\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})$. Il est surjectif et son noyau est de cardinal $(p^{\alpha-1})^3$. Comme

$$|\mathrm{SL}_2(\mathbb{Z}/p\mathbb{Z})| = \frac{|\mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z})|}{p-1} = \frac{(p^2-1)(p^2-p)}{p-1} = p(p-1)(p+1),$$

on obtient finalement $|\mathrm{SL}_2(\mathbb{Z}/p^a\mathbb{Z})| = (p-1)(p+1)p^{3a-2}$. Pour conclure, il suffit d'observer que pour tout nombre premier p on a $\nu_p(n)+\nu_p(m) = \nu_p(d)+\nu_p(e)$ où ν_p désigne la valuation p-adique. Cela provient simplement de ce que $nm = de!$ ◁

3.25. Surjection de $\mathbf{GL_2}(\mathbb{Z}/m\mathbb{Z})$ dans $\mathbf{GL_2}(\mathbb{Z}/n\mathbb{Z})$

Soient m, n dans $\mathbb{N}^*$ avec $n|m$. Trouver une surjection naturelle de $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ dans $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$. On pourra d'abord étudier le cas où m et n ont les mêmes facteurs premiers.

(**École normale supérieure**)

▷ **Solution.**

Il y a une surjection naturelle de $\mathbb{Z}/m\mathbb{Z}$ sur $\mathbb{Z}/n\mathbb{Z}$ à savoir l'application qui, pour tout entier a, associe à la classe de a modulo m sa classe modulo n. Cette application est bien définie car si $a \equiv b \pmod m$ alors $a \equiv b \pmod n$. Elle induit naturellement une application π de $\mathcal{M}_2(\mathbb{Z}/m\mathbb{Z})$ dans $\mathcal{M}_2(\mathbb{Z}/n\mathbb{Z})$. On va commencer par montrer que π envoie $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ dans $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$.

Étant donnée une matrice A à coefficients dans $\mathbb{Z}$ on notera $\pi_n(\mathrm{A})$ la matrice de $\mathcal{M}_2(\mathbb{Z}/n\mathbb{Z})$ obtenue par réduction des coefficients de A modulo n.

Considérons une matrice de $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ que l'on écrit $\pi_m(\mathrm{A})$ avec $\mathrm{A} \in \mathcal{M}_2(\mathbb{Z})$. Il existe $\mathrm{B} \in \mathcal{M}_2(\mathbb{Z})$ tel que $\pi_m(\mathrm{A})\pi_m(\mathrm{B}) = \pi_m(\mathrm{AB}) = \pi_m(\mathrm{I}_2)$. On a alors $\pi_n(\mathrm{A})\pi_n(\mathrm{B}) = \pi_n(\mathrm{AB}) = \pi_n(\mathrm{I}_2)$ car n divise m et $\pi_n(\mathrm{A})$ est inversible. Ainsi π va de $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ dans $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$. On va démontrer qu'elle est surjective.

• On suppose pour commencer que m et n ont les mêmes facteurs premiers, chaque facteur apparaissant avec un exposant plus élevé dans m. Ainsi m apparaît comme le produit de n et de nombres premiers qui divisent n. En fait, il suffit de démontrer le résultat pour $m = np$, avec p premier divisant n et de réitérer le procédé.

Considérons une matrice de $\mathrm{GL}_2(\mathbb{Z}/n\mathbb{Z})$ que l'on écrit $\pi_n(\mathrm{A})$ avec $\mathrm{A} \in \mathcal{M}_2(\mathbb{Z})$. Il existe $\mathrm{B} \in \mathcal{M}_2(\mathbb{Z})$ tel que $\pi_n(\mathrm{A})\pi_n(\mathrm{B}) = \pi_n(\mathrm{AB}) = \pi_n(\mathrm{I}_2)$. Autrement dit, il existe $\mathrm{U} \in \mathcal{M}_2(\mathbb{Z})$ tel que $\mathrm{AB} = \mathrm{I}_2 + n\mathrm{U}$. On cherche $\mathrm{A}' \in \mathcal{M}_2(\mathbb{Z})$ telle que $\pi_m(\mathrm{A}') \in \mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$ et $\pi_n(\mathrm{A}') = \pi_n(\mathrm{A})$. On cherche donc A' sous la forme $\mathrm{A}' = \mathrm{A} + n\mathrm{V}$, avec $\mathrm{V} \in \mathcal{M}_2(\mathbb{Z})$. On a alors $\mathrm{A}'\mathrm{B} = \mathrm{AB} + n\mathrm{AV} = \mathrm{I}_2 + n\mathrm{U} + n\mathrm{AV}$ et $\pi_m(\mathrm{B})$ est l'inverse de $\pi_m(\mathrm{A}')$ si et seulement si $\mathrm{A}'\mathrm{B} \equiv \mathrm{I}_2\ [m]$, c'est-à-dire si $n\mathrm{U} + n\mathrm{AV} \equiv 0\ [np]$ ou encore $\mathrm{U} + \mathrm{AV} \equiv 0\ [p]$. Par hypothèse, p divise n donc $\mathrm{AB} = \mathrm{I}_2 + n\mathrm{U}$

implique $AB \equiv I_2\ [p]$. On en déduit $ABU \equiv U\ [p]$, ce qui montre que l'on peut choisir $V = -BU$.

• Examinons maintenant le cas où m possède des facteurs premiers qui n'apparaissent pas dans n. Compte tenu de ce qui précède, il suffit de démontrer la propriété pour $m = pn$, avec p premier ne divisant pas n et de réitérer le procédé.

On reprend la démonstration précédente. Elle est inchangée jusqu'à $U + AV \equiv 0\ [p]$. Si $\pi_p(A)$ est inversible dans $\mathcal{M}_2(\mathbb{Z}/p\mathbb{Z})$ d'inverse $\pi_p(C)$, on peut prendre $V = -CU$. Montrons qu'on peut se ramener à ce cas. Il suffit de montrer, dans le cas où $\pi_p(A)$ n'est pas inversible, qu'on peut trouver $S \in \mathcal{M}_2(\mathbb{Z})$ telle que $\pi_p(A + nS)$ soit inversible. On remarque que $\overline{n}$ est inversible dans le corps $\mathbb{Z}/p\mathbb{Z}$; on note n' un entier tel que $nn' \equiv 1\ [p]$. Si $\pi_p(A) = 0$, la matrice $S = n'I_2$ convient. Reste à traiter le cas où $\pi_p(A)$ est de rang 1. Parmi les deux vecteurs colonnes C_1 et C_2 de A, l'un n'est pas nul modulo p. Supposons que c'est C_1. On a alors, pour $X \in \mathbb{Z}^2$,

$$\det(C_1, C_2 + nX) \equiv n \det(C_1, X)\ \ [p].$$

Si on choisit X tel que tel que $\pi_p(X)$ ne soit pas colinéaire à $\pi_p(C_1)$ (c'est possible : dans $(\mathbb{Z}/p\mathbb{Z})^2$, il y a $p^2 - p \geqslant 2$ vecteurs non colinéaires à $\pi_p(C_1)$), la matrice dont les colonnes sont C_1 et $C_2 + nX$ est inversible, car son déterminant n'est pas nul. Si S est la matrice dont les colonnes sont 0 et X, $\pi_p(A + nS)$ est inversible, ce qui permet de conclure la démonstration. ◁

Chapitre 4

Topologie et matrices, exponentielle

Nous avons regroupé dans ce chapitre des exercices qui concernent des questions topologiques formulées dans un contexte matriciel. Le corps de base K *sera donc toujours* $\mathbb{R}$ *ou* $\mathbb{C}$. *Comme* $\mathcal{M}_n(\mathrm{K})$ *est un espace vectoriel de dimension finie, toutes les normes sur* $\mathcal{M}_n(\mathrm{K})$ *sont équivalentes et il y a donc une unique topologie d'espace normé. On prendra toutefois souvent une norme triple induite par une norme sur* K^n.

Les premiers exercices concernent des questions d'adhérence et d'intérieur pour diverses parties remarquables de $\mathcal{M}_n(\mathrm{K})$.

4.1. Matrices de rang r

Trouver l'adhérence et l'intérieur de $\{\mathrm{M} \in \mathcal{M}_n(\mathbb{C}),\ \mathrm{rg}\,\mathrm{M} = r\}$.

(École normale supérieure)

▷ **Solution.**

Posons $\mathrm{V}_r = \{\mathrm{M} \in \mathcal{M}_n(\mathbb{C}),\ \mathrm{rg}\,\mathrm{M} = r\}$ pour $r \in [\![0, n]\!]$. L'ensemble V_0 est le singleton $\{0\}$ qui est un fermé d'intérieur vide et l'ensemble V_n est le groupe linéaire $\mathrm{GL}_n(\mathbb{C})$. Il est bien connu que ce dernier est un ouvert dense de $\mathcal{M}_n(\mathbb{C})$. Rappelons la preuve de ce fait : comme la fonction det est continue sur $\mathcal{M}_n(\mathbb{C})$ l'ensemble $\mathrm{V}_n = \mathrm{GL}_n(\mathbb{C})$ qui est l'image réciproque de l'ouvert $\mathbb{C}^*$ est un ouvert. Si $\mathrm{A} \in \mathcal{M}_n(\mathbb{C})$ est une matrice quelconque, la matrice $\mathrm{A} - \lambda \mathrm{I}_n$ est inversible sauf pour un nombre fini de valeurs de λ (à savoir les valeurs propres de A). Si $(\lambda_k)_{k\geqslant 0}$ est une suite de nombres complexes qui converge vers 0 en évitant les éléments de $\mathrm{Sp}\,\mathrm{A}$, la suite $(\mathrm{A} - \lambda_k \mathrm{I}_n)_{k\geqslant 0}$ est une suite de V_n qui converge vers A. Ainsi V_n est un ouvert dense.

On en déduit tout de suite que les ensembles V_r pour $r < n$ sont tous d'intérieur vide puisqu'ils sont inclus dans le fermé $\mathcal{M}_n(\mathbb{C}) \setminus \mathrm{V}_n$ qui est d'intérieur vide. Il nous reste donc à déterminer l'adhérence de V_r. Il est facile de voir que cette adhérence contient toutes les matrices de rang $\leqslant r$. En effet, si $\mathrm{A} \in \mathrm{V}_s$ avec $s \leqslant r$ on peut trouver P, Q inversibles telles que $\mathrm{A} = \mathrm{P}\mathrm{J}_s\mathrm{Q}$ où $\mathrm{J}_s = \begin{pmatrix} \mathrm{I}_s & 0 \\ 0 & 0 \end{pmatrix}$. La suite des matrices

$$\mathrm{A}_k = \mathrm{P}\,\mathrm{Diag}\left(1, \ldots, 1, \frac{1}{k}, \ldots, \frac{1}{k}, 0, \ldots, 0\right)\mathrm{Q}$$

où il y a $r-s$ termes $\frac{1}{k}$, est formée de matrices de rang r et converge vers A lorsque k vers l'infini. En fait, on vient d'exhiber l'adhérence de V_r. En effet, si A est une matrice de rang $\geqslant r+1$ il existe une sous-matrice carrée de A de taille $r+1$ qui est inversible. Par continuité du déterminant, cette sous-matrice reste inversible pour toute matrice assez proche de A. Autrement dit $\{M \in \mathcal{M}_n(\mathbb{C}),\ \operatorname{rg} M \geqslant r+1\}$ est ouvert et par conséquent $\{M \in \mathcal{M}_n(\mathbb{C}),\ \operatorname{rg} M \leqslant r\} = \bigcup_{s \leqslant r} V_s$ est fermé pour tout r.

Comme on a

$$\bigcup_{s \leqslant r} V_s \subset \overline{V_r} \subset \overline{\bigcup_{s \leqslant r} V_s} = \bigcup_{s \leqslant r} V_s$$

on peut conclure que $\overline{V_r} = \bigcup_{s \leqslant r} V_s$. ◁

Il est important de noter que tous les résultats de cet exercice restent vrais dans $\mathcal{M}_n(\mathbb{R})$, les preuves étant analogues.

4.2. Matrices diagonalisables complexes

Trouver l'adhérence et l'intérieur de l'ensemble des matrices diagonalisables de $\mathcal{M}_n(\mathbb{C})$.

(École normale supérieure, école polytechnique)

▷ **Solution.**

Notons $\mathcal{D}_n$ l'ensemble des matrices diagonalisables de $\mathcal{M}_n(\mathbb{C})$. On supposera $n \geqslant 2$, sans quoi la question est triviale.

• On va commencer par prouver que $\mathcal{D}_n$ est dense dans $\mathcal{M}_n(\mathbb{C})$. Pour cela, on utilise le fait que toute matrice A de $\mathcal{M}_n(\mathbb{C})$ est semblable à une matrice triangulaire T. Écrivons $A = PTP^{-1}$ avec T triangulaire supérieure et P inversible. Pour $k \in \mathbb{N}^*$, soit $T_k = (t_{ij}^{(k)})$ la matrice dont les coefficients sont égaux à ceux de T, sauf les coefficients diagonaux qu'on définit par $t_{ii}^{(k)} = t_{ii} + \frac{i}{k}$. Pour i et j indices distincts, on a $t_{ii}^{(k)} \neq t_{jj}^{(k)}$ si $t_{ii} = t_{jj}$ et sinon $\lim_{k \to +\infty} t_{ii}^{(k)} - t_{jj}^{(k)} = t_{ii} - t_{jj} \neq 0$. Pour k assez grand, tous les termes diagonaux de T_k sont distincts deux à deux. Autrement dit, à partir d'un certain rang T_k possède n valeurs propres distinctes et est donc diagonalisable. Il en est de même de PT_kP^{-1}. Comme la suite $(PT_kP^{-1})_{k \geqslant 1}$ converge vers A, on a le résultat voulu.

• L'ensemble $\mathcal{D}_n$ n'est pas ouvert : par exemple la matrice I_n n'est pas intérieure à $\mathcal{D}_n$ car $I_n + \varepsilon E_{12}$ tend vers I_n lorsque ε tend vers 0 et la matrice $I_n + \varepsilon E_{12}$ n'est pas diagonalisable pour ε non nul. Plus généralement, toute matrice diagonalisable A ayant une valeur propre au moins double n'est pas intérieure à $\mathcal{D}_n$. En effet, le même argument s'applique : A s'écrit $A = P^{-1}DP$ avec $D = \operatorname{Diag}(\lambda, \lambda, \lambda_3, \ldots, \lambda_n)$ et la

suite $A_k = P^{-1}\left(D + \frac{1}{k}E_{12}\right)P$ est formée de matrices non diagonalisables et elle converge vers A. Ainsi l'intérieur de $\mathcal{D}_n$ est inclus dans l'ensemble des matrices ayant n valeurs propres distinctes (une telle matrice est nécessairement diagonalisable). Il y a égalité et pour le voir on fait appel à la notion de discriminant (voir l'exercice 1.17). Si A possède n valeurs propres distinctes le discriminant de χ_A est non nul. Par continuité de l'application qui à une matrice A associe le discriminant de χ_A (l'exercice donné en référence montre que ce discriminant est une fonction polynomiale des coefficients de A), les matrices proches de A ont aussi n valeurs propres distinctes et sont donc dans $\mathcal{D}_n$. ◁

Le résultat de densité prouvé dans cet exercice est important. Ainsi, lorsqu'on cherche à montrer une propriété qui dépend continûment d'une matrice $A \in \mathcal{M}_n(\mathbb{C})$, *on peut se contenter de la prouver lorsque* A *est diagonalisable, ce qui est souvent beaucoup plus facile, et l'étendre par un argument de densité (voir par exemple l'exercice 2.69).*

Notons que la première partie de la solution montre que l'ensemble des matrices ayant n valeurs propres distinctes est aussi dense dans $\mathcal{M}_n(\mathbb{C})$. *On sait que la dimension du commutant d'une telle matrice est n. Cela résulte du fait qu'une matrice commute avec une matrice diagonale à coefficients diagonaux distincts si et seulement si elle est diagonale. On en déduit que l'ensemble des matrices dont le commutant est de dimension n est dense dans* $\mathcal{M}_n(\mathbb{C})$. *Cette question a également été posée à l'oral des écoles normales supérieures.*

L'exercice suivant étudie la situation sur le corps des réels.

4.3. Matrices diagonalisables réelles

1. Soit $P \in \mathbb{R}[X]$ unitaire de degré n. Montrer que le polynôme P est scindé sur $\mathbb{R}$ si, et seulement si, $|P(z)| \geqslant |\operatorname{Im} z|^n$ pour tout $z \in \mathbb{C}$.

2. Montrer que l'ensemble $\mathcal{T}_n$ des matrices trigonalisables de $\mathcal{M}_n(\mathbb{R})$ est fermé.

3. Déterminer l'adhérence de l'ensemble $\mathcal{D}_n$ des matrices diagonalisables de $\mathcal{M}_n(\mathbb{R})$ puis l'adhérence de l'ensemble Δ_n des matrices réelles ayant n valeurs propres réelles distinctes.

(**École polytechnique**)

▷ **Solution.**

1. Supposons $|P(z)| \geqslant |\operatorname{Im} z|^n$ pour tout $z \in \mathbb{C}$. Si z_0 est une racine complexe de P on a $\operatorname{Im} z_0 = 0$ donc $z_0 \in \mathbb{R}$. Ainsi toutes les racines de P sont réelles et P est scindé sur $\mathbb{R}$. Réciproquement, supposons P scindé sur $\mathbb{R}$ et posons $P = (X - x_1)\dots(X - x_n)$, les réels x_i n'étant pas

forcément distincts. Soit $z = x + iy \in \mathbb{C}$ avec x, y réels. On a

$$|P(z)| = \prod_{i=1}^{n} |z - x_i| = \prod_{i=1}^{n} |x - x_i + iy| \geqslant \prod_{i=1}^{n} |y| = |\operatorname{Im} z|^n.$$

2. Montrons que $\mathcal{T}_n$ est fermé en utilisant la caractérisation séquentielle. Soit A_k une suite de matrices trigonalisables qui converge vers une matrice B. Pour tout k le polynôme caractéristique de A_k est scindé donc, si $z \in \mathbb{C}$ on a $|\chi_{A_k}(z)| \geqslant |\operatorname{Im} z|^n$. Comme le polynôme caractéristique dépend continûment de la matrice on peut faire tendre k vers l'infini dans cette inégalité à z fixé pour obtenir $|\chi_B(z)| \geqslant |\operatorname{Im} z|^n$. La réciproque de la question précédente montre que χ_B est scindé sur $\mathbb{R}$ et donc que B est trigonalisable.

3. Comme $\Delta_n \subset \mathcal{D}_n \subset \mathcal{T}_n$ on a $\overline{\Delta_n} \subset \overline{\mathcal{D}_n} \subset \mathcal{T}_n$ et on montre exactement comme dans l'exercice précédent que les deux inclusions sont des égalités. ◁

Comme dans le cas complexe on montre que l'intérieur de $\mathcal{D}_n$ est égal à Δ_n : il est même plus facile de montrer à l'aide du théorème des valeurs intermédiaires que si χ_A est scindé à racines simples sur $\mathbb{R}$, il en est de même de χ_B pour B *assez proche de* A *(voir exercice 5.40 du tome algèbre 1).*

L'exercice suivant se pose toujours la même question mais pour l'ensemble des matrices cycliques. On rappelle que A *est dite cyclique lorsque $\mu_A = \chi_A$, ce qui équivaut à l'existence d'un vecteur* X *tel que $\mathbb{C}^n = \operatorname{Vect}(X, AX, \ldots, A^{n-1}X)$ (se reporter à l'exercice 2.52).*

4.4. Matrices cycliques

Trouver l'adhérence et l'intérieur de l'ensemble des matrices cycliques de $\mathcal{M}_n(\mathbb{C})$.

(École normale supérieure)

▷ **Solution.**

• Soit $A \in \mathcal{M}_n(\mathbb{C})$ une matrice cyclique et $X \in \mathbb{C}^n$ un vecteur tel que $\mathbb{C}^n = \operatorname{Vect}(X, AX, \ldots, A^{n-1}X)$. La famille $(X, AX, \ldots, A^{n-1}X)$ est alors une base de $\mathbb{C}^n$. Notons B la base canonique de $\mathbb{C}^n$. La fonction f qui à $M \in \mathcal{M}_n(\mathbb{C})$ associe $f(M) = \det_B(X, MX, \ldots, M^{n-1}X)$ est continue car polynomiale en les coefficients de M. Par hypothèse, elle prend une valeur non nulle en A. Donc elle ne s'annule pas sur tout un voisinage V de A et toute matrice M de V est alors cyclique. Autrement dit, l'ensemble des matrices cycliques de $\mathcal{M}_n(\mathbb{C})$ est ouvert.

• On va voir que cet ouvert est dense. Cherchons d'abord des

exemples de matrices cycliques simples. Une matrice diagonalisable n'est pas forcément cyclique : par exemple I_n n'est pas cyclique (pour $n \geqslant 2$). En fait, une matrice diagonalisable dont le spectre est de cardinal n est cyclique. En effet, notons $\lambda_1, \ldots, \lambda_n$ les valeurs propres distinctes d'une telle matrice A et $X_1, \ldots, X_n$ une base de vecteurs propres associés à ces valeurs propres. Considérons le vecteur $X = X_1 + X_2 + \cdots + X_n$. La matrice de la famille $(X, AX, \ldots, A^{n-1}X)$ dans la base $(X_1, \ldots, X_n)$ est une matrice de Vandermonde : elle est inversible et la famille $(X, AX, \ldots, A^{n-1}X)$ est une base de $\mathbb{C}^n$. Or on a vu dans la solution de l'exercice 4.2 que l'ensemble des matrices diagonalisables ayant n valeurs propres distinctes est dense dans $\mathcal{M}_n(\mathbb{C})$. Il en découle que l'ouvert des matrices cycliques est *a fortiori* dense dans $\mathcal{M}_n(\mathbb{C})$. ◁

Les trois exercices qui suivent concernent des racines de la matrice identité.

4.5. Racines de l'identité

Trouver l'adhérence de $\mathcal{A} = \{M \in \mathcal{M}_n(\mathbb{C}),\ \exists p \in \mathbb{N}^*,\ M^p = I_n\}$.
(École polytechnique)

▷ **Solution.**

Si $M \in \mathcal{A}$ et vérifie $M^p = I_n$, les valeurs propres de M sont des racines p-ièmes de l'unité. De plus M est diagonalisable car le polynôme $X^p - 1$ annule A et est scindé à racines simples. On remarque que réciproquement, si M est une matrice diagonalisable de $\mathcal{M}_n(\mathbb{C})$ dont les valeurs propres sont toutes des racines p-ièmes de 1, on a $M^p = I_n$ et M appartient à $\mathcal{A}$.

Regardons le cas $n = 1$ qui va nous donner une idée du résultat général. En identifiant $\mathcal{M}_1(\mathbb{C})$ avec $\mathbb{C}$, $\mathcal{A}$ est l'ensemble des racines de l'unité. Son adhérence est alors le cercle unité S^1 de $\mathbb{C}$. En effet, $\mathcal{A}$ est l'image de $\pi\mathbb{Q}$ par le morphisme surjectif $f : \mathbb{R} \to S^1$ défini par $f(\theta) = e^{i\theta}$. Comme f est continu, et comme $\pi\mathbb{Q}$ est dense dans $\mathbb{R}$, on a $\overline{\mathcal{A}} = f(\mathbb{R}) = S^1$. Revenons au cas général. Nous allons montrer que l'adhérence de $\mathcal{A}$ est exactement l'ensemble F des matrices de $\mathcal{M}_n(\mathbb{C})$ dont les valeurs propres sont toutes de module 1. On a évidemment $\mathcal{A} \subset F$. Montrons que $F \subset \overline{\mathcal{A}}$. Soit $M \in F$. La matrice M peut s'écrire $P^{-1}TP$, avec P inversible et T triangulaire supérieure. Comme $\mathcal{A}$ est stable par conjugaison, il suffit de prouver que T est dans l'adhérence de $\mathcal{A}$: si (T_p) est une suite de $\mathcal{A}$ qui converge vers T, alors la suite $(P^{-1}T_pP)$ converge vers M et les matrices $P^{-1}T_pP$ sont toutes dans $\mathcal{A}$. Les termes diagonaux de $T = (t_{ij})$ sont de module 1. Pour tout $k \in [\![1, n]\!]$, on considère une suite $t_{kk}^{(p)}$ de racines de l'unité qui converge vers t_{kk} (une telle suite existe

d'après le cas $n = 1$). On considère alors la matrice T_p dont les termes non diagonaux sont les termes de T et les termes diagonaux sont les $t_{kk}^{(p)}$. La suite (T_p) converge vers T. Pour que les matrices T_p soient dans $\mathcal{A}$ il suffit de faire en sorte que les coefficients diagonaux soient deux à deux distincts, car dans ce cas T_p est diagonalisable et l'une de ses puissances vaut I_n. Or cela n'est pas très difficile à réaliser. Pour $1 \leqslant k \leqslant n$, considérons $\theta_k \in \mathbb{R}$ tel que $t_{kk} = e^{i\theta_k}$. Pour $p \in \mathbb{N}^*$, on prend $t_{kk}^{(p)} = e^{2i\pi\mu_k^{(p)}}$ avec $\mu_k^{(p)} = \dfrac{k}{p} + \dfrac{1}{p}\mathrm{E}\left(\dfrac{p\theta_k}{2\pi}\right)$. On a

$$|e^{2i\pi\mu_k^{(p)}} - e^{i\theta_k}| \leqslant |2\pi\mu_k^{(p)} - \theta_k| \leqslant \frac{2\pi k}{p} + \frac{2\pi}{p},$$

ce qui montre que $t_{kk}^{(p)}$ converge vers t_{kk} pour tout k. Vérifions que les termes diagonaux de T_p sont deux à deux distincts pour p assez grand. Soient k et ℓ deux entiers distincts dans $[\![1, n]\!]$.

- Si $\theta_k \equiv \theta_\ell\ [2\pi]$, alors $2\pi\mu_k - 2\pi\mu_\ell \equiv \dfrac{2\pi(k-\ell)}{p}\ [2\pi]$.
- Si $\theta_k \not\equiv \theta_\ell\ [2\pi]$, alors $2\pi\mu_k - 2\pi\mu_\ell \underset{p\to\infty}{\sim} \theta_k - \theta_\ell$.

Cela prouve le résultat.

Pour montrer que F est l'adhérence de $\mathcal{A}$, il suffit de vérifier que F est fermé. Soit (M_p) une suite de matrices de F qui converge vers $M \in \mathcal{M}_n(\mathbb{C})$. On note $\lambda_{1,p}, \ldots, \lambda_{n,p}$ les valeurs propres de M_p (comptées avec multiplicité et rangées dans un ordre quelconque). Les coefficients du polynôme caractéristique étant des fonctions polynomiales des coefficients de la matrice, on a $\chi_M = \lim\limits_{p\to+\infty} \chi_{M_p}$ (*i.e.* les coefficients de χ_M sont les limites des coefficients de χ_{M_p}). D'autre part, la suite $(\lambda_{1,p}, \ldots, \lambda_{n,p})_{p\in\mathbb{N}}$ de $\mathbb{C}^n$ étant bornée, on peut en extraire une suite convergente. Soit $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ une telle extraction et

$$(\lambda_1, \ldots, \lambda_n) = \lim_{p\to+\infty} (\lambda_{1,\varphi(p)}, \ldots, \lambda_{n,\varphi(p)}).$$

Pour tout $p \in \mathbb{N}$, on a $\chi_{M_{\varphi(p)}} = \prod_{i=1}^{n}(X - \lambda_{i,\varphi(p)})$. On en déduit par passage à la limite que $\chi_M = \prod_{i=1}^{n}(X - \lambda_i)$. Les valeurs propres de M sont donc $\lambda_1, \ldots, \lambda_n$. Comme, pour tout $i \in [\![1, n]\!]$, $|\lambda_i| = \lim\limits_{p\to+\infty} |\lambda_{i,\varphi(p)}| = 1$, les valeurs propres de M sont de module 1 et M appartient à F qui est fermé. ◁

Il est facile de généraliser le résultat de cet exercice. Si X *est une partie de* $\mathbb{C}$, *l'adhérence de l'ensemble des matrices* $A \in \mathcal{M}_n(\mathbb{C})$ *telles que* $\mathrm{Sp}\,A \subset X$ *est l'ensemble des matrices* A *telles que* $\mathrm{Sp}\,A \subset \overline{X}$.

Dans l'énoncé suivant on prouve que l'identité est isolée dans l'ensemble des racines q-ièmes de l'identité.

4.6. Point isolé dans l'ensemble des racines q-ièmes de l'identité

Soit $q \in \mathbb{N}^*$. On note E_q l'ensemble des matrices $A \in GL_n(\mathbb{C})$ telles que $A^q = I_n$.

1. Que dire de $A \in E_q$ si 1 est la seule valeur propre de A ?

2. Montrer que I_n est un point isolé de E_q.

3. Soit $A_0 \in E_q$. Montrer l'existence de $\varepsilon > 0$ tel que si $A \in E_q$ et $\|A - A_0\| < \varepsilon$, alors A et A_0 sont semblables.

(École normale supérieure)

▷ **Solution.**

1. Toute matrice A de E_q est annulée par $X^q - 1$ qui est scindé à racines simples, donc est diagonalisable. Si 1 est sa seule valeur propre, la matrice A est semblable, donc égale, à la matrice identité I_n.

2. On munit $\mathcal{M}_n(\mathbb{C})$ d'une norme triple associée à une norme quelconque de $\mathbb{C}^n$. Si $A \in \mathcal{M}_n(\mathbb{C})$ et $\lambda \in \mathrm{Sp}(A)$, de vecteur propre associé X, on a $\|AX\| = |\lambda|\|X\| \leqslant |||A|||\|X\|$ et donc $|\lambda| \leqslant |||A|||$.

Soit alors $A \in E_q$ distincte de I_n. D'après la question 1. A admet une valeur propre λ différente de 1. Cette valeur propre λ est une racine q-ième de l'unité. Il existe donc $k \in \mathbb{N}$, $0 < |k| \leqslant \dfrac{q}{2}$ tel que $\lambda = e^{\frac{2ik\pi}{q}}$. On a alors $|\lambda - 1| = 2\left|\sin\dfrac{k\pi}{q}\right| \geqslant 2\sin\dfrac{\pi}{q}$ car la fonction sinus croît sur $\left[0, \dfrac{\pi}{2}\right]$. Comme $\lambda - 1$ est une valeur propre de $A - I_n$ on a

$$|||A - I_n||| \geqslant |\lambda - 1| \geqslant 2\sin\frac{\pi}{q}.$$

Autrement dit la boule ouverte de centre I_n et de rayon $2\sin\dfrac{\pi}{q}$ ne contient pas d'autre élément de E_q en dehors de l'identité. Cela montre que I_n est un point isolé de E_q.

3. Raisonnons par l'absurde et supposons pour tout $k \in \mathbb{N}^*$, il existe $M_k \in E_q$ tel que $\|M_k - A_0\| < \dfrac{1}{k}$, sans que M_k et A_0 soient semblables. Comme $(M_k)_{k\geqslant 0}$ converge vers A_0, χ_{M_k} converge vers χ_{A_0} (dans $\mathbb{C}_n[X]$) puisque les coefficients de χ_M sont des fonctions continues de M. Or, l'ensemble des polynômes caractéristiques des matrices de E_q est fini : en effet, le spectre de ces matrices est contenu dans l'ensemble des racines q-ième de l'unité. Il s'ensuit que la suite χ_{M_k} est constante à partir d'un rang k_0 et égale à χ_{A_0}. Dans ces conditions, M_{k_0} est semblable à A_0, car ces matrices sont diagonalisables et comme elles ont même polynôme caractéristique, elles ont même spectre et même ordre pour les valeurs propres. C'est la contradiction cherchée. ◁

L'énoncé suivant étudie plus particulièrement le cas des symétries.

4.7. Ensemble des symétries

Soit $\mathcal{S} = \{A \in \mathcal{M}_n(\mathbb{R}), A^2 = I_n\}$.

1. Montrer que I_n est un point isolé de $\mathcal{S}$.

2. Montrer que, si $n \geqslant 2$, une symétrie par rapport à un hyperplan de $\mathbb{R}^n$ n'est pas isolée dans $\mathcal{S}$.

3. Montrer que $\mathcal{S}$ n'est pas compact.

4. Pour $A \in \mathcal{S} \setminus \{\pm I_n\}$ on pose $\delta(A) = \inf_{(X,Y)\in E} \|X - Y\|$, où

$$E = \{(X, Y), X \in \mathrm{Ker}(A - I_n), Y \in \mathrm{Ker}(A + I_n), \|X\| = \|Y\| = 1\}$$

et où $\|\ \|$ est la norme euclidienne sur $\mathbb{R}^n$. Pour $\varepsilon > 0$ on pose $\mathcal{S}_\varepsilon = \{A \in \mathcal{S} \setminus \{\pm I_n\}, \delta(A) \geqslant \varepsilon\}$. Montrer que $\mathcal{S}_\varepsilon$ est compact.

(École normale supérieure)

▷ **Solution.**

1. On note $|||\ |||$ la norme sur $\mathcal{M}_n(\mathbb{R})$ subordonnée à la norme euclidienne de $\mathbb{R}^n$. Soit $A \in \mathcal{S}$ distincte de I_n. Alors -1 est une valeur propre de A, et si X est un vecteur propre unitaire associé, on a l'inégalité $2 = \|(A - I_n)X\| \leqslant |||A - I_n|||$. Cela prouve que I_n est un point isolé de $\mathcal{S}$.

On voit de même que $-I_n$ est aussi un point isolé de $\mathcal{S}$.

2. Si A est la matrice d'une symétrie par rapport à un hyperplan de $\mathbb{R}^n$, elle est semblable à la matrice $D = \mathrm{Diag}(-1, 1, \ldots, 1)$. La suite des matrices $D_k = D + \dfrac{1}{k}E_{12}$ pour $k \in \mathbb{N}^*$ est formée d'éléments de $\mathcal{S}$ et elle converge vers D. Donc D n'est pas isolée dans $\mathcal{S}$ et il est immédiat d'en déduire que A n'est pas isolée non plus.

Le même argument permet de montrer que $\pm I_n$ sont les seuls points isolés de $\mathcal{S}$.

3. Avec les notations de la question précédente on observe que les matrices $D + kE_{12}$ sont dans $\mathcal{S}$ pour tout entier naturel k. Il en découle que $\mathcal{S}$ n'est pas borné et il ne peut donc être compact. Notons toutefois que $\mathcal{S}$ est fermé en tant qu'image réciproque du singleton $\{I_n\}$ par la fonction continue $M \mapsto M^2$.

4. Comme $\mathcal{M}_n(\mathbb{R})$ est un espace de dimension finie, les parties compactes sont les parties fermées et bornées. Montrons tout d'abord que $\mathcal{S}_\varepsilon$ est borné pour la norme triple introduite en 1. Soit $A \in \mathcal{S}_\varepsilon$ et $X \in \mathbb{R}^n$. On note X_1 le projeté de X sur $\mathrm{Ker}(A - I_n)$ parallèlement à $\mathrm{Ker}(A + I_n)$ et $X_2 = X - X_1$ le projeté de X sur $\mathrm{Ker}(A + I_n)$ parallèlement à $\mathrm{Ker}(A - I_n)$. On a donc $AX = X_1 - X_2$ et, comme on travaille avec la norme euclidienne, $\|AX\|^2 = \|X_1\|^2 + \|X_2\|^2 - 2\langle X_1, X_2\rangle$. Essayons d'estimer $|||A|||$. Elle est forcément $\geqslant 1$ puisque A admet 1 pour valeur propre. Pour $\lambda \geqslant 1$ l'inégalité $\|AX\|^2 \leqslant \lambda\|X\|^2$ est équivalente à

$$-\langle X_1, X_1\rangle \leqslant \frac{\lambda-1}{2(\lambda+1)}(\|X_1\|^2+\|X_2\|^2).$$

Si X_1 ou X_2 est nul cela est vérifié. Si X_1 et X_2 sont non nuls, alors $\dfrac{X_1}{\|X_1\|}$ et $-\dfrac{X_2}{\|X_2\|}$ sont des vecteurs propres unitaires (respectivement pour 1 et pour -1) et on peut utiliser l'hypothèse $A \in \mathcal{S}_\varepsilon$. On a

$$\left\|\frac{X_1}{\|X_1\|}+\frac{X_2}{\|X_2\|}\right\|^2 \geqslant \varepsilon^2$$

ce qui donne en développant

$$(*) \qquad -\langle X_1, X_2\rangle \leqslant \left(1-\frac{\varepsilon^2}{2}\right)\|X_1\|\|X_2\|.$$

On a donc $-\langle X_1, X_2\rangle \leqslant \left(\frac{1}{2}-\frac{\varepsilon^2}{4}\right)(\|X_1\|^2+\|X_2\|^2)$ avec l'inégalité arithmético-géométrique. Un petit calcul montre que $\dfrac{\lambda-1}{2(\lambda+1)}$ est égal à $\dfrac{1}{2}-\dfrac{\varepsilon^2}{4}$ si, et seulement si, $\lambda = \dfrac{4-\varepsilon^2}{\varepsilon^2}$. Cela montre donc que $|||A||| \leqslant \dfrac{\sqrt{4-\varepsilon^2}}{\varepsilon}$ et donc que $\mathcal{S}_\varepsilon$ est inclus dans la boule fermée de centre l'origine et de rayon $\dfrac{\sqrt{4-\varepsilon^2}}{\varepsilon}$.

En fait dans l'inégalité (*) on peut mettre des valeurs absolues sur le produit scalaire (il suffit le cas échéant de remplacer X_1 par $-X_1$). Le coefficient $1-\dfrac{\varepsilon^2}{2}$ doit donc être positif. Autrement dit pour $\varepsilon > \sqrt{2}$ l'ensemble $\mathcal{S}_\varepsilon$ est tout simplement vide. Lorsque $\varepsilon = \sqrt{2}$ on a forcément $\mathrm{Ker}(A-I_n)$ *et* $\mathrm{Ker}(A+I_n)$ *qui sont orthogonaux et* $|||A||| = 1$ *: A est une symétrie orthogonale donc une isométrie. Enfin lorsque $\varepsilon < \sqrt{2}$ on vérifie qu'on a bien $\dfrac{\sqrt{4-\varepsilon^2}}{\varepsilon} \geqslant 1$.*

Il nous reste à montrer que $\mathcal{S}_\varepsilon$ est fermé. On supposera bien entendu $\varepsilon \leqslant \sqrt{2}$. L'ensemble $\mathcal{S}$ est fermé comme dit plus haut et comme $\pm I_n$ sont isolés l'ensemble $\mathcal{S} \setminus \{\pm I_n\}$ est aussi fermé. On va montrer que $\mathcal{S}_\varepsilon$ est exactement l'intersection de cet ensemble et de la boule fermée de rayon $\dfrac{\sqrt{4-\varepsilon^2}}{\varepsilon}$. Soit $A \in \mathcal{S} \setminus \{\pm I_n\}$ vérifiant $|||A||| \leqslant \dfrac{\sqrt{4-\varepsilon^2}}{\varepsilon}$. Pour tout $X \in \mathbb{R}^n$ on a donc $\|AX\|^2 \leqslant \dfrac{4-\varepsilon^2}{\varepsilon^2}\|X\|^2$. D'après les calculs précédents, on en déduit que pour tout vecteur propre X_1 (resp. X_2) associé à la valeur propre 1 (resp. -1) on a

$$\langle X_1, X_2\rangle \leqslant \left(\frac{1}{2}-\frac{\varepsilon^2}{4}\right)(\|X_1\|^2+\|X_2\|^2).$$

En particulier si on prend deux vecteurs propres unitaires, on obtient $\langle X_1, X_2\rangle \leqslant 1 - \dfrac{\varepsilon^2}{2}$ ce qui équivaut à $\|X_1 - X_2\|^2 \geqslant \varepsilon^2$. On a donc $\delta(A) \geqslant \varepsilon$ et $A \in \mathcal{S}_\varepsilon$. On peut conclure que $\mathcal{S}_\varepsilon$ est fermé et borné donc compact. ◁

La compacité de $\mathcal{S}_{\sqrt{2}}$ est évidente, puisqu'il s'agit de l'intersection du fermé $\mathcal{S} \setminus \{\pm I_n\}$ et du groupe orthogonal $O_n(\mathbb{R})$ qui est compact.

Le petit exercice suivant porte sur la compacité. Il généralise le résultat facile suivant : un sous-ensemble fini d'un groupe G stable par la loi du groupe est forcément un sous-groupe de G.

4.8. Partie compacte stable par produit

Montrer qu'une partie non vide et compacte de $GL_n(\mathbb{C})$, stable par produit, est un sous-groupe de $GL_n(\mathbb{C})$.

(**École normale supérieure**)

▷ **Solution.**

Soit G une partie de $GL_n(\mathbb{C})$ compacte, non vide et stable par le produit. Il nous faut montrer que G contient la matrice I_n et est stable par l'inverse. Soit $A \in G$. De la suite $(A^k)_{k\geqslant 1}$, qui est dans G, on peut extraire, par compacité de G, une sous-suite $(A^{\varphi(k)})_{k\geqslant 1}$ qui converge vers une matrice $B \in G$. La suite $A^{\varphi(k+1)-\varphi(k)}$ est formée d'éléments de G (car $\varphi(k+1) - \varphi(k) \geqslant 1$) et converge vers $BB^{-1} = I_n$ par continuité de l'inversion. Comme G est fermé on a donc $I_n \in G$. Quitte à prendre encore une sous-suite, on peut très bien supposer que $\varphi(k+1)-\varphi(k) \geqslant 2$ pour tout k et ainsi la suite $A^{\varphi(k+1)-\varphi(k)-1}$ est encore formée d'éléments de G et converge vers A^{-1}, qui est donc aussi dans G. On peut conclure que G est un sous-groupe de $GL_n(\mathbb{C})$. ◁

Les exercices qui suivent s'intéressent à la suite des puissances successives d'une matrice donnée. Il est naturel de penser que le rayon spectral de la matrice, c'est-à-dire le plus grand module des valeurs propres, joue un rôle important.

4.9. Rayon spectral

Pour $A \in \mathcal{M}_n(\mathbb{C})$, on note $\rho(A) = \max\limits_{\lambda\in \mathrm{Sp}(A)} |\lambda|$ le *rayon spectral* de A.

1. Montrer l'équivalence des conditions suivantes :

(i) il existe une norme $\|\ \|$ sur $\mathbb{C}^n$ telle que $\|\!|A|\!\| < 1$, où $\|\!|\ |\!\|$ désigne la norme triple subordonnée à $\|\ \|$;

(ii) $\rho(A) < 1$;

(iii) la suite $(A^k)_{k\geqslant 0}$ converge vers 0.

2. Montrer que $\sum A^k$ converge si, et seulement si, $\rho(A) < 1$.

3. Soit N une norme quelconque sur $\mathcal{M}_n(\mathbb{C})$. Montrer que

$$\lim_{k\to\infty} N(A^k)^{1/k} = \rho(A).$$

(**École normale supérieure**)

▷ **Solution.**

1. • Montrons pour commencer que (i) $\Longrightarrow$ (iii). Soit $\|\ \|$ une norme sur $\mathbb{C}^n$ telle que $\|\!|A|\!\| < 1$, où $|\!|\!|\ |\!|\!|$ désigne la norme triple subordonnée à $\|\ \|$. Comme la norme triple est sous-multiplicative on a $|\!|\!|A^k|\!|\!| \leqslant |\!|\!|A|\!|\!|^k$ pour tout $k \in \mathbb{N}^*$ et cela montre que la suite $(A^k)_{k\geqslant 0}$ converge vers 0, puisque la suite géométrique $|\!|\!|A|\!|\!|^k$ tend vers 0.

• Il est aussi facile de voir que (iii) $\Longrightarrow$ (ii). En effet soit $\lambda \in \operatorname{Sp} A$ et X un vecteur propre associé qu'on peut choisir unitaire. On a alors $A^k X = \lambda^k X$ pour tout k et comme la suite $(A^k X)$ tend vers 0, on en déduit que la suite (λ^k) converge vers 0. Cela impose $|\lambda| < 1$. Comme il n'y a qu'un nombre fini de valeurs propres, on a bien $\rho(A) < 1$.

• La dernière implication (ii) $\Longrightarrow$ (i) est la plus difficile. Il nous suffit de prouver que si $A \in \mathcal{M}_n(\mathbb{C})$ est fixée ainsi que $\varepsilon > 0$, on peut trouver une norme triple telle que $|\!|\!|A|\!|\!| \leqslant \rho(A)+\varepsilon$. Cela résulte des trois résultats suivants qui sont intéressants pour eux-mêmes.

(i) Soit $\|\ \|_1$ une norme sur $\mathbb{C}^n$ et $P \in \mathrm{GL}_n(\mathbb{C})$. Alors l'application $\|\ \|_2$ définie par $\|X\|_2 = \|PX\|_1$ est une norme sur $\mathbb{C}^n$ et pour $M \in \mathcal{M}_n(\mathbb{C})$ on a aisément $|\!|\!|M|\!|\!|_2 = |\!|\!|PMP^{-1}|\!|\!|_1$. Ce lemme permet de remplacer A par une matrice qui lui est semblable puisque le rayon spectral est aussi conservé par similitude.

(ii) On sait que A est semblable à une matrice triangulaire supérieure T. En fait, il est possible d'imposer que les coefficients non diagonaux de T soient inférieurs en module à n'importe quel réel $\eta > 0$ fixé. En effet, soit $p \in \mathbb{N}^*$ et $D = \operatorname{Diag}\left(\frac{1}{p}, \frac{1}{p^2}, \ldots, \frac{1}{p^n}\right)$. On a $D^{-1} = (p, p^2, \ldots, p^n)$. Regardons les coefficients de la matrice $D^{-1}TD$, où T est triangulaire supérieure. Le coefficient de la i-ième ligne et de la j-ième colonne est multiplié par p^{i-j}. Ainsi les coefficients diagonaux sont inchangés, mais tous ceux qui sont au-dessus de la diagonale sont au moins divisés par p. Il suffit de prendre l'entier p assez grand pour que ces coefficients aient un module majoré par $\eta > 0$ fixé.

Ce lemme est très utile et le lecteur le retrouvera notamment dans les exercices 4.16 et 4.17.

(*iii*) Le dernier point est le suivant : la norme infinie sur $\mathbb{C}^n$ définie par $\|X\|_\infty = \max_{1\leqslant i\leqslant n} |x_i|$ si $X = (x_1, \dots, x_n)$ induit sur $\mathcal{M}_n(\mathbb{C})$ la norme triple définie par $|||M|||_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |m_{ij}|$, où $M = (m_{ij})$.

Combinons alors ces trois résultats. Soit $A \in \mathcal{M}_n(\mathbb{C})$ et $\varepsilon > 0$. On commence par choisir $P \in \mathrm{GL}_n(\mathbb{C})$ telle que $T = P^{-1}AP$ soit triangulaire supérieure avec des coefficients non diagonaux inférieurs à $\dfrac{\varepsilon}{n}$. Bien entendu les coefficients diagonaux de T sont les valeurs propres de A. D'après le dernier point ci-dessus on a $|||T|||_\infty \leqslant \rho(A) + \varepsilon$. D'après le point (*i*) la norme définie par $\|X\| = \|P^{-1}X\|_\infty$ induit une norme triple pour laquelle on a $|||A||| = |||T|||_\infty$ et donc $|||A||| \leqslant \rho(A) + \varepsilon$. Si $\rho(A) < 1$ il est clairement possible d'avoir $|||A||| < 1$ en prenant ε assez petit.

2. Si la série $\sum A^k$ converge, alors la suite A^k tend vers 0 et on a donc $\rho(A) < 1$. Inversement, supposons $\rho(A) < 1$. Les valeurs propres de A sont de module strictement inférieur à 1. En particulier, 1 n'est pas valeur propre et la matrice $B = I_n - A$ est inversible. On a pour tout $p \geqslant 0$,

$$B(I + A + A^2 + \cdots + A^p) = I_n - A^{p+1}$$

et donc

$$I + A + A^2 + \cdots + A^p = B^{-1}(I - A^{p+1}).$$

Comme A^{p+1} tend vers 0 on a la convergence de la série $\sum A^k$ et on dispose même de sa somme : $\sum_{k=0}^{+\infty} A^k = (I_n - A)^{-1}$.

3. On commence par prouver le résultat pour une norme triple $|||\ |||$ induite par une norme $\|\ \|$ sur $\mathbb{C}^n$. Soit $A \in \mathcal{M}_n(\mathbb{C})$. Si $\lambda \in \mathrm{Sp}\, A$ et si X est un vecteur propre associé unitaire, on a $|\lambda| = \|AX\| \leqslant |||A|||$. Comme cela vaut pour toute valeur propre de A, on a donc $\rho(A) \leqslant |||A|||$. Il est clair, en trigonalisant A, que $\rho(A^k) = \rho(A)^k$ pour tout $k \in \mathbb{N}$. On a donc déjà pour tout k,

$$\rho(A) = \rho(A^k)^{1/k} \leqslant |||A^k|||^{1/k}.$$

Soit maintenant $\varepsilon > 0$ et $A' = \dfrac{A}{\rho(A) + \varepsilon}$. On a $\rho(A') < 1$ de sorte que la suite (A'^k) converge vers 0 d'après la question 1 ci-dessus. Il existe donc un rang N tel que, pour tout $k \geqslant N$,

$$\left|\!\left|\!\left| \frac{A^k}{(\rho(A) + \varepsilon)^k} \right|\!\right|\!\right| \leqslant 1$$

On a donc, pour tout $k \geqslant N$, $\rho(A) \leqslant |||A^k|||^{1/k} \leqslant \rho(A) + \varepsilon$. Cela prouve bien que $\rho(A) = \lim_{k\to+\infty} |||A^k|||^{1/k}$.

Si N est une autre norme quelconque elle est équivalente à $\|\ \|$ et si a, b sont des réels strictement positifs tels que $a\|\ \| \leqslant \mathrm{N} \leqslant b\|\ \|$ on a pour tout k,

$$a^{1/k}\|\mathrm{A}^k\|^{1/k} \leqslant \mathrm{N}(\mathrm{A}^k)^{1/k} \leqslant b^{1/k}\|\mathrm{A}^k\|^{1/k}.$$

Le théorème d'encadrement permet de conclure que $\mathrm{N}(\mathrm{A}^k)^{1/k}$ converge aussi vers $\rho(\mathrm{A})$. ◁

La série géométrique est utilisée dans l'exercice suivant qui offre une jolie preuve analytique du théorème de Cayley-Hamilton.

4.10. Une preuve du théorème de Cayley-Hamilton

1. Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{C})$ et $k \in \mathbb{N}^*$. Montrer qu'il existe $\mathrm{R} \geqslant 0$ tel que pour $r \geqslant \mathrm{R}$,

$$\mathrm{A}^{k-1} = \frac{1}{2\pi}\int_0^{2\pi} (re^{it})^k (re^{it}\mathrm{I}_n - \mathrm{A})^{-1}\mathrm{d}t.$$

2. En déduire le théorème de Cayley-Hamilton.

(École normale supérieure)

▷ **Solution.**

1. On munit $\mathcal{M}_n(\mathbb{C})$ d'une norme triple induite par une norme quelconque de $\mathbb{C}^n$. Pour $r > 0$ on peut écrire $re^{it}\mathrm{I}_n - \mathrm{A} = re^{it}(\mathrm{I}_n - r^{-1}e^{-it}\mathrm{A})$. Pour $r > \|\mathrm{A}\|$ la matrice $r^{-1}e^{-it}\mathrm{A}$ a une norme triple strictement inférieure à 1 et la matrice $\mathrm{I}_n - r^{-1}e^{-it}\mathrm{A}$ est inversible, son inverse étant la somme d'une série géométrique (voir l'exercice précédent) :

$$(re^{it}\mathrm{I}_n - \mathrm{A})^{-1} = r^{-1}e^{-it}(\mathrm{I}_n - r^{-1}e^{-it}\mathrm{A})^{-1} = \sum_{p=1}^{+\infty} r^{-p}e^{-ipt}\mathrm{A}^{p-1}.$$

En mettant cette expression dans l'intégrale il vient

$$\int_0^{2\pi} (re^{it})^k (re^{it}\mathrm{I}_n - \mathrm{A})^{-1}\mathrm{d}t = \int_0^{2\pi} \sum_{p=1}^{+\infty} r^{k-p}e^{-i(k-p)t}\mathrm{A}^{p-1}\mathrm{d}t.$$

Comme la convergence de la série est normale en t, il est possible d'intervertir l'intégrale et la sommation en série. On obtient donc

$$\int_0^{2\pi} (re^{it})^k (re^{it}\mathrm{I}_n - \mathrm{A})^{-1}\mathrm{d}t = \sum_{p=1}^{+\infty} r^{k-p}\int_0^{2\pi} e^{-i(k-p)t}\mathrm{d}t\ \mathrm{A}^{p-1} = 2\pi\mathrm{A}^{k-1},$$

car seul le terme d'indice k donne une intégrale non nulle. En divisant par 2π on a la formule demandée.

2. Posons $\chi_A = a_0 + a_1X + \cdots + a_nX^n$. En multipliant la formule précédente par a_{k-1} et en sommant les égalités obtenues pour $k = 1, 2, \ldots, n+1$ on obtient, toujours pour r assez grand,

$$\chi_A(A) = \frac{1}{2\pi} \int_0^{2\pi} \chi_A(re^{it}) re^{it} (re^{it} I_n - A)^{-1} dt.$$

D'après la formule fondamentale vérifiée par la comatrice on a

$${}^t\mathrm{Com}(XI_n - A)(XI_n - A) = \chi_A(X) I_n.$$

En évaluant cela en re^{it}, on obtient, pour r assez grand,

$$\chi_A(re^{it})(re^{it} I_n - A)^{-1} = {}^t\mathrm{Com}(re^{it} I_n - A).$$

L'intégrale ci-dessus est donc nulle car les coefficients de la matrice sous le signe intégral sont des polynômes trigonométriques de valeur moyenne nulle : $re^{it}\,({}^t\mathrm{Com}(re^{it} I_n - A)) = e^{it} A_1(r) + e^{2it} A_2(r) + \cdots + e^{int} A_n(r)$. On a donc $\chi_A(A) = 0$ ce qui prouve le théorème de Cayley-Hamilton. ◁

L'exercice suivant complète l'exercice 4.9 en étudiant les matrices A *dont la suite des puissances est bornée.*

4.11. Suite des puissances bornée

Déterminer l'ensemble des matrices $A \in \mathcal{M}_n(\mathbb{C})$ telles que la suite $(A^k)_{k \geqslant 0}$ soit bornée.

(École normale supérieure)

▷ **Solution.**

On note $\rho(A) = \max_{\lambda \in \mathrm{Sp}\, A} |\lambda|$ le rayon spectral de A. Lorsque $n = 1$ on est ramené à une simple suite géométrique de $\mathbb{C}$ et il est bien connu que, pour $a \in \mathbb{C}$, la suite a^k est bornée si, et seulement si, $|a| \leqslant 1$. La position du rayon spectral par rapport à 1 joue donc un rôle important.

• Si $\rho(A) < 1$, la suite $(A^k)_{k \geqslant 0}$ converge vers 0 (voir l'exercice 4.9) donc elle est bornée.

• Supposons $\rho(A) > 1$. Alors A possède une valeur propre $\lambda \in \mathbb{C}$ de module > 1. Si X est un vecteur propre associé, on a

$$\|A^k X\| = |\lambda|^k \|X\| \xrightarrow[k \to +\infty]{} +\infty,$$

et la suite $(A^k)_{k \geqslant 0}$ ne saurait être bornée.

• Il reste à étudier le cas où $\rho(A) = 1$. Dans ce cas la suite $(A^k)_{k \geqslant 0}$ n'est pas forcément bornée : elle l'est si $A = I_n$, mais ne l'est pas pour

$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ par exemple, car $A^k = \begin{pmatrix} 1 & k \\ 0 & 1 \end{pmatrix}$ pour tout $k \in \mathbb{N}$. Commençons à nous ramener au cas où il n'y a qu'une seule valeur propre en utilisant la décomposition en sous-espaces caractéristiques. L'espace $\mathbb{C}^n$ est somme directe des sous-espaces caractéristiques $F_1, \dots, F_r$ respectivement relatifs aux valeurs propres deux à deux distinctes $\lambda_1, \dots, \lambda_r$. Si $\mathcal{B}$ est une base obtenue par recollement de bases de $F_1, \dots, F_r$, la matrice de l'endomorphisme canoniquement associé à A est, dans cette nouvelle base, diagonale par blocs de la forme

$$\begin{pmatrix} \boxed{A_1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \boxed{A_r} \end{pmatrix}.$$

Les puissances de A sont bornées si, et seulement si, toutes les suites $(A_i^k)_{k\geqslant 0}$ sont bornées (pour $1 \leqslant i \leqslant r$). Comme nous l'avons déjà dit, lorsque $|\lambda_i| < 1$ la suite $(A_i^k)_{k\geqslant 0}$ est bornée. Regardons ce qui se passe lorsque $|\lambda_i| = 1$. En remarquant que $(A_i^k)_{k\geqslant 0}$ est bornée si, et seulement si, la suite $\left((\overline{\lambda_i} A_i)^k\right)_{k\geqslant 0}$ est bornée, on peut supposer $\lambda_i = 1$. Autrement dit, A_i est une matrice unipotente que l'on peut écrire $A_i = I + N$, avec I matrice identité et N nilpotente. Soit p le plus petit entier naturel tel que $N^p = 0$. Nous savons alors que $(I, N, \dots, N^{p-1})$ est une famille libre que nous pouvons compléter en une base $\mathcal{M}$ de l'espace des matrices complexes de même taille que A_i. Pour $k \geqslant p$, la formule du binôme donne

$$A_i^k = (I + N)^k = I + kN + C_k^2 N^2 + \dots + C_k^{p-1} N^{p-1}.$$

On obtient alors les coordonnées de A_i^k dans la base $\mathcal{M}$ en lisant les coefficients de ce polynôme en N. Si $p \geqslant 2$, le coefficient de N est k, qui n'est pas borné. Pour que $(A_i^k)_{k\geqslant 0}$ soit bornée, il est nécessaire que p soit égal à 1, c'est-à-dire que A_i soit diagonale, ce qui signifie encore que le sous-espace propre relatif à λ_i est égal au sous-espace caractéristique : $\mathrm{Ker}(A - \lambda_i I) = F_i$. Réciproquement, si A_i est diagonale, ses puissances sont clairement bornées.

Conclusion. La suite des puissances de la matrice A est bornée si, et seulement si, $\rho(A) \leqslant 1$ et pour toute valeur propre complexe de module 1, on a $\mathrm{Ker}(A - \lambda I) = F_\lambda$, où F_λ désigne le sous-espace caractéristique relatif à λ. ◁

Dans l'exercice suivant on obtient une majoration du rayon spectral à partir d'une information sur la suite des traces des puissances successives de la matrice.

4.12. Suite des traces des puissances itérées

Soit $A \in \mathcal{M}_n(\mathbb{C})$. On suppose que $\lim_{k\to\infty} \operatorname{Tr} A^k = 0$. Montrer que les valeurs propres de A sont de module < 1.

(École polytechnique)

▷ **Solution.**

Notons $\lambda_1, \ldots, \lambda_p$ les valeurs propres deux à deux distinctes de A et $n_1, \ldots, n_p$ les ordres de multiplicités respectifs. En trigonalisant A, il est clair que $\operatorname{Tr} A^k = n_1\lambda_1^k + \cdots + n_k\lambda_p^k$ pour tout k. Par hypothèse cette suite tend vers 0 lorsque $k \to +\infty$ et on doit en déduire que nécessairement $|\lambda_i| < 1$ pour tout i. Le résultat est évident lorsque $p = 1$ ce qui incite à faire une démonstration par récurrence sur p. On montre plus généralement la propriété (H_p) suivante : si $\lambda_1, \ldots, \lambda_p$ sont des nombres complexes deux à deux distincts et $\alpha_1, \ldots, \alpha_p$ des nombres complexes non nuls, alors la convergence vers 0 de la suite de terme général $u_k = \alpha_1\lambda_1^k + \cdots + \alpha_p\lambda_p^k$ implique $|\lambda_i| < 1$ pour tout i.

Cela est toujours évident pour $p = 1$. Supposons le résultat vrai au rang $p - 1$. Pour se ramener à ce cas il nous suffit de faire une petite combinaison de deux termes consécutifs de la suite (u_k). Posons

$$v_k = u_{k+1} - \lambda_p u_k = \alpha_1(\lambda_1 - \lambda_p)\lambda_1^k + \cdots + \alpha_{p-1}(\lambda_{p-1} - \lambda_p)\lambda_{p-1}^k$$

pour tout k. La suite (v_k) converge toujours vers 0. En appliquant l'hypothèse de récurrence, on en déduit que $|\lambda_i| < 1$ pour $1 \leqslant i \leqslant p-1$ (car $\lambda_1, \ldots, \lambda_p$ sont deux à deux distincts). On a alors $\alpha_p\lambda_p^k$ qui tend vers 0 et donc on a aussi $|\lambda_p| < 1$, ce qui termine la récurrence. ◁

Les prochains exercices ont pour thème les classes de similitude. Le premier exercice, très classique, montre qu'il n'existe pas de norme sur $\mathcal{M}_n(\mathbb{R})$ qui soit constante sur toutes les classes de similitudes et s'intéresse ensuite aux semi-normes qui ont cette propriété. Rappelons qu'une semi-norme N *sur un espace* E *vérifie tous les axiomes d'une norme, sauf l'implication* $N(x) = 0 \Longrightarrow x = 0$.

4.13. Semi-normes invariantes par similitude

1. Soit $n \geqslant 2$. Montrer qu'il n'existe pas sur $\mathcal{M}_n(\mathbb{R})$ de norme $\|\ \|$ invariante par similitude, *i.e.* telle que :

$$\forall (A, P) \in \mathcal{M}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{R}), \quad \|P^{-1}AP\| = \|A\|.$$

2. Montrer que $A \mapsto |\operatorname{Tr} A|$ est une semi-norme sur $\mathcal{M}_n(\mathbb{R})$ qui est invariante par similitude.

3. Déterminer toutes les semi-normes sur $\mathcal{M}_n(\mathbb{R})$ invariantes par similitude.

(**École normale supérieure**)

▷ **Solution.**

1. Notons que dans le cas $n = 1$ toutes les normes sont invariantes par similitude. Supposons l'existence d'une telle norme $\|\ \|$ sur $\mathcal{M}_n(\mathbb{R})$ avec $n \geqslant 2$. Par hypothèse on a $\|AP\| = \|PA\|$ pour tout couple (A, P) de $\mathcal{M}_n(\mathbb{R}) \times \mathrm{GL}_n(\mathbb{R})$. Par densité de $\mathrm{GL}_n(\mathbb{R})$ (voir l'exercice 4.1) et continuité de la norme on a $\|AB\| = \|BA\|$ pour tout couple (A, B) de $\mathcal{M}_n(\mathbb{R})^2$. Or, si $n \geqslant 2$, il est aisé d'exhiber A et B telles que $AB = 0$ et $BA \neq 0$, par exemple en prenant deux matrices de la base canonique. D'où la contradiction.

2. Il est clair que pour A, B dans $\mathcal{M}_n(\mathbb{R})$ et $\lambda \in \mathbb{R}$ on a $|\operatorname{Tr}(A)| \geqslant 0$, $|\operatorname{Tr}(\lambda A)| = |\lambda||\operatorname{Tr}(A)|$ et $|\operatorname{Tr}(A + B)| \leqslant |\operatorname{Tr}(A)| + |\operatorname{Tr}(B)|$. On a donc bien une semi-norme. De plus, si $P \in \mathrm{GL}_n(\mathbb{R})$ on a par propriété de la trace $\operatorname{Tr}(P^{-1}AP) = \operatorname{Tr}(APP^{-1}) = \operatorname{Tr} A$, donc cette semi-norme est invariante par similitude.

3. Soit N une semi-norme invariante par similitude. Notons que N est continue : en effet, si $\|\ \|$ est une norme quelconque sur $\mathcal{M}_n(\mathbb{R})$ il en est de même de $N + \|\ \|$, donc N s'écrit comme la différence de deux normes et est continue. En particulier on en déduit comme dans la question 1 que $N(AB) = N(BA)$ pour tout couple $(A, B) \in \mathcal{M}_n(\mathbb{R})^2$.

Considérons $F = \{A \in \mathcal{M}_n(\mathbb{R}),\ N(A) = 0\}$. Il résulte de l'inégalité triangulaire que F est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$. Observons que si $A \in \mathcal{M}_n(\mathbb{R})$ et $B \in F$, on a $N(A + B) \leqslant N(A) + N(B) = N(A)$ et $N(A) = N(A + B - B) \leqslant N(A + B) + N(-B) = N(A + B)$ de sorte que $N(A + B) = N(A)$. On va montrer que F contient l'hyperplan des matrices de trace nulle. Soit $i \neq j$ deux entiers de $[\![1, n]\!]$. On a

$$N(E_{ij}) = N(E_{ij}E_{jj}) = N(E_{jj}E_{ij}) = N(0) = 0$$

donc $E_{ij} \in F$. On en déduit que N contient toutes les matrices de diagonale nulle. Or il est classique que toute matrice de trace nulle est semblable à une matrice de diagonale nulle (voir l'exercice 7.12 du tome algèbre 1). Donc F contient l'hyperplan des matrices de trace nulle. Si $A \in \mathcal{M}_n(\mathbb{R})$, alors $A - \dfrac{\operatorname{Tr} A}{n} I_n \in F$ donc

$$N(A) = N\left(\frac{\operatorname{Tr} A}{n} I_n\right) = \frac{N(I_n)}{n} |\operatorname{Tr}(A)|$$

et N est positivement colinéaire à la semi-norme de la question précédente.

Conclusion. Les semi-normes invariantes par similitude sont les applications $A \mapsto \lambda |\operatorname{Tr} A|$ avec $\lambda \in \mathbb{R}_+$. ◁

L'exercice suivant traduit la condition $\operatorname{rg} A = \operatorname{rg} A^2$ *en termes de limite.*

4.14. Condition pour que rg A = rg A²

Soit $A \in \mathcal{M}_n(\mathbb{R})$. Montrer l'équivalence entre les propositions suivantes :

(*i*) $\operatorname{rg} A = \operatorname{rg} A^2$;

(*ii*) $a(A + aI_n)^{-1}$ admet une limite finie quand a tend vers 0.

(**École normale supérieure**)

▷ **Solution.**

On remarque que $(A + aI_n)^{-1}$ est définie si $0 < |a| < |\lambda|$, où λ est la valeur propre réelle de A de plus petite valeur absolue (si $\operatorname{Sp} A = \emptyset$, on peut prendre a quelconque).

Supposons que (*i*) est vérifié. Il est classique qu'on a alors

$$\mathbb{R}^n = \operatorname{Ker} A \dotplus \operatorname{Im} A.$$

En effet, la restriction de A à Im A a pour image Im A tout entier, c'est donc un isomorphisme de Im A et son noyau $\operatorname{Ker} A \cap \operatorname{Im} A$ est réduit à 0. On conclut grâce au théorème du rang. En se plaçant dans une base de $\mathbb{R}^n$ constituée d'une base de Ker A et d'une base de Im A, on voit qu'il existe $P \in \mathrm{GL}_n(\mathbb{R})$ et $B \in \mathrm{GL}_r(\mathbb{R})$, où r est le rang de A, telles que $A = P\begin{pmatrix} 0 & 0 \\ 0 & B \end{pmatrix}P^{-1}$. On a donc

$$A + aI_n = P\begin{pmatrix} aI_{n-r} & 0 \\ 0 & B + aI_r \end{pmatrix}P^{-1}$$

et pour a assez petit,

$$a(A + aI_n)^{-1} = P\begin{pmatrix} I_{n-r} & 0 \\ 0 & a(B + aI_r)^{-1} \end{pmatrix}P^{-1}.$$

La matrice B est inversible donc $\lim\limits_{a \to 0} a(B + aI_r)^{-1} = 0B^{-1} = 0$, par continuité de l'inverse, et $\lim\limits_{a \to 0} a(A + aI_r)^{-1} = P\begin{pmatrix} I_{n-r} & 0 \\ 0 & 0 \end{pmatrix}P^{-1}$. Ainsi (*ii*) est vérifié.

Pour la réciproque, on démontre la contraposée. Si $\operatorname{rg} A \neq \operatorname{rg} A^2$, on a, par le théorème du rang, $\operatorname{Ker} A \neq \operatorname{Ker} A^2$. Soit $X \in \operatorname{Ker} A^2 \setminus \operatorname{Ker} A$.

On a alors, pour a assez petit et non nul,

$$(A + aI_n)(A - aI_n)X = (A^2 - a^2 I_n)X = -a^2 X$$

et donc

$$a(A + aI_n)^{-1}X = -\frac{1}{a}(AX - aX).$$

Quand a tend vers 0, $AX - aX$ tend vers $AX \neq 0$ et $a(A + aI_n)^{-1}X$ n'a pas une limite finie. Il en est de même *a fortiori* de $a(A + aI_n)^{-1}$. ◁

Il est remarquable de pouvoir lire certaines propriétés algébriques d'une matrice complexe par des caractéristiques topologiques de sa classe de similitude. L'exercice suivant s'intéresse aux matrices dont la classe de similitude est bornée et le suivant concernera les matrices dont la classe de similitude est fermée.

4.15. Classes de similitude bornées

Déterminer les matrices $A \in \mathcal{M}_n(\mathbb{C})$ dont la classe de similitude est bornée.

(École polytechnique)

▷ **Solution.**

Il est clair que si la matrice A est scalaire, sa classe de similitude ne contient qu'elle-même, donc est bornée.

Montrons que les matrices scalaires sont les seules matrices ayant cette propriété. Toutes les normes sur $\mathcal{M}_n(\mathbb{C})$ étant équivalentes, on va travailler avec la norme infinie. Soit $A \in \mathcal{M}_n(\mathbb{C})$ et f l'endomorphisme de $\mathbb{C}^n$ canoniquement associé à A. Si A n'est pas scalaire, f n'est pas une homothétie vectorielle. C'est un résultat classique qu'il existe alors un vecteur $u \in \mathbb{C}^n$ tel que la famille $(u, f(u))$ soit libre. Soit $\lambda \in \mathbb{C}^*$ et $\mathcal{B} = (e_1, \ldots, e_n)$ une base de $\mathbb{C}^n$ telle que $e_1 = \lambda u$, $e_2 = f(u)$. On a alors $f(e_1) = \lambda e_2$. La matrice B de f dans la base $\mathcal{B}$ est semblable à A. Sa première colonne est $\begin{pmatrix} 0 \\ \lambda \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. On a donc $\|B\|_\infty \geqslant |\lambda|$. Comme λ est quelconque, la classe de similitude de A n'est pas bornée. ◁

L'énoncé suivant montre que les matrices diagonalisables complexes se caractérisent par le fait que leur classe de similitude est fermée.

4.16. Classe de similitude d'une matrice diagonalisable

Si $A \in \mathcal{M}_n(\mathbb{C})$, on note $S(A)$ la classe de similitude de A.

1. Montrer que si A est inversible, l'adhérence de $S(A)$ est incluse dans $GL_n(\mathbb{C})$.

2. Montrer que A est diagonalisable si, et seulement si, $S(A)$ est fermée.

(**École normale supérieure**)

▷ **Solution.**

1. Soit $A \in GL_n(\mathbb{C})$ et $(A_p)_{p\geqslant 0}$ une suite d'éléments de $S(A)$ qui converge vers une matrice B. On a alors $\det A_p = \det A$ pour tout $p \in \mathbb{N}$. Le déterminant est continu donc on a $\det B = \lim\limits_{p\to+\infty} \det A_p = \det A \neq 0$ et $B \in GL_n(\mathbb{C})$.

2. On suppose que A est diagonalisable. On se donne encore $(A_p)_{p\geqslant 0}$ une suite de $S(A)$ qui converge vers une matrice B. Comme A_p est semblable à A, on a $\chi_{A_p} = \chi_A$. Les coefficients du polynôme caractéristique sont des fonctions continues. On en déduit que $\chi_B = \lim\limits_{p\to+\infty} \chi_{A_p} = \chi_A$. Le polynôme minimal μ_A de A est à racines simples car A est diagonalisable. Puisque A_p est semblable à A, on a $\mu_A(A_p) = 0$ pour tout p. Par passage à la limite, on en déduit que $\mu_A(B) = 0$. La matrice B possède un polynôme annulateur à racines simples donc est diagonalisable. Les matrices A et B sont diagonalisables et ont même polynôme caractéristique donc mêmes valeurs propres comptées avec multiplicité ; elles sont semblables. Cela montre que $B \in S(A)$ et donc que $S(A)$ est fermé.

Montrons la réciproque. Soit $A = (a_{ij}) \in \mathcal{M}_n(\mathbb{C})$ telle que $S(A)$ soit fermée. Quitte à remplacer A par une matrice semblable, on peut supposer que A est triangulaire supérieure. Notons $\mathcal{B} = (e_1, \dots e_n)$ la base canonique de $\mathbb{C}^n$. Pour $p \in \mathbb{N}^*$, écrivons la matrice A_p de l'endomorphisme u canoniquement associé à A dans la base $\left(\frac{1}{p}e_1, \frac{1}{p^2}e_2, \dots, \frac{1}{p^n}e_n\right)$. On a, pour tout $j \in [\![1, n]\!]$,

$$u\left(\frac{1}{p^j}e_j\right) = \frac{1}{p^j}u(e_j) = \frac{1}{p^j}\sum_{i=1}^{j} a_{i,j}e_i = \sum_{i=1}^{j} p^{i-j}a_{i,j}\frac{1}{p^i}e_i.$$

Le coefficient (i, j) de A_p est donc égal à $p^{i-j}a_{ij}$ si $i \leqslant j$ et est nul sinon. Autrement dit, A_p est aussi triangulaire supérieure et a la même diagonale que A, mais tous les coefficients strictement au-dessus de la diagonale sont divisés par une puissance de p. Il en découle que A_p converge vers la matrice diagonale $D = \mathrm{Diag}(a_{1,1}, \dots, a_{n,n})$ lorsque p tend vers $+\infty$. Or, A_p est semblable à A pour tout p. Comme $S(A)$ est fermé, D est semblable à A, qui est donc diagonalisable. ◁

La première implication reste vraie dans $\mathcal{M}_n(\mathbb{R})$: la classe de similitude d'une matrice diagonalisable est fermée (la preuve est la même). En revanche la réciproque devient fausse. En effet, $A \in \mathcal{M}_n(\mathbb{R})$ on a

$$\{P^{-1}AP, P \in GL_n(\mathbb{C})\} \cap \mathcal{M}_n(\mathbb{R}) = \{P^{-1}AP, P \in GL_n(\mathbb{R})\},$$

car deux matrices réelles semblables sur $\mathbb{C}$ sont semblables sur $\mathbb{R}$ (voir l'exercice 2.71). Autrement dit, la classe de similitude réelle de A est l'intersection de sa classe de similitude complexe avec $\mathcal{M}_n(\mathbb{R})$. Ainsi, si A est diagonalisable sur $\mathbb{C}$ et pas sur $\mathbb{R}$, comme par exemple $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, la classe de similitude réelle de A est fermée.

Notons en passant qu'une classe de similitude ne peut pas être ouverte. Voici un argument simple : la trace étant constante sur une classe de similitude, celle-ci est incluse dans un hyperplan affine de $\mathcal{M}_n(\mathbb{C})$.

Dans l'exercice ci-après, on montre que les matrices nilpotentes sont les matrices dont la classe de similitude est adhérente à la matrice nulle. La démarche est analogue.

4.17. Classe de similitude d'une matrice nilpotente

Soit $n \in \mathbb{N}^*$ et $A \in \mathcal{M}_n(\mathbb{C})$. Montrer que A est nilpotente si, et seulement si, il existe une suite de matrices semblables à A qui converge vers la matrice nulle.

(École polytechnique)

▷ **Solution.**

Supposons qu'une suite $(A_p)_{p\in\mathbb{N}}$, composée de matrices semblables à A, converge vers 0. Alors, comme le polynôme caractéristique de M est une fonction continue de la matrice M, χ_{A_p} converge dans $\mathbb{C}_n[X]$ vers $\chi_0 = X^n$. Or la suite χ_{A_p} est constante et égale à χ_A. On en déduit que $\chi_A = X^n$ et $A^n = 0$, en vertu du théorème de Cayley-Hamilton : la matrice A est nilpotente.

Réciproquement, supposons que A est nilpotente. Il existe une base $(e_1, \dots e_n)$ de $\mathbb{C}^n$ telle que la matrice $A_0 = (a_{ij})_{1\leqslant i,j\leqslant n}$ dans cette base de l'endomorphisme canoniquement associé à A soit triangulaire supérieure à diagonale nulle. Comme dans l'exercice 4.16, on considère, pour $p \geqslant 1$, la base $\mathcal{B}_p = \left(\frac{1}{p}e_1, \frac{1}{p^2}e_2, \dots, \frac{1}{p^n}e_n\right)$. On a

$$A(e_k) = a_{k-1,k}e_{k-1} + a_{k-2,k}e_{k-2} + \cdots + a_{1,k}e_1,$$

et donc

$$A\left(\frac{e_k}{p^k}\right) = \frac{a_{k-1,k}}{p}\frac{e_{k-1}}{p^{k-1}} + \frac{a_{k-2,k}}{p^2}\frac{e_{k-2}}{p^{k-2}} + \cdots + \frac{a_{1,k}}{p^{k-1}}\frac{e_1}{p}.$$

Notons A_p la matrice de l'endomorphisme canoniquement associé à A dans la base $\mathcal{B}_p$. C'est une matrice semblable à A. La suite de matrices (A_p) converge bien vers 0, car chaque coefficient tend vers 0 quand p tend vers l'infini. ◁

Nous changeons maintenant de thème et abordons une série d'exercices consacrés à des questions de connexité. Le lecteur aura déjà rencontré quelques résultats : connexité par arcs du sous-groupe $\mathrm{SL}_n(\mathrm{K})$ *dans 3.1, connexité par arcs du groupe* $\mathrm{SO}_n(\mathbb{R})$ *dans l'exercice 1.35 du tome algèbre 3. L'énoncé suivant s'intéresse au groupe linéaire.*

4.18. Groupe linéaire

1. Est-ce que $\mathrm{GL}_n(\mathbb{R})$ est connexe par arcs ?

2. Montrer que le groupe $\mathrm{GL}_n^+(\mathbb{R}) = \{A \in \mathcal{M}_n(\mathbb{R}),\ \det A > 0\}$ est connexe par arcs.

3. Soit $X \subset \mathcal{M}_n(\mathbb{R})$ un ensemble contenant strictement $\mathrm{GL}_n(\mathbb{R})$. Montrer que X est connexe par arcs.

(École normale supérieure)

▷ **Solution.**

1. L'application det est continue et l'image de $\mathrm{GL}_n(\mathbb{R})$ est l'ensemble $\mathbb{R}^*$ qui n'est pas connexe par arcs. On en déduit que $\mathrm{GL}_n(\mathbb{R})$ n'est pas connexe par arcs.

2. Le résultat peut s'obtenir de diverses manières en fonction des connaissances dont on dispose :

• On peut utiliser la décomposition polaire (voir l'exercice 2.27 du tome algèbre 3) et la connexité par arcs de $\mathrm{SO}_n(\mathbb{R})$ (voir exercice 1.35 du même tome). Il est toutefois peu probable que ce soit la solution attendue de la part d'un candidat qui peut très bien ignorer ces résultats.

• On peut adapter la démonstration de la connexité par arcs de $\mathrm{SL}_n(\mathbb{R})$ donnée dans l'exercice 3.1 : on écrivant une matrice $A \in \mathrm{GL}_n^+(\mathbb{R})$ sous la forme $T_1 \dots T_p D T'_1 \dots T'_q$ où les matrices T_i et T'_j sont des matrices de transvection et où D est la matrice diagonale $(1, 1, \dots, 1, \det A)$ il est facile de construire un chemin continu joignant A à l'identité inclus dans $\mathrm{GL}_n^+(\mathbb{R})$.

• Pour changer, nous allons donner une autre preuve, par récurrence sur n, preuve que pourrait trouver un candidat ignorant les résultats précédents le jour de l'oral. Pour $n = 1$, le groupe $\mathrm{GL}_1^+(\mathbb{R})$ est homéomorphe à $\mathbb{R}_+^*$ et il est clairement connexe par arcs. Supposons le résultat vrai au rang $n-1$. Soit $A \in \mathrm{GL}_n^+(\mathbb{R})$. On va construire un chemin

continu entre A et la matrice I_n, ce qui suffit clairement pour conclure. Notons $(C_1, \dots, C_n)$ les colonnes de A et $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$. Il y a au moins un des vecteurs e_i qui n'est pas dans l'hyperplan $\mathrm{Vect}(C_2, \dots, C_n)$ et, soit e_i, soit $-e_i$, est dans le même demi-espace ouvert délimité par cet hyperplan que le vecteur C_1. Comme un demi-espace ouvert est connexe par arcs on peut trouver $\varphi : [0,1] \to \mathbb{R}^n$ continue telle que $\varphi(0) = C_1$, $\varphi(1) = \pm e_i$ et $(\varphi(t), C_2, \dots, C_n)$ inversible pour tout t. On a un chemin continu de matrices inversibles entre A et une matrice inversible B dont la première colonne contient $\pm e_i$. Comme le déterminant garde un signe constant sur ce chemin, celui-ci est inclus dans $\mathrm{GL}_n^+(\mathbb{R})$.

Notons C la matrice obtenue à partir de B en échangeant les lignes L_1 et L_i et en multipliant une de ces lignes par -1, de manière à obtenir un coefficient $(1,1)$ égal à 1. Il est encore possible de passer de B à C par un chemin continu inclus dans $\mathrm{GL}_n^+(\mathbb{R})$. Cela provient de l'existence d'un chemin continu entre $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ et $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ (resp. $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$) : il suffit de prendre $t \mapsto \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$ avec $t \in \left[0, \frac{\pi}{2}\right]$ (resp. $t \in \left[-\frac{\pi}{2}, 0\right]$). On en déduit un chemin φ entre I_n et la matrice $I_n - E_{11} - E_{ii} - E_{1i} + E_{i1}$ (resp. $I_n - E_{11} - E_{ii} + E_{1i} - E_{i1}$). Alors $t \longmapsto \varphi(t)\mathrm{B}$ est un chemin inclus dans $\mathrm{GL}_n^+(\mathbb{R})$ de B à C.

On vient donc de relier continûment A à une matrice de la forme $\mathrm{C} = \begin{pmatrix} 1 & \mathrm{L} \\ 0 & \mathrm{A}' \end{pmatrix}$ où $\mathrm{A}' \in \mathrm{GL}_{n-1}^+(\mathbb{R})$. Par hypothèse de récurrence, on peut relier continûment A' à I_{n-1} avec un chemin inclus dans $\mathrm{GL}_{n-1}^+(\mathbb{R})$. Il est donc possible de relier A à la matrice $\begin{pmatrix} 1 & \mathrm{L} \\ 0 & I_{n-1} \end{pmatrix}$ puis celle-ci à la matrice I_n avec le chemin continu : $t \in [0,1] \mapsto \begin{pmatrix} 1 & t\mathrm{L} \\ 0 & I_{n-1} \end{pmatrix}$. D'où le résultat.

3. L'ensemble $\mathrm{GL}_n^-(\mathbb{R}) = \{\mathrm{A} \in \mathcal{M}_n(\mathbb{R}), \det \mathrm{A} < 0\}$ est aussi connexe par arcs puisque si M est une matrice fixée de cet ensemble l'application $\mathrm{A} \mapsto \mathrm{MA}$ réalise un homéomorphisme entre $\mathrm{GL}_n^+(\mathbb{R})$ et $\mathrm{GL}_n^-(\mathbb{R})$. Soit $\mathrm{X} \subset \mathcal{M}_n(\mathbb{R})$ un ensemble contenant strictement $\mathrm{GL}_n(\mathbb{R})$. Pour montrer que X est connexe par arcs il suffit de prouver que toute matrice non inversible A de X peut être reliée par un chemin continu inclus dans X à une matrice de $\mathrm{GL}_n^+(\mathbb{R})$ et à une matrice de $\mathrm{GL}_n^-(\mathbb{R})$. Soit $r < n$ le rang de X. On peut trouver P, Q inversibles telles que $\mathrm{A} = \mathrm{P}J_r\mathrm{Q}$ où J_r est la matrice diagonale dont les r premiers coefficients diagonaux valent 1, les autres étant nuls. Pour $t \in [0,1]$, posons

$$\psi(t) = \mathrm{P}\ \mathrm{Diag}(1, \dots, 1, t, \dots, t)\mathrm{Q} \text{ et } \varphi(t) = \mathrm{P}\ \mathrm{Diag}(1, \dots, 1, -t, t, \dots, t)\mathrm{Q}.$$

On a $\psi(0) = \varphi(0) = \mathrm{A}$, $\psi(t)$ et $\varphi(t)$ inversibles pour tout $t \in]0,1]$,

donc éléments de X. Enfin, $\det\varphi(1) = -\det\psi(1)$, donc $\psi(1)$ et $\varphi(1)$ ne sont pas dans la même composante connexe de $\mathrm{GL}_n(\mathbb{R})$ et le résultat est prouvé. ◁

4.19. Connexité d'un ensemble de matrices

Soit X l'ensemble des matrices $A \in \mathcal{M}_n(\mathbb{R})$ telles que $\mathrm{Ker}\, A$ n'est pas inclus dans $\mathrm{Im}\, A$.

1. Montrer que X est connexe par arcs.

2. Quelle est l'adhérence de X ?

(École normale supérieure)

▷ **Solution.**

1. L'ensemble X est connexe par arcs car il est étoilé par rapport à la matrice nulle. En effet, si $A \in X$ alors pour tout $t \in [0,1]$ la matrice tA est dans X puisque si $t > 0$ on a $\mathrm{Ker}\, tA = \mathrm{Ker}\, A$ et $\mathrm{Im}\, tA = \mathrm{Im}\, A$.

2. Comme X ne contient aucune matrice inversible il est inclus dans le complémentaire de $\mathrm{GL}_n(\mathbb{R})$. Cet ensemble étant fermé, on a $\overline{X} \subset \mathcal{M}_n(\mathbb{R}) \setminus \mathrm{GL}_n(\mathbb{R})$. Montrons qu'il y a égalité. Soit A une matrice non inversible. Si $\mathrm{Ker}\, A$ n'est pas inclus dans $\mathrm{Im}\, A$ alors $A \in X$ et c'est fini. Supposons donc que $\mathrm{Ker}\, A \subset \mathrm{Im}\, A$. En prenant une base de $\mathrm{Ker}\, A$ complétée en une base de $\mathrm{Im}\, A$ puis complétée en une base de $\mathbb{R}^n$ on voit que A est semblable à une matrice de la forme $A' = \begin{pmatrix} 0 & B \\ 0 & 0 \end{pmatrix}$ avec $B \in \mathcal{M}_r(\mathbb{R})$ où $r = \mathrm{rg}\, A$. Il suffit de montrer que A' est adhérente à X. Notons $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$. Pour $\varepsilon > 0$, notons A_ε la matrice obtenue à partir de A' en mettant dans la première colonne sa dernière colonne multipliée par ε, les autres colonnes étant inchangées. Il est clair que A_ε tend vers A' lorsque ε tend vers 0^+. L'image n'est pas modifiée par l'opération précédente : pour tout $\varepsilon > 0$ on a $\mathrm{Im}\, A_\varepsilon = \mathrm{Im}\, A' = \mathrm{Vect}(e_1, \ldots, e_r)$. En revanche $e_1 - \varepsilon e_n \in \mathrm{Ker}\, A_\varepsilon$ et comme ce vecteur n'est pas dans l'image on a $A_\varepsilon \in X$. ◁

4.20. Chemin continu de projecteurs

Soit $f : [0,1] \to \mathcal{L}(\mathbb{R}^n)$ continue et telle que $f(t) = f(t) \circ f(t)$ pour tout $t \in [0,1]$.

1. Montrer que $\mathrm{rg}\, f(t)$ ne dépend pas de $t \in [0,1]$. On le note r et on suppose $r \geqslant 1$.

2. Montrer qu'il existe $v_1, \ldots, v_r$ des fonctions continues de $[0,1]$ dans $\mathbb{R}^n$ telles que $(v_1(t), \ldots, v_r(t))$ est une base de $\mathrm{Im}\, f(t)$ pour tout $t \in [0,1]$.

(École normale supérieure)

▷ **Solution.**

1. Pour tout $t \in [0,1]$, $f(t)$ est un projecteur de $\mathbb{R}^n$ donc son rang est égal à sa trace. Il en découle que la fonction $t \mapsto \operatorname{rg} f(t) = \operatorname{Tr} f(t)$ est continue sur $[0,1]$ et, comme elle prend ses valeurs dans l'ensemble discret $\{0,1,\dots,n\}$, elle est constante.

2. On va commencer par résoudre le problème localement au voisinage d'un point $t_0 \in [0,1]$. Supposons qu'on dispose d'une base $(e_1,\dots,e_r)$ de $\operatorname{Im} f(t_0)$. On la complète avec une base de $\operatorname{Ker} f(t_0)$ pour avoir une base $(e_1,\dots,e_n)$ de $\mathbb{R}^n$. La matrice de $f(t_0)$ dans cette base est $\begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$. Pour tout t, notons $\mathrm{M}(t) = \begin{pmatrix} \mathrm{A}(t) & \mathrm{B}(t) \\ \mathrm{C}(t) & \mathrm{D}(t) \end{pmatrix}$ la matrice de $f(t)$ dans cette même base. Comme $\mathrm{A}(t_0) = \mathrm{I}_r$ la continuité de $t \mapsto \mathrm{A}(t)$ et celle du déterminant permettent de dire que $\mathrm{A}(t)$ est inversible pour t au voisinage de t_0. En particulier la famille $(f(t)(e_1),\dots,f(t)(e_r))$ est libre pour de tels t et c'est donc une base de $\operatorname{Im} f(t)$, puisqu'elle comporte r vecteurs. On vient donc d'exhiber une base de $\operatorname{Im} f(t)$ qui dépend continûment de t pour t au voisinage de t_0, et cette base coïncide en t_0 avec la base initialement choisie de $\operatorname{Im} f(t_0)$.

Il reste à passer de cette solution locale à une solution globale. Considérons pour cela l'ensemble J des réels $\mathrm{T} \in [0,1]$ ayant la propriété suivante : il existe des fonctions continues $v_1,\dots,v_r$ de $[0,\mathrm{T}]$ dans $\mathbb{R}^n$ telles que $(v_1(t),\dots,v_r(t))$ est une base de $\operatorname{Im} f(t)$ pour tout $t \in [0,\mathrm{T}]$. Il est clair que J est un intervalle de $[0,1]$ dont l'une des extrémités est 0, car si $\mathrm{T} \in \mathrm{J}$ alors $[0,\mathrm{T}] \subset \mathrm{J}$. L'application de ce qui précède avec $t_0 = 0$ montre que J n'est pas réduit au singleton $\{0\}$. Il ne peut pas non plus être de la forme $[0,\alpha]$ avec $0 < \alpha < 1$ car, en appliquant ce qui précède avec $t_0 = \alpha$, on pourrait prolonger continûment la base de $\operatorname{Im} f(\alpha)$ à droite de α. Il nous reste à exclure le cas où J est de la forme $[0,\alpha[$ avec $0 < \alpha \leqslant 1$. Appliquons le résultat local précédent en α. Il existe $\eta > 0$ (avec $\eta < \alpha$) et des fonctions continues $w_1,\dots,w_r$ définies sur $[\alpha-\eta,\alpha]$ telles que $(w_1(t),\dots,w_r(t))$ soit une base de $\operatorname{Im} f(t)$ pour tout $t \in [\alpha-\eta,\alpha]$. Par ailleurs le réel $\alpha-\eta$ est dans J donc on dispose de fonctions continues $v_1,\dots,v_r$ définies sur $[0,\alpha-\eta]$ ayant la même propriété. On dispose alors de deux bases de l'espace $\operatorname{Im} f(\alpha-\eta)$. Il existe donc une matrice inversible $\mathrm{P} \in \mathrm{GL}_r(\mathbb{R})$ telle que

$$(v_1(\alpha-\eta),\dots,v_r(\alpha-\eta)) = (w_1(\alpha-\eta),\dots,w_r(\alpha-\eta))\mathrm{P},$$

où les r-uplets de vecteurs de $\mathbb{R}^n$ sont identifiés à des matrices de taille (n,r). Pour $t \in]\alpha-\eta,\alpha]$, on pose alors

$$(v_1(t),\dots,v_r(t)) = (w_1(t),\dots,w_r(t))\mathrm{P}.$$

On a toujours une base de $\operatorname{Im} f(t)$ et les fonctions v_k ainsi prolongées à $[0,\alpha]$ sont toutes continues d'après ce qui précède. Ainsi $\alpha \in \mathrm{J}$, ce qui est contradictoire.

On a donc finalement $J = [0,1]$, ce qui est le résultat demandé. ◁

Le dernier thème de ce chapitre est consacré à l'exponentielle des matrices. Les premiers exercices concernent les propriétés fondamentales de l'exponentielle. Ils sont suivis de quelques exercices sur les sous-groupes à un paramètre et de diverses recherches de morphismes continus de groupes. Le chapitre se termine par des exercices issus de la théorie des groupes de Lie. Le lecteur trouvera les exercices liés aux équations différentielles linéaires dans le tome analyse 3.

Rappelons que pour $A \in \mathcal{M}_n(\mathbb{C})$ *on pose*

$$\exp A = \sum_{k=0}^{+\infty} \frac{A^k}{k!}$$

et cette série est absolument convergente. À l'aide d'un produit de Cauchy, on montre que si A *et* B *commutent, on a* $\exp(A+B) = \exp A \exp B$. *En particulier, en prenant* $B = -A$, *on voit que* $\exp A$ *est toujours inversible d'inverse* $\exp(-A)$.

4.21. Exponentielle d'une matrice à coefficients non diagonaux positifs

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ et $B = \exp(A) = (b_{ij})_{1 \leqslant i,j \leqslant n}$. On suppose que pour tout $(i,j) \in [\![1,n]\!]^2$, $i \neq j$, $a_{ij} \geqslant 0$. Montrer que pour tout $(i,j) \in [\![1,n]\!]^2$, $b_{ij} \geqslant 0$.

(École polytechnique)

▷ **Solution.**

Si C et D sont des matrices à coefficients positifs et λ un réel positif, les matrices λC, CD, C + D sont clairement à coefficients positifs. Il en découle que $\exp C = \sum_{n=0}^{+\infty} \frac{C^n}{n!}$ est à coefficients positifs.

Soit $\lambda = \max_{1 \leqslant i \leqslant n} |a_{ii}|$. La matrice $A + \lambda I_n$ est à coefficients positifs, donc la matrice $\exp(A + \lambda I_n)$ également. Mais il en est de même de la matrice $\exp(-\lambda I_n) = \mathrm{Diag}(e^{-\lambda}, \ldots, e^{-\lambda})$. Comme λI_n et $A + \lambda I_n$ commutent, on obtient

$$\exp A = \exp(A + \lambda I_n) \exp(-\lambda I_n)$$

donc, comme produit de matrices à coefficients positifs, $\exp A$ est à coefficients positifs. ◁

L'exercice suivant montre que l'exponentielle restreinte aux matrices diagonalisables réelles est injective.

4.22. Exponentielle de matrices réelles diagonalisables

Soient A, B deux matrices diagonalisables de $\mathcal{M}_n(\mathbb{R})$. On suppose que $\exp A = \exp B$. Montrer que $A = B$.

(École polytechnique)

▷ **Solution.**

Si A et B sont toutes les deux diagonales, le résultat est évident car la fonction exponentielle est injective sur $\mathbb{R}$. Il nous suffit donc de montrer que A et B sont codiagonalisables pour conclure. On sait pour cela qu'il suffit de prouver que A et B commutent (cf. exercice 2.21). Il est clair que A commute avec $\exp A = \exp B$. Or, B est un polynôme en $\exp B$. Pour le voir il suffit de se placer dans une base diagonalisant B : si $B = P^{-1}DP$ avec P inversible et $D = \mathrm{Diag}(d_1, \ldots, d_n)$, alors $\exp B = P^{-1}D'P$ avec $D' = \mathrm{Diag}(e^{d_1}, \ldots, e^{d_n})$. Or il est aisé de trouver un polynôme Q tel que $Q(e^{d_k}) = d_k$ pour tout k (en utilisant un polynôme d'interpolation de Lagrange). On a alors $Q(D') = D$ et donc $Q(\exp B) = B$. Ainsi A commute avec B. ◁

Dans l'exercice qui suit, on utilise encore le fait que si M *est une matrice réelle diagonalisable alors* M *est un polynôme en* $\exp(M)$.

4.23. Exponentielle des matrices symétriques réelles

Montrer que l'exponentielle induit un homéomorphisme de $\mathcal{S}_n(\mathbb{R})$ sur $\mathcal{S}_n^{++}(\mathbb{R})$.

(École normale supérieure)

▷ **Solution.**

Il est clair, par continuité de la transposition, que si $M \in \mathcal{M}_n(\mathbb{R})$ on a ${}^t(\exp M) = \exp({}^tM)$, donc l'exponentielle d'une matrice symétrique est symétrique. En fait, si $M \in \mathcal{S}_n(\mathbb{R})$, on sait d'après le théorème spectral qu'il existe $P \in O_n(\mathbb{R})$ et $D = \mathrm{Diag}(d_1, \ldots, d_n)$ telles que $M = PDP^{-1}$. On a alors $\exp M = P\,\mathrm{Diag}(e^{d_1}, \ldots, e^{d_n})P^{-1}$, ce qui prouve que $\exp M$ est symétrique et définie positive.

Inversement, toute matrice N de $\mathcal{S}_n^{++}(\mathbb{R})$ peut s'écrire sous la forme $P\,\mathrm{Diag}(d'_1, \ldots, d'_n)P^{-1}$, avec P orthogonale et les d'_i strictement positifs. Ce qui précède montre que $N = \exp M$ où M est la matrice symétrique $M = P\,\mathrm{Diag}(\ln d'_1, \ldots, \ln d'_n)P^{-1}$. C'est en fait la seule solution car si $M' \in \mathcal{S}_n(\mathbb{R})$ vérifie $\exp M' = N = \exp M$ alors M' commute avec $\exp M$, donc avec M qui est un polynôme en $\exp M$ (voir exercice 4.22). Par suite M' et M se diagonalisent dans une même base et il est facile d'en déduire qu'elles sont égales.

On a ainsi montré que exp établit une bijection entre $\mathcal{S}_n(\mathbb{R})$ et $\mathcal{S}_n^{++}(\mathbb{R})$. Cette bijection est clairement continue. Montrons que la bijection réciproque est également continue. On va le faire séquentiellement. Soit $(N_k)_{k\geqslant 0}$ une suite de matrices de $\mathcal{S}_n^{++}(\mathbb{R})$ qui converge vers une matrice $N \in \mathcal{S}_n^{++}(\mathbb{R})$. Considérons, pour $k \in \mathbb{N}$ l'unique matrice symétrique M_k telle que $N_k = \exp M_k$ et également l'unique matrice symétrique M telle que $N = \exp M$. On doit montrer que la suite $(M_k)_{k\geqslant 0}$ converge vers M. Il est clair que la seule valeur d'adhérence de la suite $(M_k)_{k\geqslant 0}$ est M. En effet, si une suite extraite $(M_{\varphi(k)})_{k\geqslant 0}$ converge vers $M' \in \mathcal{S}_n(\mathbb{R})$, la continuité de l'exponentielle montre que $N_{\varphi(k)}$ converge vers $\exp M'$. Or cette suite convergeant vers $N = \exp M$, on a $M' = M$ par injectivité. Pour conclure, il suffit de prouver que la suite $(M_k)_{k\geqslant 0}$ est bornée (car une suite bornée d'un espace normé de dimension finie qui possède une unique valeur d'adhérence converge vers cette valeur). Travaillons avec la norme triple induite par la norme euclidienne. Il est classique que pour $A \in \mathcal{S}_n(\mathbb{R})$ on a $|||A||| = \rho(A) = \max_{\lambda\in \mathrm{Sp}\, A} |\lambda|$. Comme la suite $(N_k)_{k\geqslant 0}$ converge vers N, elle est bornée et il existe une constante $C > 0$ telle que $|||N_k||| \leqslant C$ pour tout k. En particulier les valeurs propres (positives) des matrices N_k sont toutes majorées par C et il en découle que les valeurs propres des matrices M_k sont toutes majorées par $\ln C$. Cela ne suffit pas pour conclure, car les valeurs propres de M_k peuvent être négatives ! Il nous faut aussi les minorer. On utilise le fait que la suite $(N_k^{-1})_{k\geqslant 0}$ converge vers N^{-1} par continuité de l'inverse. On en déduit l'existence de $c > 0$ tel que toutes les valeurs propres des matrices N_k^{-1} soient majorées par c. Par suite toutes les valeurs propres des matrices M_k sont minorées par $-\ln c$. On en déduit que $|||M_k||| \leqslant \max(|\ln c|, |\ln C|)$ pour tout k et la suite $(M_k)_{k\geqslant 0}$ est bornée. ◁

4.24. Image de l'exponentielle

1. Montrer que l'exponentielle réalise une bijection entre les matrices nilpotentes et les matrices unipotentes (c'est-à-dire les matrices qui s'écrivent $I_n + N$, avec N nilpotente) de $\mathcal{M}_n(\mathbb{C})$.

2. Quelle est l'image de l'application $M \longmapsto \exp M$ de $\mathcal{M}_n(\mathbb{C})$ dans lui-même ?

(**École polytechnique**)

▷ **Solution.**

1. Si N est nilpotente, alors $\exp N = \sum_{k=0}^{n-1} \frac{N^k}{k!}$. La matrice $N' = \sum_{k=1}^{n-1} \frac{N^k}{k!}$ est nilpotente puisque de la forme NP(N) avec P polynôme. Donc $\exp N$ est unipotente.

Notons $\mathcal{N}$ l'ensemble des matrices nilpotentes. Il s'agit de démontrer que l'application $\psi : \mathrm{N} \longmapsto \exp \mathrm{N} - \mathrm{I}_n$ est une bijection de $\mathcal{N}$ sur $\mathcal{N}$. Pour écrire la bijection réciproque, on pense bien entendu au logarithme et cela ne pose aucun problème particulier car on travaille avec des matrices nilpotentes. Considérons donc

$$\varphi : \mathrm{N} \longmapsto \ln(\mathrm{I}_n + \mathrm{N}) = \sum_{k=1}^{n-1} (-1)^{k-1} \frac{\mathrm{N}^k}{k}.$$

Notons d'abord que φ est, comme ψ, une application de $\mathcal{N}$ dans $\mathcal{N}$ car $\varphi(\mathrm{N}) = \mathrm{NR}(\mathrm{N})$ où R est un polynôme. Il ne reste plus qu'à justifier que $\psi \circ \varphi = \varphi \circ \psi = \mathrm{Id}$. Posons $\mathrm{P}(\mathrm{X}) = \sum_{k=1}^{n-1} \frac{\mathrm{X}^k}{k!}$ et $\mathrm{Q}(\mathrm{X}) = \sum_{k=1}^{n-1} (-1)^{k-1} \frac{\mathrm{X}^k}{k}$. L'application ψ associe à N la matrice $\mathrm{P}(\mathrm{N})$ et l'application φ associe à N la matrice $\mathrm{Q}(\mathrm{N})$. Ainsi, $\psi \circ \varphi$ associe à N la matrice $\mathrm{P}(\mathrm{Q}(\mathrm{N})) = (\mathrm{P} \circ \mathrm{Q})(\mathrm{N})$. Pour la calculer on a seulement besoin du polynôme $\mathrm{P} \circ \mathrm{Q}$ tronqué à la puissance $n-1$. Or, comme

$$e^x - 1 = \mathrm{P}(x) + o(x^{n-1}) \quad \text{et} \quad \ln(1+x) = \mathrm{Q}(x) + o(x^{n-1}),$$

celui-ci est exactement la partie polynomiale du développement limité à l'ordre $n-1$ en 0 de la fonction $e^{\ln(1+x)} - 1 = x$, c'est-à-dire X. On a donc $(\psi \circ \varphi)(\mathrm{N}) = (\mathrm{P} \circ \mathrm{Q})(\mathrm{N}) = \mathrm{N}$ pour toute matrice nilpotente N. On obtient de même $(\varphi \circ \psi)(\mathrm{N}) = (\mathrm{Q} \circ \mathrm{P})(\mathrm{N}) = \mathrm{N}$, puisque $\ln(1+e^x-1) = x$.

Conclusion. Pour toute matrice nilpotente N′, il existe une unique matrice nilpotente N telle que $\exp \mathrm{N} = \mathrm{I}_n + \mathrm{N}'$.

L'exponentielle induit un homéomorphisme de l'ensemble des matrices nilpotentes sur l'ensemble des matrices unipotentes. Le résultat reste vrai dans $\mathcal{M}_n(\mathbb{R})$.

2. Nous savons que si $\mathrm{M} \in \mathcal{M}_n(\mathbb{C})$, alors $\exp(\mathrm{M})$ est inversible puisque, M et $-\mathrm{M}$ commutant, on a

$$\exp(\mathrm{M}) \exp(-\mathrm{M}) = \exp(0) = \mathrm{I}_n.$$

On sait également que l'image de l'exponentielle sur $\mathbb{C}$ est $\mathbb{C}^*$: il est donc raisonnable de penser que l'image de $\mathcal{M}_n(\mathbb{C})$ par l'exponentielle est $\mathrm{GL}_n(\mathbb{C})$ tout entier. C'est ce que nous allons prouver.

• Soit B une matrice inversible et supposons que le spectre de B soit réduit à un seul élément $\lambda \in \mathbb{C}^*$. La matrice $\mathrm{N} = \frac{1}{\lambda}(\mathrm{B} - \lambda \mathrm{I}_n)$ est donc nilpotente puisqu'en vertu du théorème de Cayley-Hamilton, B est annulée par son polynôme caractéristique $(\mathrm{X} - \lambda)^n$. On a $\mathrm{B} = \lambda(\mathrm{I}_n + \mathrm{N})$. D'après la question précédente, la matrice unipotente $\mathrm{I}_n + \mathrm{N}$ est l'exponentielle d'une matrice M. De plus, λ non nul s'écrit e^μ avec $\mu \in \mathbb{C}$. Dans ces conditions, on a

$$\mathrm{B} = e^\mu \exp(\mathrm{M}) = e^\mu \mathrm{I}_n \exp(\mathrm{M}) = \exp(\mu \mathrm{I}_n) \exp \mathrm{M} = \exp(\mu \mathrm{M}),$$

puisque μI_n commute avec M. On a donc $B \in \operatorname{Im}\exp$.

• Soit B une matrice inversible quelconque. On va se ramener au cas précédent : on écrit $\chi_B = (X-\lambda_1)^{n_1}\dots(X-\lambda_r)^{n_r}$ avec les $\lambda_i \in \mathbb{C}^*$ deux à deux distincts et les $n_i \in \mathbb{N}^*$. Comme B est annulée par χ_B, le théorème de décomposition des noyaux assure que

$$\mathbb{C}^n = \operatorname{Ker}\chi_B(B) = \bigoplus_{i=1}^{r} \operatorname{Ker}(B-\lambda_i I_n)^{n_i}.$$

On considère la concaténation de bases de $\operatorname{Ker}(B-\lambda_i I_n)^{n_i}$: c'est une base $\mathcal{B}$ de $\mathbb{C}^n$. Notons b l'endomorphisme de $\mathbb{C}^n$ canoniquement associé à B. Chaque sous-espace $F_i = \operatorname{Ker}(B-\lambda_i I_n)^{n_i}$ (il s'agit des sous-espaces caractéristiques) est stable par b et l'endomorphisme b_i induit par b sur F_i est annulé par $(X-\lambda_i)^{n_i}$. Son spectre est donc réduit à λ_i. La matrice B′ de l'endomorphisme b dans la base $\mathcal{B}$ est donc diagonale par blocs : $B' = \operatorname{Diag}(B_1,\dots,B_r)$ avec $\operatorname{Sp} B_i = \{\lambda_i\}$. D'après ce qui précède, B_i s'écrit $\exp(M_i)$ et si l'on pose $M = \operatorname{Diag}(M_1,\dots,M_r)$, on a $B' = \exp M$. Enfin, en notant P la matrice de passage de la base canonique à $\mathcal{B}$, on a $B = PB'P^{-1} = P\exp(M)P^{-1} = \exp(PMP^{-1})$ et B est effectivement dans l'image de l'exponentielle.

Conclusion. L'image de $\mathcal{M}_n(\mathbb{C})$ par l'exponentielle est $\mathrm{GL}_n(\mathbb{C})$. ◁

On notera qu'un élément $B \in \mathrm{GL}_n(\mathbb{C})$ *donné a toujours une infinité d'antécédents puisque à un bloc* M_i *on peut ajouter* $2ik\pi I$ *avec* $k \in \mathbb{Z}$.

Du résultat de l'exercice, on déduit que $\mathrm{GL}_n(\mathbb{C})$ *est connexe par arcs, car c'est l'image par la fonction* exp, *qui est continue, de* $\mathcal{M}_n(\mathbb{C})$ *qui est connexe par arcs. On peut aussi démontrer que si* $A \in \mathrm{GL}_n(\mathbb{C})$ *et* $p \in \mathbb{N}^*$, *il existe* $B \in \mathrm{GL}_n(\mathbb{C})$ *tel que* $A = B^p$. *On écrit* $A = \exp A'$ *et alors* $B = \exp\left(\frac{1}{p}A'\right)$ *convient.*

Lorsqu'on se place sur $\mathbb{R}$, *la situation se complique et l'image n'est pas égale à* $\mathrm{GL}_n(\mathbb{R})$, *ce qui se voit déjà pour* $n = 1$. *On peut en fait démontrer que l'image est l'ensemble des carrés de* $\mathrm{GL}_n(\mathbb{R})$, *qui est un sous-ensemble strict de* $\mathrm{GL}_n^+(\mathbb{R})$ *(si* $n \geqslant 2$*).*

Dans l'exercice suivant on cherche les antécédents de l'identité par l'exponentielle. On y utilisera la décomposition de Dunford étudiée dans l'exercice 2.40 : toute matrice A *de* $\mathcal{M}_n(\mathbb{C})$ *se décompose de manière unique sous la forme* $A = D + N$ *avec* D *diagonalisable,* N *nilpotente et* $DN = ND$.

4.25. Antécédent de l'identité par l'exponentielle

1. Déterminer la décomposition de Dunford de $\exp A$ en fonction de celle de A.

2. Déterminer les matrices A de $\mathcal{M}_n(\mathbb{C})$ telles que $\exp A = I_n$. **(École polytechnique)**

▷ **Solution.**

1. Posons $A = D + N$ avec D diagonalisable, N nilpotente et $DN = ND$. Comme D et N commutent on a $\exp A = \exp D \exp N$. Posons $\exp N = I_n + N'$ avec $N' = \sum_{k=1}^{n-1} \frac{N^k}{k!}$. La matrice N' est nilpotente puisque de la forme $NP(N)$ avec P polynôme. La matrice $\exp D$ est diagonalisable car si $D = P^{-1}\,\mathrm{Diag}(d_1, \ldots, d_n)P$ alors $\exp D = P^{-1}\,\mathrm{Diag}(e^{d_1}, \ldots, e^{d_n})P$. De plus, comme $\exp D$ et $\exp N$ commutent, N' commute avec $\exp D$. L'écriture

$$\exp A = \exp D + (\exp D)N'$$

est donc la décomposition de Dunford de $\exp A$ car $(\exp D)N'$ est nilpotente et commute avec $\exp D$.

2. Soit $A \in \mathcal{M}_n(\mathbb{C})$ vérifiant $\exp A = I_n$. On reprend les notations de la question précédente. L'unicité de la décomposition de Dunford permet de dire que

$$\exp A = I_n \iff (\exp D = I_n \text{ et } (\exp D)N' = 0)$$

soit encore $\exp A = I_n \iff (\exp D = I_n \text{ et } N' = 0)$. Or, $N' = NP(N)$ avec $P(X) = \sum_{k=1}^{n-1} \frac{X^{k-1}}{k!}$. Le coefficient constant de P est égal à 1. Il en découle que la matrice $P(N)$ est inversible : en effet, dans une base trigonalisant N, on obtient une matrice unipotente (triangulaire supérieure avec des 1 sur la diagonale). Comme $N' = 0$, on a $N = 0$ et la matrice A est diagonalisable. En se plaçant dans une base de diagonalisation on voit que les valeurs propres de A doivent être dans $2i\pi\mathbb{Z}$. Réciproquement, toute matrice diagonalisable dont le spectre est contenu dans $2i\pi\mathbb{Z}$ a comme image I_n par l'exponentielle. ◁

La seule solution réelle est la matrice I_n.

L'exponentielle permet de décrire les sous-groupes à un paramètre du groupe linéaire. On utilisera le fait que si $A \in \mathcal{M}_n(\mathbb{C})$ *la fonction* $t \mapsto \exp(tA)$ *est dérivable de dérivée* $t \mapsto A\exp(tA)$.

4.26. Sous-groupes à un paramètre de $\mathrm{GL}_n(\mathbb{C})$

1. Déterminer les morphismes de groupes dérivables de $(\mathbb{R}, +)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$.

2. En déduire les morphismes de groupes continus de $(\mathbb{R}, +)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$.

(**École polytechnique**)

▷ **Solution.**

1. Soit $f : \mathbb{R} \to \mathrm{GL}_n(\mathbb{C})$ un morphisme dérivable. On a, pour tout coupe $(s, t) \in \mathbb{R}^2$,

$$f(s + t) = f(s)f(t).$$

En dérivant par rapport à t, on obtient $f'(s + t) = f(s)f'(t)$ pour $(s, t) \in \mathbb{R}^2$. puis $f'(s) = f(s)f'(0)$ en prenant $t = 0$. Posons $\mathrm{A} = f'(0)$ et considérons la fonction $g : \mathbb{R} \longrightarrow \mathcal{M}_n(\mathbb{C})$ définie par $g(s) = f(s)e^{-s\mathrm{A}}$. Les fonctions f et $s \longmapsto e^{-s\mathrm{A}}$ sont dérivables, donc g également. Pour tout $s \in \mathbb{R}$,

$$g'(s) = f'(s)e^{-s\mathrm{A}} + f(s)(-\mathrm{A}e^{-s\mathrm{A}}) = (f'(s) - f(s)\mathrm{A})e^{-s\mathrm{A}} = 0.$$

L'application g est donc constante. De plus, $g(0) = f(0) = \mathrm{I}_n$, donc, pour tout $s \in \mathbb{R}$, $g(s) = f(s)e^{-s\mathrm{A}} = \mathrm{I}_n$. On en déduit que pour tout $s \in \mathbb{R}$, $f(s) = e^{s\mathrm{A}}$.

Réciproquement, A étant une matrice quelconque de $\mathcal{M}_n(\mathbb{C})$, on vérifie que l'application

$$f : t \in \mathbb{R} \longmapsto e^{t\mathrm{A}} \in \mathrm{GL}_n(\mathbb{C})$$

est un morphisme dérivable de $(\mathbb{R}, +)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$. Cela résulte immédiatement des propriétés de l'exponentielle.

2. Soit $f : \mathbb{R} \to \mathrm{GL}_n(\mathbb{C})$ un morphisme continu. On va montrer en fait que f est nécessairement dérivable. L'idée est d'utiliser une primitive de f. Posons pour $x \in \mathbb{R}$, $\mathrm{F}(x) = \displaystyle\int_0^x f(t)dt$. La fonction F est de classe $\mathcal{C}^1$ sur $\mathbb{R}$. Imaginons qu'il existe $a > 0$ tel que la matrice $\mathrm{F}(a)$ soit inversible. On a $f(x + u) = f(x)f(u)$ pour tout $(x, u) \in \mathbb{R}^2$. Intégrons cette égalité par rapport à u sur le segment $[0, a]$. Il vient

$$\int_0^a f(x + u)\mathrm{d}u = f(x) \int_0^a f(u)\mathrm{d}u = f(x)\mathrm{F}(a).$$

Or, $g(x) = \displaystyle\int_0^a f(x+u)\mathrm{d}u = \int_x^{x+a} f(t)\mathrm{d}t$ est une fonction dérivable de x car f est continue. Comme on a $f(x) = g(x)\mathrm{F}(a)^{-1}$, f est aussi dérivable et on peut conclure avec la question précédente.

Il ne reste plus qu'à prouver l'existence de a. Pour cela on observe que $F'(0) = f(0) = I_n$. Comme le taux d'accroissement $\frac{1}{t}F(t)$ tend vers I_n lorsque t tend vers 0^+, et comme le groupe $GL_n(\mathbb{C})$ est ouvert, on est certain que $\frac{1}{t}F(t)$ est inversible pour t assez petit. Il en est alors de même de $F(t)$.

Conclusion. Les morphismes continus de $(\mathbb{R},+)$ dans $(GL_n(\mathbb{C}),\times)$ sont les applications $t \longmapsto e^{tA}$, où A est une matrice quelconque. ◁

Il a aussi été proposé à l'oral des ENS la détermination des morphismes continus de $(\mathbb{R},+)$ *dans* $(SL_2(\mathbb{R}),\times)$. *Il est facile de voir qu'il s'agit des* $t \mapsto \exp(tA)$ *où* A *est une matrice de* $\mathcal{M}_2(\mathbb{R})$ *de trace nulle puisque* $\det(\exp M) = e^{\mathrm{Tr}\, M}$ *pour toute matrice* M.

Voici une variante où on n'impose plus d'avoir des matrices inversibles à l'arrivée.

4.27. Groupes à un paramètre dans $\mathcal{M}_n(\mathbb{R})$

Soit $\varphi : \mathbb{R} \longrightarrow \mathcal{M}_n(\mathbb{R})$ de classe $\mathcal{C}^1$ telle que pour tout $(s,t) \in \mathbb{R}^2$, $\varphi(s+t) = \varphi(s)\varphi(t)$. Montrer qu'il existe $(P,A) \in \mathcal{M}_n(\mathbb{R})$ telles que $P^2 = P$ et pour tout $t \in \mathbb{R}$, $\varphi(t) = P\exp(tA)$.

(École polytechnique)

▷ **Solution.**

Notons $P = \varphi(0)$. On a $P^2 = \varphi(0)\varphi(0) = \varphi(0+0) = P$. De plus, pour tout $(s,t) \in \mathbb{R}^2$, en dérivant $\varphi(s+t) = \varphi(s)\varphi(t)$ par rapport à t, on obtient $\varphi'(s+t) = \varphi'(t)\varphi(s)$ et en particulier, $\varphi'(s) = \varphi'(0)\varphi(s) = A\varphi(s)$ en posant $A = \varphi'(0)$. On en déduit que, pour tout s de $\mathbb{R}$,

$$\varphi(s) = \varphi(0)\exp(sA) = P\exp(sA). \quad \triangleleft$$

Dans les deux énoncés suivants, on remplace le groupe additif $\mathbb{R}$ *par le groupe des nombres complexes de module* 1, *qui est isomorphe à* $\mathbb{R}/\mathbb{Z}$.

4.28. Morphismes continus de S^1 dans $GL_n(\mathbb{C})$

On note S^1 le groupe des nombres complexes de module 1.

1. Déterminer les morphismes continus de $(S^1,\times)$ dans $(S^1,\times)$.

2. Déterminer les morphismes de groupes continus de $(S^1,\times)$ dans $(GL_n(\mathbb{C}),\times)$.

(École normale supérieure)

▷ **Solution.**

1. Soit $f : \mathrm{S}^1 \to \mathrm{S}^1$ un morphisme continu et $g : t \in \mathbb{R} \mapsto f(e^{it})$. La fonction g est un morphisme continu du groupe $(\mathbb{R}, +)$ dans le groupe $(\mathrm{S}^1, \times)$. On peut alors utiliser l'exercice 4.26 dans le cas $n = 1$: il existe $a \in \mathbb{C}$ tel que $g(t) = e^{at}$ pour tout t. Mais comme g est 2π-périodique a est nécessairement de la forme ik avec $k \in \mathbb{Z}$, ce qui donne $f(z) = z^k$ pour tout $z \in \mathrm{S}^1$. Réciproquement les applications $z \mapsto z^k$ sont pour tout $k \in \mathbb{Z}$ des morphismes continus de $(\mathrm{S}^1, \times)$ dans $(\mathrm{S}^1, \times)$.

Donnons une solution alternative, qui ne fait pas appel au résultat de l'exercice 4.26, mais qui utilise les coefficients de Fourier. La fonction g est 2π-périodique et continue et vérifie $g(t + t') = g(t)g(t')$, pour tout $(t, t') \in \mathbb{R}^2$. Prenons les coefficients de Fourier en regardant cette égalité comme une égalité de fonctions de la variable t, t' étant fixé. On a pour tout $n \in \mathbb{Z}$, $e^{int'} c_n(g) = g(t')c_n(g)$. Comme la fonction continue g n'est pas nulle il y a au moins un de ses coefficients de Fourier qui n'est pas nul. Et si $c_k(g) \neq 0$ on a $g(t') = e^{ikt'}$ pour tout $t' \in \mathbb{R}$. On retrouve le fait que f est la fonction $z \mapsto z^k$.

2. Soit f un morphisme continu de $(\mathrm{S}^1, \times)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$. Ici aussi, on pourrait relever f en un morphisme continu de $(\mathbb{R}, +)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$ et utiliser l'exercice 4.26, mais nous allons donner une solution plus élémentaire consistant à se ramener au cas $n = 1$ étudié dans la question précédente. Soit $z \in \mathrm{S}^1$, de la forme $z = e^{it}$, avec $\frac{t}{\pi} \in \mathbb{Q}$. Il existe alors $\mathrm{N} \in \mathbb{N}^*$ tel que $z^{\mathrm{N}} = 1$ et par conséquent on a $f(z)^{\mathrm{N}} = \mathrm{I}_n$. Comme le polynôme $\mathrm{X}^{\mathrm{N}} - 1$ est scindé à racines simples, la matrice $f(z)$ est diagonalisable. On note que le sous-groupe G de S^1 formé des nombres complexes de la forme $z = e^{it}$ avec $\frac{t}{\pi} \in \mathbb{Q}$ est commutatif. Ainsi, les matrices $(f(z))_{z \in \mathrm{G}}$ sont diagonalisables et commutent, il est donc possible de les codiagonaliser (voir l'exercice 2.21). Il existe donc une matrice $\mathrm{P} \in \mathrm{GL}_n(\mathbb{C})$ telle que pour tout $z \in \mathrm{G}$, $f(z) = \mathrm{P}\mathrm{D}(z)\mathrm{P}^{-1}$, où $\mathrm{D}(z)$ est diagonale. Or G est dense dans S^1 et f est continue. On en déduit que $\mathrm{D}(z) = \mathrm{P}^{-1} f(z) \mathrm{P}$ reste diagonale pour tout $z \in \mathrm{S}^1$, puisque l'espace des matrices diagonales est fermé. Posons alors $\mathrm{D}(z) = \mathrm{Diag}(\lambda_1(z), \ldots, \lambda_n(z))$. Les applications $z \in \mathrm{S}^1 \mapsto \lambda_i(z)$ sont des morphismes continus de $(\mathrm{S}^1, \times)$ dans $(\mathbb{C}^*, \times)$. Or la solution de la question 1 montre que les morphismes de $(\mathrm{S}^1, \times)$ dans $(\mathbb{C}^*, \times)$ sont les mêmes que les morphismes de $(\mathrm{S}^1, \times)$ dans $(\mathrm{S}^1, \times)$ (cela est dû au le fait qu'un sous-groupe compact de $\mathbb{C}^*$ est inclus dans S^1). Il existe donc des entiers $k_i \in \mathbb{Z}$ tels que $\lambda_i(z) = z^{k_i}$ pour tout $z \in \mathrm{S}^1$. La réciproque étant évidente on peut conclure que les morphismes de groupe continus de $(\mathrm{S}^1, \times)$ dans $(\mathrm{GL}_n(\mathbb{C}), \times)$ sont les applications $z \mapsto \mathrm{P}\,\mathrm{Diag}(z^{k_1}, \ldots, z^{k_n})\mathrm{P}^{-1}$, où $\mathrm{P} \in \mathrm{GL}_n(\mathbb{C})$ et où les k_i sont des entiers quelconques. ◁

Voici le même exercice, mais sur le corps des réels. Il y a une difficulté supplémentaire car les éléments de l'image d'un morphisme ne sont pas forcément diagonalisables.

4.29. Morphismes continus de S^1 dans $GL_n(\mathbb{R})$

Soit $\varphi : (S^1, \times) \to (GL_n(\mathbb{R}), \times)$ un morphisme de groupes continu où S^1 désigne le groupe des nombres complexes de module 1.

1. Montrer que pour tout $z \in S^1$, $\det(\varphi(z)) = 1$.

2. Montrer que pour tout $z \in S^1$, les valeurs propres complexes de $\varphi(z)$ sont de module 1.

3. Achever la description de φ.

(**École normale supérieure**)

▷ **Solution.**

1. Posons $\psi = \det \circ \varphi$. Il s'agit d'un morphisme continu de groupes de S^1 dans $\mathbb{R}^*$. On va montrer que ψ est trivial. Puisque S^1 est connexe, que ψ est continu et que $\psi(1) = 1$, $\psi(S^1)$ est un intervalle inclus dans $\mathbb{R}_+^*$. Comme S^1 est compact, $\psi(S^1)$ est même un segment. De plus, comme ψ est un morphisme, $\psi(S^1)$ est un sous-groupe de $\mathbb{R}_+^*$. Or, le seul segment de $\mathbb{R}_+^*$ qui est un sous-groupe est le singleton $\{1\}$ (un sous-groupe non trivial de $\mathbb{R}_+^*$ n'est pas borné). On a donc $\varphi(S^1) \subset SL_n(\mathbb{R})$.

2. On munit $\mathcal{M}_n(\mathbb{C})$ d'une norme triple quelconque associée à une norme de $\mathbb{C}^n$. Comme S^1 est compact, et φ continu, le groupe image $\varphi(S^1)$ est compact et en particulier borné. Il existe donc une constante $M > 0$ telle que $|||\varphi(z)||| \leqslant M$ pour tout $z \in S^1$. Or si λ est une valeur propre complexe de $\varphi(z)$ on a $|\lambda| \leqslant |||\varphi(z)|||$. Ainsi, l'ensemble des valeurs propres des éléments de $\varphi(S^1)$ est borné. De plus, si λ est dans le spectre de $\varphi(z)$ alors, pour tout $p \in \mathbb{Z}$, λ^p est valeur propre de $\varphi(z)^p = \varphi(z^p) \in \varphi(S^1)$. La suite $(\lambda^p)_{p\in\mathbb{Z}}$ est donc bornée et cela impose que $|\lambda| = 1$.

3. L'application $\psi : t \longmapsto \varphi(e^{it})$ est un morphisme continu de $(\mathbb{R}, +)$ dans $(GL_n(\mathbb{R}), \times)$. D'après l'exercice 4.26, il existe $A \in \mathcal{M}_n(\mathbb{C})$ telle que, pour tout $t \in \mathbb{R}$, $\psi(t) = e^{tA}$. Comme ψ est à valeurs dans $GL_n(\mathbb{R})$, la matrice $A = \psi'(0)$ est réelle. Par construction, ψ est périodique de période 2π. On a donc, pour tout réel t,

$$e^{tA} = e^{2\pi A + tA} = e^{2\pi A} e^{tA}$$

et donc $e^{2\pi A} = I_n$.

La résolution de l'équation $e^M = I_n$ a été menée dans l'exercice 4.25 mais nous allons la refaire ici, et aussi formuler la conclusion de façon un peu plus précise. Les valeurs propres complexes de $e^{2\pi A}$ sont les $e^{2\pi\lambda}$ avec $\lambda \in \operatorname{Sp} A$. Il faut $e^{2\pi\lambda} = 1$ donc $\lambda \in i\mathbb{Z}$.

Montrons que la matrice A est diagonalisable dans $\mathbb{C}$. D'après la décomposition de Dunford, il existe D diagonalisable et N nilpotente

telles que $A = D + N$ et $DN = ND$. Puisque N et D commutent, on a $e^{2\pi A} = e^{2\pi N}e^{2\pi D}$. La matrice $e^{2\pi D}$ est diagonalisable. Comme les valeurs propres de D sont égales à celles de A, les valeurs propres de $e^{2\pi D}$ sont celles de $e^{2\pi A}$ soit 1. Ainsi $e^{2\pi D} = I_n$ et $e^{2\pi N} = I_n$. Supposons que N n'est pas la matrice nulle. On a alors $\operatorname{Ker} N \neq \operatorname{Ker} N^2$. Soit $X \in \operatorname{Ker} N^2 \setminus \operatorname{Ker} N$. Il vérifie $e^{2\pi N}X = X + 2\pi NX \neq X$, ce qui contredit l'égalité $e^{2\pi N} = I_n$. On a donc $N = 0$ et A diagonalisable dans $\mathbb{C}$.

Les valeurs propres de A non nulles sont deux à deux conjuguées. Il existe donc $k_1, \ldots, k_r$ dans $\mathbb{Z}^*$ et $P \in \mathrm{GL}_n(\mathbb{C})$ tels que

$$A = P\,\mathrm{Diag}(ik_1, -ik_1, \ldots, ik_r, -ik_r, 0, \ldots, 0)P^{-1}.$$

On obtient pour tout réel t,

$$e^{tA} = P\,\mathrm{Diag}(e^{itk_1}, e^{-itk_1}, \ldots, e^{itk_r}, e^{-itk_r}, 1, \ldots, 1)P^{-1}.$$

On en déduit qu'il existe une matrice $Q \in \mathrm{GL}_n(\mathbb{C})$, indépendante de t telle que

$$e^{tA} = Q \begin{pmatrix} R_{tk_1} & & & & & \\ & \ddots & & & & \\ & & R_{tk_r} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix} Q^{-1},$$

où $R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$. En effet, pour tout $\theta \in \mathbb{R}$, on peut écrire

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix} = \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix} R_\theta \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix}^{-1}.$$

On peut choisir Q dans $\mathrm{GL}_n(\mathbb{R})$. En effet, en notant B la matrice

$$\begin{pmatrix} R_{tk_1} & & & & & \\ & \ddots & & & & \\ & & R_{tk_r} & & & \\ & & & 1 & & \\ & & & & \ddots & \\ & & & & & 1 \end{pmatrix}$$

et en écrivant $Q = Q_1 + iQ_2$, on a $e^{tA}(Q_1 + iQ_2) = (Q_1 + iQ_2)B$ et donc $e^{tA}Q_1 = Q_1B$ et $e^{tA}Q_2) = Q_2B$. Le polynôme $P = \det(Q_1 + Q_2X)$ n'est pas nul car $P(i) = \det(Q) \neq 0$, donc il existe $x \in \mathbb{R}$ tel que $P(x) \neq 0$, c'est-à-dire $Q_1 + xQ_2$ inversible. On a alors $e^{tA}(Q_1 + xQ_2) = (Q_1 + xQ_2)B$ et $e^{tA} = (Q_1 + xQ_2)B(Q_1 + xQ_2)^{-1}$.

Réciproquement toute application de la forme

$$e^{it} \longmapsto Q \begin{pmatrix} R_{tk_1} & & & & \\ & \ddots & & & \\ & & R_{tk_r} & & \\ & & & 1 & \\ & & & & \ddots \\ & & & & & 1 \end{pmatrix} Q^{-1}$$

est un morphisme continu de S_1 dans $GL_n(\mathbb{R})$. D'une part, une telle application est bien définie car pour $k \in \mathbb{Z}$, R_{tk} ne dépend que de t modulo 2π. De plus c'est un morphisme de groupes car $R_{(t+t')k} = R_{tk}R_{t'k}$ pour t, t' réels et $k \in \mathbb{Z}$. Il est continu car pour tout $(t, t') \in \mathbb{R}^2$ et tout $k \in \mathbb{Z}$, $|e^{kit} - e^{kit'}| \leqslant |k||e^{it} - e^{it'}|$ et donc $|\cos(kt) - \cos(kt')| \leqslant |k||e^{it} - e^{it'}|$ et $|\sin(kt) - \sin(kt')| \leqslant |k||e^{it} - e^{it'}|$. ◁

Nous terminons cette série d'exercices par quelques énoncés provenant de la théorie des groupes de Lie. Pour une introduction en langue française à cette théorie le lecteur pourra consulter le livre de R.MNEIMNE, F. TESTARD, Introduction à la théorie des groupes de Lie classiques, *Hermann, 1986 ou le récent livre de* JACQUES FARAUT, Analyse sur les groupes de Lie, *Calvage et Mounet, 2006.*

4.30. Formules de Lie

On pose, pour $A \in \mathcal{M}_n(\mathbb{R})$ et $p \in \mathbb{N}^*$, $f_p(A) = \left(I_n + \dfrac{A}{p}\right)^p$.

1. Montrer que la suite $(f_p)_{p \geqslant 1}$ converge uniformément sur tout compact de $\mathcal{M}_n(\mathbb{R})$. Quelle en est la limite ?

2. Montrer que

$$\lim_{p \to +\infty} \left(\exp\left(\frac{A}{p}\right) \exp\left(\frac{B}{p}\right)\right)^p = \exp(A + B),$$

$$\lim_{p \to \infty} \left(\exp\left(\frac{A}{p}\right) \exp\left(\frac{B}{p}\right) \exp\left(-\frac{A}{p}\right) \exp\left(-\frac{B}{p}\right)\right)^{p^2} = \exp(AB - BA).$$

3. Soit G un sous-groupe fermé de $GL_n(\mathbb{R})$ et $\mathcal{G}$ l'ensemble des matrices $M \in \mathcal{M}_n(\mathbb{R})$ telles que, pour tout réel t, $\exp(tM) \in G$. Montrer que $\mathcal{G}$ est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$. Par quelle autre opération $\mathcal{G}$ est-il stable ?

(École normale supérieure)

▷ **Solution.**

1. Dans le cas $n = 1$ il est facile de voir que la suite de fonction $f_p : x \mapsto \left(1 + \frac{x}{p}\right)^p$ converge simplement vers la fonction $x \mapsto e^x$ sur $\mathbb{R}$. On va donc essayer de montrer que $f_p(\mathrm{A})$ converge vers $\exp \mathrm{A}$ et de manière uniforme sur un compact de $\mathcal{M}_n(\mathbb{R})$. On fixe une norme triple quelconque sur $\mathcal{M}_n(\mathbb{R})$ On développe $f_p(\mathrm{A})$ par la formule du binôme de Newton et on remplace $\exp \mathrm{A}$ par sa série que l'on tronque au rang p. Il vient :

$$\|\exp \mathrm{A} - f_p(\mathrm{A})\| \leqslant \left\| \sum_{k=0}^{p} \left(\frac{1}{k!} - \frac{p!}{k!(p-k)!p^k} \right) \mathrm{A}^k \right\| + \sum_{k=p+1}^{+\infty} \frac{\|\mathrm{A}\|^k}{k!}.$$

Il s'avère que le coefficient

$$\frac{1}{k!} - \frac{p!}{k!(p-k)!p^k} = \frac{1}{k!}\left(1 - \left(1 - \frac{1}{p}\right) \ldots \left(1 - \frac{p-1}{p}\right)\right)$$

est positif. Supposons que A soit dans un compact K de $\mathcal{M}_n(\mathbb{R})$ et donnons-nous $\mathrm{M} > 0$ tel que $\|\mathrm{A}\| \leqslant \mathrm{M}$ pour tout $\mathrm{A} \in \mathrm{K}$. Par inégalité triangulaire on a alors

$$\begin{aligned} \|\exp \mathrm{A} - f_p(\mathrm{A})\| &\leqslant \sum_{k=0}^{p} \left(\frac{1}{k!} - \frac{p!}{k!(p-k)!p^k} \right) \|\mathrm{A}\|^k + \sum_{k=p+1}^{+\infty} \frac{\|\mathrm{A}\|^k}{k!} \\ &\leqslant \sum_{k=0}^{p} \left(\frac{1}{k!} - \frac{p!}{k!(p-k)!p^k} \right) \mathrm{M}^k + \sum_{k=p+1}^{+\infty} \frac{\mathrm{M}^k}{k!} \\ &\leqslant e^{\mathrm{M}} - \left(1 + \frac{\mathrm{M}}{p}\right)^p. \end{aligned}$$

Le majorant tend vers 0 lorsque p tend vers l'infini.

2. On utilise des développements limités. Lorsque p tend vers l'infini, on a

$$\exp\left(\frac{\mathrm{A}}{p}\right) = \mathrm{I}_n + \frac{\mathrm{A}}{p} + o\left(\frac{1}{p}\right).$$

Il en découle que

$$\left(\exp\left(\frac{\mathrm{A}}{p}\right) \exp\left(\frac{\mathrm{B}}{p}\right)\right)^p = f_p(\mathrm{A} + \mathrm{B} + o(1)) = f_p(\mathrm{A} + \mathrm{B} + \mathrm{C}_p),$$

où (C_p) est une suite de matrices convergente vers 0. Soit $\varepsilon > 0$ et $\mathrm{M} > \|\mathrm{A}+\mathrm{B}\|$. La suite (f_p) converge uniformément sur la boule centrée en 0 de rayon M. Fixons p_0 tel que pour tout $p \geqslant p_0$, $\|f_p - \exp\|_{\infty,\overline{\mathrm{B}}(0,\mathrm{M})} \leqslant \varepsilon$. Or pour p assez grand, $\mathrm{A} + \mathrm{B} + \mathrm{C}_p$ est dans la boule de centre 0 et de rayon M. Par l'inégalité triangulaire, pour p assez grand, on a

$$\begin{aligned}\|f_p(\mathrm{A}+\mathrm{B}+\mathrm{C}_p)-e^{\mathrm{A}+\mathrm{B}}\| &\leqslant \|f_p(\mathrm{A}+\mathrm{B}+\mathrm{C}_p)-e^{\mathrm{A}+\mathrm{B}+\mathrm{C}_p}\|+\\ &\qquad \|e^{\mathrm{A}+\mathrm{B}+\mathrm{C}_p}-e^{\mathrm{A}+\mathrm{B}}\|\\ &\leqslant \varepsilon+\|e^{\mathrm{A}+\mathrm{B}+\mathrm{C}_p}-e^{\mathrm{A}+\mathrm{B}}\|\end{aligned}$$

et par continuité de l'exponentielle, pour p assez grand, on a $\|f_p(\mathrm{A}+\mathrm{B}+\mathrm{C}_p)-e^{\mathrm{A}+\mathrm{B}}\|\leqslant 2\varepsilon$. Cela permet d'affirmer que $f_p(\mathrm{A}+\mathrm{B}+o(1))$ converge vers $\exp(\mathrm{A}+\mathrm{B})$. On procède de même pour la seconde limite mais avec un développement limité à l'ordre 2. On a

$$\exp\left(\frac{\mathrm{A}}{p}\right)\exp\left(\frac{\mathrm{B}}{p}\right)=\mathrm{I}_n+\frac{\mathrm{A}+\mathrm{B}}{p}+\frac{\mathrm{A}^2+2\mathrm{AB}+\mathrm{B}^2}{2p^2}+o\left(\frac{1}{p^2}\right)$$

et de même pour le second terme en replaçant A et B par $-\mathrm{A}$ et $-\mathrm{B}$. En faisant le produit il vient

$$\exp\left(\frac{\mathrm{A}}{p}\right)\exp\left(\frac{\mathrm{B}}{p}\right)\exp\left(-\frac{\mathrm{A}}{p}\right)\exp\left(-\frac{\mathrm{B}}{p}\right)=\mathrm{I}_n+\frac{\mathrm{AB}-\mathrm{BA}}{p^2}+o\left(\frac{1}{p^2}\right).$$

On conclut comme précédemment.

3. Il est clair que $\mathcal{G}$ contient la matrice nulle car $\mathrm{I}_n\in\mathrm{G}$ et que $\mathcal{G}$ est stable par produit par un scalaire. Montrons la stabilité par la somme. Soit A,B deux éléments de $\mathcal{G}$. Pour tout $t\in\mathbb{R}$ et tout $p\in\mathbb{N}^*$ la matrice $\left(\exp\left(\frac{t\mathrm{A}}{p}\right)\exp\left(\frac{t\mathrm{B}}{p}\right)\right)^p$ est dans G, puisque G est stable par produit. Comme G est fermé, la limite lorsque p tend vers l'infini qui vaut $\exp(t(\mathrm{A}+\mathrm{B}))$ est encore dans G. Ainsi $\mathrm{A}+\mathrm{B}\in\mathcal{G}$ et $\mathcal{G}$ est un sous-espace vectoriel de $\mathcal{M}_n(\mathbb{R})$. À l'aide de la seconde limite ci-dessus on voit de même que si A,B sont dans $\mathcal{G}$, alors $\mathrm{AB}-\mathrm{BA}=[\mathrm{A},\mathrm{B}]\in\mathcal{G}$. ◁

Par définition, $\mathcal{G}$ est l'algèbre de Lie du groupe G.

L'exercice suivant détermine ainsi l'algèbre de Lie du sous-groupe des rotations.

4.31. Algèbre de Lie de $\mathrm{SO}_n(\mathbb{R})$

Déterminer $\mathcal{S}=\{\mathrm{A}\in\mathcal{M}_n(\mathbb{R}),\forall t\in\mathbb{R},\ \exp(t\mathrm{A})\in\mathrm{SO}_n(\mathbb{R})\}$.
(École polytechnique)

▷ **Solution.**

• Soit $\mathrm{A}\in\mathcal{S}$. Pour tout $t\in\mathbb{R}$, $\exp(t\mathrm{A})\exp(t({}^t\mathrm{A}))=\mathrm{I}$. En dérivant par rapport à t, on obtient

$$0=\mathrm{A}\exp(t\mathrm{A})\exp(t({}^t\mathrm{A}))+\exp(t\mathrm{A})\exp(t({}^t\mathrm{A}))({}^t\mathrm{A})$$

car pour toute matrice B, $\dfrac{\mathrm{d}\exp(\lambda \mathrm{B})}{\mathrm{d}\lambda} = \mathrm{B}\exp(\lambda \mathrm{B}) = \exp(\lambda \mathrm{B})\mathrm{B}$. Pour $t = 0$, on obtient $\mathrm{A} + {}^t\mathrm{A} = 0$. Donc A est antisymétrique.

• Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ antisymétrique. Alors $\exp \mathrm{A} \in \mathrm{SO}_n(\mathbb{R})$. En effet, on a

$$\exp \mathrm{A}\, {}^t(\exp \mathrm{A}) = \exp \mathrm{A} \exp({}^t\mathrm{A}) = \exp \mathrm{A} \exp(-\mathrm{A}) = \mathrm{I}_n$$

$$\text{et } \det \exp \mathrm{A} = \exp(\mathrm{Tr}\, \mathrm{A}) = \exp 0 = 1.$$

Comme pour tout $t \in \mathbb{R}$, $t\mathrm{A}$ est antisymétrique, $\exp t\mathrm{A} \in \mathrm{SO}_n(\mathbb{R})$ et donc $\mathrm{A} \in \mathcal{S}$.

Conclusion. $\mathcal{S} = \{\mathrm{A} \in \mathcal{M}_n(\mathbb{R}), \mathrm{A} = -{}^t\mathrm{A}\}$ est l'espace des matrices antisymétriques. ◁

Dans l'exercice 1.36 du tome algèbre 3 le lecteur trouvera la preuve du fait que l'exponentielle établit une surjection de l'espace des matrices antisymétriques sur $\mathrm{SO}_n(\mathbb{R})$.

4.32. Sous-groupe fermé de $\mathrm{SL}_2(\mathbb{R})$

Soit G un sous-groupe fermé de $\mathrm{SL}_2(\mathbb{R})$, non commutatif, tel que $\mathrm{G} = -\mathrm{G}$ et pour tout $\mathrm{M} \in \mathrm{G} \setminus \{\mathrm{I}_2, -\mathrm{I}_2\}$, $|\mathrm{Tr}(\mathrm{M})| > 2$.

Soit $\mathrm{G}_0 = \mathrm{G} \setminus \{\mathrm{I}_2, -\mathrm{I}_2\}$.

1. On munit $\mathcal{M}_2(\mathbb{C})$ d'une norme d'algèbre quelconque. Pour $\mathrm{A} \in \mathcal{M}_2(\mathbb{C})$ telle que $\|\mathrm{A}\| < 1$, on pose

$$\ln(\mathrm{I}_2 + \mathrm{A}) = \sum_{n=0}^{+\infty} (-1)^{n+1} \frac{\mathrm{A}^n}{n}.$$

Montrer que, pour toute matrice $\mathrm{A} \in \mathcal{M}_2(\mathbb{C})$ telle que $\|\mathrm{A} - \mathrm{I}_2\| < 1$, on a $\exp(\ln \mathrm{A}) = \mathrm{A}$.

On suppose qu'il existe une suite (A_n) de G_0 qui converge vers I_n.

2. Montrer qu'il existe une matrice $\mathrm{H} \in \mathcal{M}_2(\mathbb{R})$ de norme 1 telle que, pour tout $t \in \mathbb{R}$,

$$\exp(t\mathrm{H}) \in \mathrm{G}.$$

3. En déduire qu'il existe $\mathrm{P} \in \mathrm{GL}_2(\mathbb{R})$ tel que $\mathrm{P}^{-1}\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}\mathrm{P}$ soit dans G pour tout $\lambda \in \mathbb{R}^*$.

Montrer que cela est contradictoire avec G non commutatif. En déduire que G_0 est fermé.

(**École polytechnique**)

▷ **Solution.**

1. La série qui définit le logarithme converge car elle est absolument convergente (en effet, $\|A\| < 1$ et $\left\|(-1)^{n+1}\dfrac{A^n}{n}\right\| \leqslant \dfrac{1}{n}\|A^n\| \leqslant \|A\|^n$ pour $n \geqslant 1$) et $\mathcal{M}_2(\mathbb{C})$ est complet. Soit $A \in B(I_2, 1)$. On imagine que l'égalité demandée est facile à démontrer lorsque A est diagonalisable. Supposons donc qu'il existe $P \in \mathrm{GL}_2(\mathbb{C})$ et $(\lambda, \mu) \in \mathbb{C}^2$ tels que

$$A = P^{-1}\begin{pmatrix} \lambda & 0 \\ 0 & \mu \end{pmatrix}P.$$

Alors $\mathrm{sp}(A - I_2) = \{\lambda - 1, \mu - 1\}$ et comme $\|A - I_2\| < 1$, $|\lambda - 1| < 1$ et $|\mu - 1| < 1$. On a ainsi pour $N \geqslant 1$,

$$\begin{aligned}\sum_{n=1}^{N}(-1)^{n+1}\frac{(A-I_2)^n}{n} &= P^{-1}\left(\sum_{n=1}^{N}\frac{(-1)^{n+1}}{n}\begin{pmatrix}\lambda-1 & 0\\ 0 & \mu-1\end{pmatrix}^n\right)P\\ &= P^{-1}\begin{pmatrix}\displaystyle\sum_{n=1}^{N}(-1)^{n+1}\frac{(\lambda-1)^n}{n} & 0\\ 0 & \displaystyle\sum_{n=1}^{N}(-1)^{n+1}\frac{(\mu-1)^n}{n}\end{pmatrix}P\\ &\xrightarrow[N\to+\infty]{} P^{-1}\begin{pmatrix}\ln\lambda & 0\\ 0 & \ln\mu\end{pmatrix}P\end{aligned}$$

puisque la multiplication par P et P^{-1} est continue. Ainsi, on a

$$\begin{aligned}\exp\ln A &= \exp\left[P^{-1}\begin{pmatrix}\ln\lambda & 0\\ 0 & \ln\mu\end{pmatrix}P\right] = P^{-1}\exp\left[\begin{pmatrix}\ln\lambda & 0\\ 0 & \ln\mu\end{pmatrix}\right]P\\ &= P^{-1}\begin{pmatrix}\exp(\ln\lambda) & 0\\ 0 & \exp(\ln\mu)\end{pmatrix}P = P^{-1}\begin{pmatrix}\lambda & 0\\ 0 & \mu\end{pmatrix}P = A.\end{aligned}$$

Il reste à vérifier le résultat pour A quelconque. Pour cela, nous allons utiliser la densité des matrices diagonalisables dans $\mathcal{M}_2(\mathbb{C})$ (voir l'exercice 4.2). Il existe une suite $(A_n)_{n\in\mathbb{N}}$ de $\mathcal{M}_2(\mathbb{C})$ formée de matrices diagonalisables avec $\lim\limits_{n\to+\infty} A_n = A$. Pour n assez grand, A_n est dans l'ouvert $B(I_2, 1)$ qui contient A. Quitte à tronquer cette suite, on peut supposer que pour $n \geqslant 0$, $\|A_n - I_2\| < 1$. Soit $r \in]0, 1[$. Montrons que $M \longmapsto \ln M$ est continue sur $B(I_2, r)$. Si $M \in B(I_2, r)$, $\left\|(-1)^{n+1}\dfrac{(M-I_2)^n}{n}\right\| \leqslant \dfrac{r^n}{n}$ qui est le terme général d'une série convergente. La série du logarithme est donc normalement convergente, donc la série est continue. La continuité étant une propriété locale, le logarithme est continu sur $B(I_2. 1)$. Ainsi, on a

$$\ln A_n \xrightarrow[n\to+\infty]{} \ln A, \quad \text{et} \quad A_n = \exp\ln A_n \xrightarrow[n\to+\infty]{} \exp\ln A.$$

Par unicité de la limite $A = \exp \ln A$.

2. Pour n assez grand, on a $\|A_n - I_2\| < 1$. Dans ces conditions, $\ln A_n$ est défini et $\ln A_n \neq 0$ car $\exp \ln A_n = A_n \neq I_2$. On pose $H_n = \dfrac{\ln A_n}{\|\ln A_n\|}$. Étant donné que nous sommes en dimension finie et que $\|H_n\| = 1$, d'après le théorème de Bolzano-Weierstrass, il existe une sous-suite $(H_{\varphi(n)})_{n\geqslant 0}$ convergente vers une matrice H. Montrons que H convient. Soit $t \in \mathbb{R}$. On a

$$\exp(tH) = \lim_{n\to+\infty} \exp(tH_{\varphi(n)}) = \lim_{n\to\infty} \exp\left[\frac{t}{\|\ln A_{\varphi(n)}\|} \ln A_{\varphi(n)}\right]$$

Notons pour $P_n = \exp\left[E\left(\dfrac{t}{\|\ln A_{\varphi(n)}\|}\right) \ln A_{\varphi(n)}\right]$. On a

$$\left\|\left(\frac{t}{\|\ln A_{\varphi(n)}\|} - E\left(\frac{t}{\|\ln A_{\varphi(n)}\|}\right)\right) \ln A_{\varphi(n)}\right\| \leqslant \|\ln A_{\varphi(n)}\| \xrightarrow[n\to+\infty]{} \|\ln I_2\| = 0$$

par continuité de ln. Ainsi par continuité de exp, on obtient

$$\exp\left[-\left(\frac{t}{\|\ln A_{\varphi(n)}\|} - E\left(\frac{t}{\|\ln A_{\varphi(n)}\|}\right)\right) \ln A_{\varphi(n)}\right] \xrightarrow[n\to+\infty]{} \exp 0 = I_2.$$

Ainsi, P_n s'écrit

$$P_n = \exp(tH_{\varphi(n)}) \exp\left(-\left(\frac{t}{\|\ln A_{\varphi(n)}\|} - E\left(\frac{t}{\|\ln A_{\varphi(n)}\|}\right)\right) \ln A_{\varphi(n)}\right)$$

et P_n converge vers $e^{tH} I_2 = e^{tH}$. Or $P_n = [A_{\varphi(n)}]^{E\left(\frac{t}{\|\ln A_{\varphi(n)}\|}\right)} \in G$ et comme G est fermé, $e^{tH} = \lim\limits_{n\to+\infty} P_n \in G$.

3. Si $M \in G_0$, on a $\chi_M = X^2 - (\operatorname{Tr} M)X + \det M = X^2 - (\operatorname{Tr} M)X + 1$ qui est de discriminant $\Delta = (\operatorname{Tr} M)^2 - 4 > 0$. Donc χ_M est scindé à racines simples ce qui prouve que M est diagonalisable.

Montrons que H est diagonalisable. Comme $\exp H \in G$, on a

$$1 = \det \exp H = \exp \operatorname{Tr} H,$$

et la trace de H étant réelle, elle est nulle. Il existe $k \in \mathbb{C}$ tel que le spectre complexe de H soit $\{k, -k\}$. En fait, k est nécessairement réel. En effet, pour tout t, le spectre de $\exp tH$ est $\{e^{tk}, e^{-tk}\} \subset \mathbb{R}$, car les éléments de G sont diagonalisables. En dérivant par rapport à t, e^{tk}, on a $ke^{tk} \in \mathbb{R}$ et k est réel.

Si $k = 0$, alors H est nilpotente, $H^2 = 0$ et $\exp H = I_2 + H$. On en déduit $\operatorname{Tr} \exp H = \operatorname{Tr} I_2 + \operatorname{Tr} H = 2 + 0 = 2$ et comme $\exp H \in G$,

$\exp \mathrm{H} = \pm \mathrm{I}_2$ *i.e.* $\mathrm{H} = 0$ ou $\mathrm{H} = -2\mathrm{I}_2$. Ces deux cas sont exclus, puisque H est de norme 1. Donc k est non nul et H est diagonalisable. Il existe $\mathrm{P} \in \mathrm{GL}_2(\mathbb{R})$ tel que $\mathrm{H} = \mathrm{P}^{-1}\begin{pmatrix} k & 0 \\ 0 & -k \end{pmatrix}\mathrm{P}$ et, pour tout $t \in \mathbb{R}$,

$$\exp t\mathrm{H} = \mathrm{P}^{-1}\begin{pmatrix} e^{kt} & 0 \\ 0 & e^{-kt} \end{pmatrix}\mathrm{P} \in \mathrm{G}.$$

Finalement, pour tout $\lambda \in \mathbb{R}^*$,

$$\mathrm{P}^{-1}\begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}\mathrm{P} \in \mathrm{G}.$$

On pose $\mathrm{A}_\lambda = \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix}$. Quitte à remplacer G par un groupe conjugué, on peut supposer que $\mathrm{A}_\lambda \in \mathrm{G}$, pour tout $\lambda \in \mathbb{R}^*$.

Comme G est non commutatif, il existe $\mathrm{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{G}$ tel que $b \neq 0$ ou $c \neq 0$. Considérons une telle matrice. Montrons qu'alors $b \neq 0$ et $c \neq 0$. Supposons que $b = 0$. Alors le produit $\mathrm{MA}_{1/a} = \begin{pmatrix} 1 & 0 \\ c/a & 1 \end{pmatrix}$ est dans G, ce qui est impossible car cette matrice est distincte de I_2 et $-\mathrm{I}_2$ et pourtant sa trace vaut 2. On en déduit que $b \neq 0$ et $c \neq 0$.

Soit $\lambda > 0$. Un petit calcul nous donne

$$\mathrm{A}_\lambda \mathrm{M} \mathrm{A}_\lambda = \begin{pmatrix} a\lambda^2 & b \\ c & d/\lambda^2 \end{pmatrix}.$$

À l'aide d'une étude de fonction, on constate que si $ad \leqslant 0$, il existe $\lambda > 0$ tel que $\mathrm{Tr}(\mathrm{A}_\lambda \mathrm{M} \mathrm{A}_\lambda) = |a\lambda^2 + d/\lambda^2| < 2$, ce qui est impossible car $\mathrm{A}_\lambda \mathrm{M} \mathrm{A}_\lambda \in \mathrm{G}$. Par conséquent $ad > 0$. Comme $\mathrm{G} = -\mathrm{G}$, on peut supposer $a > 0$ et $d > 0$.

En prenant $\lambda = (d/a)^{1/4}$, on prouve que

$$\begin{pmatrix} \sqrt{ad} & b \\ c & \sqrt{ad} \end{pmatrix} \in \mathrm{G}.$$

On peut donc supposer $a = d > 0$. On a alors $\mathrm{M} = \begin{pmatrix} a & b \\ c & a \end{pmatrix}$ et $a > 1$ car $\mathrm{Tr}\,\mathrm{M} > 2$. Comme $1 = \det \mathrm{M} = a^2 - bc$, $bc = a^2 - 1 > 0$.

Soit $\lambda > 0$. Calculons $\mathrm{M}' = \mathrm{M}^2 \mathrm{A}_\lambda \mathrm{M}^{-1}$ $(\lambda > 0)$. On a

$$\begin{aligned} \mathrm{M}' &= \begin{pmatrix} a^2+bc & 2ab \\ 2ac & a^2+bc \end{pmatrix} \mathrm{A}_\lambda \mathrm{M}^{-1} \\ &= \begin{pmatrix} 2a^2-1 & 2ab \\ 2ac & 2a^2-1 \end{pmatrix} \begin{pmatrix} \lambda & 0 \\ 0 & 1/\lambda \end{pmatrix} \begin{pmatrix} a & -b \\ -c & a \end{pmatrix} \\ &= \begin{pmatrix} \times & -b(2a^2-1)\lambda + 2a^2 b/\lambda \\ 2a^2 c\lambda - c(2a^2-1)/\lambda & \times \end{pmatrix} \\ &= \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \mathrm{G}. \end{aligned}$$

Pour que $c' = 0$, il suffit que $\lambda^2 = 1 - \dfrac{1}{2a^2} > 1$. Pour un tel choix de λ, on a $b' \neq 0$, car $b' = 0$ équivaut à $\lambda = \dfrac{2a^2}{2a^2-1} > 1$. Or nous avons déjà démontré un peu plus haut que si $\mathrm{M}' = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix} \in \mathrm{G}$ avec $b' \neq 0$ alors, nécessairement $c' \neq 0$: voilà la contradiction attendue.

Il ne peut donc pas exister de suite (A_n) de G_0 qui converge vers I_n. En considérant la suite opposée, on montre de même qu'une suite de G_0 ne peut converger vers I_n. Si (A_n) est une suite de G_0 qui converge, sa limite appartient à G, car G est fermé et d'après ce qui précède, elle appartient à G_0. Ainsi G_0 est fermé. $\lhd$

Index

A

adjointe d'une application linéaire, 136–140
algèbre de Lie, 255
Auerbach (base d'), 12
Auerbach (théorème d'), 10

B

Bezout
  relation de, 91
  théorème de, 105, 165
Burnside (théorème de), 185

C

Cauchy (déterminant de), 29
Cauchy-Binet (formule de), 47
Cayley-Hamilton (théorème de), 80, 112, 121, 134, 149, 151, 161, 165, 229, 237
classes de similitude (topologie), 235–238
comatrice, 37–47
commutant, 159, 160
commutateur dans un groupe, 188, 189
connexité, connexité par arcs, 238–242
Cramer (système de), 5, 93, 111, 155

D

décomposition LU, 61–68
dérangements, 55
diagonalisation, 92–110
discriminant d'un polynôme, 26
Dunford (décomposition de), 70, 134, 135, 137, 187, 247, 251

E

endomorphisme
  cyclique, 147–156
  nilpotent, 95, 134–144, 163
  semi-simple, 142, 144
  simple, 151
espace de suites (opérateur dans), 72–75, 122–127
Euler (indicatrice d'), 56
exponentielle (de matrice), 242–260
exposant d'un groupe, 185

F

Faddeev (algorithme de), 78
Fermat (petit théorème de), 35
forme n-linéaire alternée, 5–9
Frobenius (théorème de), 156

G

Gauss (algorithme du pivot de), 50, 61, 177
Gershgörin (disques de), 80
Gram (matrice de), 29
graphe, 109
groupe dérivé, 188

H

Hadamard (lemme d'), 80
Hoffman-Singleton (théorème de), 107
Hürwitz (déterminant de), 32

K

Kronecker (produit de), 106

L

Lie
  algèbre de, 255
  crochet de, 138, 163
  formules de, 253

M

matrice
  à coefficients entiers, 54–61
  circulante, 34, 99–103
  cyclique, 155, 220
  d'incidence, 54
  de dilatation, 177
  de permutation, 34, 36, 65
  de transvection, 177
  irréductible, 87
  nilpotente, 116, 121, 135–144, 185
  par blocs, 48–54
  stochastique, 35, 82–87
  tridiagonale, 65
matrices anticommutantes, 130
Moore (graphe de), 110

N

Newton (sommes de), 78, 111
noyaux (lemme des), 96, 112, 120, 128, 132, 134, 148, 157, 170, 184

O

opérateur intégral (réduction), 71

P

Perron-Frobenius (théorème de), 87
polynôme
  annulateur, 90
  caractéristique, 76, 78, 80, 111, 112, 114, 120, 135
  d'endomorphisme, 91
  minimal, 105, 117, 128, 147, 152, 156, 160, 174, 181, 184, 199, 236
  minimal ponctuel, 148, 154–156

R

rayon spectral, 226
Rolle (théorème de), 19

S

Smith (déterminant de), 56
sous-espace
  caractéristique, 134
  stable, 141–147, 151, 152
Sylvester (équation de), 164 169

T

trigonalisation, 110–135

V

Vandermonde (déterminant de), 16–30, 111
Vandermonde (matrice de), 221

W

Williamson (formule de), 49

Achevé d'imprimer sur les presses de l'Imprimerie BARNÉOUD
53960 BONCHAMP-LÈS-LAVAL
Dépôt légal : novembre 2009 - N° d'imprimeur : 910086
Imprimé en France

Ce livre est le second tome d'algèbre du recueil d'exercices résolus de S. Francinou, H. Gianella et S. Nicolas.

Comme dans les volumes précédents, les auteurs ont tenu à présenter les solutions les plus pédagogiques possible, essayant « d'exposer clairement les idées et démarches de raisonnement, sans pour autant escamoter les détails ou calculs qui peuvent paraître évidents. »

Le premier chapitre de ce tome est consacré au déterminant et peut être abordé dès la première année en classe préparatoire. Le second chapitre sur la réduction des endomorphismes constitue le cœur du programme d'algèbre de seconde année, et il est le plus riche des quatre. Le troisième chapitre, aux exercices plus difficiles, est dédié à l'étude du groupe linéaire. Enfin, un nouveau chapitre, consacré à la topologie des ensembles de matrices et à l'exponentielle de matrice a été ajouté dans cette deuxième édition. Dans chacun de ces quatre chapitres la difficulté est en général croissante : on commence par des questions techniques ou des savoir-faire indispensables (calculs de déterminants, recherche de valeurs propres, générateurs du groupe linéaire, questions d'adhérence et d'intérieur...) et on termine par des exercices plus théoriques.

Aux concours de l'X et des ENS, les énoncés proposés aux candidats sont souvent des résultats intéressants par eux-mêmes. Les auteurs ont fait l'effort de les identifier et de les présenter, autant que possible, dans leur contexte naturel. C'est ainsi qu'à l'intérieur d'un chapitre les exercices sont regroupés par thèmes et insérés dans un texte de présentation qui tour à tour dégage des idées générales, apporte des prolongements ou effectue quelques rappels de cours.

Collection enseignement des mathématiques

ISBN 978-2-84225-142-0

Graphisme : Massin

exercices de mathématiques

oraux x-ens

algèbre 3

Serge Francinou
Hervé Gianella
Serge Nicolas

CASSINI

EXERCICES DE MATHÉMATIQUES
ORAUX X-ENS

Enseignement des mathématiques

1. J.-Y. Ouvrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouvrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter. *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron. *Analyse complexe*
19. B. Randé. *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*

SERGE FRANCINOU
HERVÉ GIANELLA
SERGE NICOLAS

Exercices de mathématiques des oraux de l'École polytechnique et des Écoles normales supérieures

Algèbre. Tome III

Deuxième édition, revue et augmentée

CASSINI

SERGE FRANCINOU, ancien élève de l'École normale supérieure et agrégé de Mathématiques, est actuellement professeur en classe préparatoire au lycée Marcelin Berthelot.

HERVÉ GIANELLA, ancien élève de l'École normale supérieure et agrégé de Mathématiques, est actuellement professeur en classe préparatoire au lycée Saint-Louis.

SERGE NICOLAS, ancien élève de l'École normale supérieure et agrégé de Mathématiques, est actuellement professeur en classe préparatoire au lycée Henri IV.

ISBN 978-2-84225-092-8

© Cassini. Paris, 2008.

Table des matières

Introduction 1

Chapitre 1. Espaces euclidiens. Espaces hermitiens 5

1.1. Bornes d'une fonction où intervient le produit scalaire 5
1.2. Boule contenant n points à distances mutuelles supérieures à 2 7
1.3. Familles obtusangles 8
1.4. Caractérisation des normes euclidiennes 10
1.5. Norme d'un endomorphisme symétrique 13
1.6. Fonctions additives sur deux vecteurs orthogonaux 15
1.7. Applications de $\mathbb{Z}^2$ dans $\mathbb{R}$ additives sur les vecteurs orthogonaux 18
1.8. Forme linéaire sur un sous-espace de $\mathcal{L}(\mathrm{E})$ 20
1.9. Caractérisation des projecteurs orthogonaux (1) 21
1.10. Caractérisation des projecteurs orthogonaux (2) 22
1.11. Norme des projetés d'une base orthonormale 23
1.12. Pseudo-inverse 25
1.13. Condition pour que deux projecteurs orthogonaux commutent 26
1.14. Composition de projecteurs orthogonaux 27
1.15. Contraction d'un espace euclidien 28
1.16. Distance à un sous-espace de $\mathbb{R}[\mathrm{X}]$ 30
1.17. Norme euclidienne canonique sur $\mathcal{M}_n(\mathbb{R})$ 31
1.18. Distance d'une matrice à l'espace des matrices symétriques 33
1.19. Problèmes de minimisation dans $\mathcal{M}_{p,q}(\mathbb{R})$ 34
1.20. Simplexes réguliers 37
1.21. Décomposition QR et inégalité d'Hadamard 40
1.22. Base orthonormale dans $\mathbb{R}[\mathrm{X}]$ 42
1.23. Polynômes de Laguerre 45
1.24. Polynômes d'Hermite 48
1.25. Méthode de Gauss 51
1.26. Matrices de Gram 53
1.27. Familles équiangulaires 55
1.28. Familles isométriques 56
1.29. Image d'une base orthonormée par un projecteur orthogonal 57
1.30. Une matrice orthogonale 58
1.31. Convergence en moyenne des puissances d'une matrice orthogonale 59
1.32. Orbites sous l'action d'un endomorphisme orthogonal . . . 60
1.33. Générateurs de $\mathrm{O}(\mathrm{E})$ 61
1.34. Théorème de Maschke (1898) 62

1.35. Réduction des matrices orthogonales 63
1.36. Exponentielle de matrices antisymétriques réelles 65
1.37. Simplicité de SO_3 67
1.38. Polynômes de quatre variables invariants sous l'action de $O_2(\mathbb{R})$ 70
1.39. Équation fonctionnelle faisant intervenir le groupe orthogonal 72
1.40. Endomorphismes conservant le produit vectoriel 73
1.41. Le groupe des quaternions 74
1.42. Une équation fonctionnelle 76
1.43. Projection sur un convexe fermé 78
1.44. Lemme de Farkas 80
1.45. Éléments de $\mathbb{R}^n$ à composantes positives 82
1.46. Inégalités 84
1.47. Étude de normes sur $\mathcal{L}(E)$ où E est hermitien 86
1.48. Condition suffisante pour que deux matrices unitaires commutent 89

Chapitre 2. Réduction des endomorphismes auto-adjoints **91**

2.1. Codiagonalisation 92
2.2. Puissances d'une matrice symétrique 92
2.3. Méthode itérative pour une équation linéaire 93
2.4. Décomposition en somme de droites et plans stables 95
2.5. Spectre de la différence de deux projecteurs orthogonaux . . 95
2.6. Produit de deux projecteurs orthogonaux 96
2.7. Frame d'un espace euclidien 97
2.8. Matrices à termes diagonaux égaux 99
2.9. Caractérisation des matrices positives avec la trace 101
2.10. Étude d'un sous-ensemble convexe de $\mathbb{R}^n \times \mathcal{M}_n(\mathbb{R})$ 102
2.11. Caractérisation des symétries orthogonales 105
2.12. Racine carrée d'un endomorphisme auto-adjoint positif . . 107
2.13. Produit de deux matrices symétriques positives 108
2.14. Condition de diagonalisabilité d'une matrice réelle 109
2.15. Condition d'existence de m vecteurs propres indépendants . 110
2.16. Une équation matricielle 111
2.17. Équation dans $\mathcal{L}(E)$ 112
2.18. Caractérisation de Sylvester des matrices définies positives 114
2.19. Caractérisation des matrices positives 116
2.20. Matrice symétrique de taille 2 à coefficients positifs 117
2.21. Théorème de Perron-Frobenius pour les matrices symétriques 118
2.22. Théorème de Suleimanova (1949) 119
2.23. Matrice de passage à coefficients positifs 121
2.24. Décomposition en valeurs singulières 123
2.25. Dilatation isométrique d'une contraction 125

2.26. Décomposition polaire (1) 126
2.27. Décomposition polaire (2) 128
2.28. Points extrémaux de la boule unité de $\mathcal{L}(E)$ 130
2.29. Décomposition de Choleski. Décomposition QR 131
2.30. Inégalité d'Hadamard 133
2.31. Inégalité de Fischer 134
2.32. Recherche d'un minimum 136
2.33. Image de S^2 par une fonction, où S est la boule unité . . . 137
2.34. Majoration et minoration d'une fonction 138
2.35. Inégalité de Kantorovich 139
2.36. Inégalité de convexité 141
2.37. Théorème du minimax 142
2.38. Théorème d'entrelacement de Cauchy 144
2.39. Une réciproque au théorème d'entrelacement de Cauchy . . 146
2.40. Théorème de perturbation de Weyl 147
2.41. Théorème de majoration de Schur 148
2.42. Inégalités de Ky-Fan 150
2.43. Majoration de Tr(AB) 152
2.44. Écart maximal entre deux valeurs propres 153
2.45. Valeurs propres d'endomorphismes dont la somme est positive 154
2.46. Étude de $\mathrm{Tr}(AB - BA)^4$, pour A et B antisymétriques . . 156
2.47. Parité du rang d'un endomorphisme antisymétrique 156
2.48. Endomorphismes antisymétriques en dimension 3 157
2.49. Carré d'une matrice antisymétrique 158
2.50. Réduction des endomorphismes antisymétriques 159
2.51. Matrices strictement positives 161
2.52. Théorème de Liapounov 162
2.53. Condition pour que $\mathrm{Im}\,B \subset \mathrm{Im}\,A$ 164
2.54. Application convexe 167
2.55. Condition pour qu'une matrice hermitienne soit définie positive 168
2.56. Calcul de l'inverse d'une matrice par une méthode itérative 169
2.57. Encadrement des valeurs propres de AB 170
2.58. Inégalité d'Hadamard pour une matrice hermitienne 171
2.59. Produit de matrices hermitiennes positives 173
2.60. Produit de Schur de deux matrices hermitiennes positives . 175
2.61. Inégalités 176
2.62. Décomposition polaire dans $\mathcal{M}_n(\mathbb{C})$ 177
2.63. Écriture d'un endomorphisme comme combinaison linéaire d'endomorphismes unitaires 179
2.64. Matrices normales 181
2.65. Théorème de Mirman (1968) 182
2.66. Matrices unitairement congruentes à une matrice triangulaire 185

Chapitre 3. Formes quadratiques 189

3.1. Formes bilinéaires réflexives 189
3.2. Caractérisation des applications bilinéaires symétriques ou antisymétriques 190
3.3. Base formée de vecteurs isotropes 191
3.4. Formes quadratiques ayant même cône isotrope 192
3.5. Composantes connexes par arcs 193
3.6. Théorème de Pfister (1965) 195
3.7. Indice d'une forme quadratique 197
3.8. Famille de formes quadratiques 199
3.9. Théorème de Fischer-Cochran 200
3.10. Une forme quadratique définie positive 201
3.11. Une forme quadratique positive 202
3.12. Une forme quadratique définie négative 203
3.13. Matrices symétriques à diagonale positive 203
3.14. Restriction à un plan définie positive 205
3.15. Caractérisation de Sylvester des matrices définies positives 206
3.16. Comparaison de noyaux 207
3.17. Une forme quadratique positive 208
3.18. Condition pour qu'une matrice soit positive 208
3.19. Inégalité de Bergström 209
3.20. Convergence d'une suite croissante et majorée 211
3.21. Décroissance de la fonction inverse 211
3.22. Signature d'une forme quadratique réelle 213
3.23. Composantes connexes 214
3.24. Un calcul de signature (1) 216
3.25. Un calcul de signature (2) 216
3.26. Un calcul de signature (3) 217
3.27. Formes quadratiques sur un corps fini 217
3.28. Matrices symétriques positives telles que $A \leqslant B$ 220
3.29. Inégalité sur le déterminant de matrices symétriques . . . 220
3.30. Comparaison d'endomorphismes symétriques définis positifs 221
3.31. Convexité logarithmique 222
3.32. Inégalité de Minkowski 223
3.33. Inégalité sur le déterminant de matrices symétriques 224
3.34. Diagonalisation simultanée de deux formes quadratiques . . 225
3.35. Commutant du groupe orthogonal d'une forme quadratique non dégénérée 226
3.36. Minimalité du groupe orthogonal d'une forme non dégénérée 228
3.37. Ellipsoïde de John-Loewner 229
3.38. Sous-groupes compacts maximaux de GL(E) 231
3.39. Endomorphisme auto-adjoint sur un espace quadratique complexe 232

3.40. Réduction des matrices symétriques complexes 233
3.41. Une décomposition 235
3.42. Maximum d'une forme quadratique sur un compact 238
3.43. Forme quadratique et réseau 239
3.44. Lemme de Davenport-Cassels 241
3.45. Image d'un cône 242

Chapitre 4. Géométrie affine et euclidienne **245**

4.1. Partage équitable 245
4.2. Théorème de Pappus 246
4.3. Théorème de Ménélaüs 248
4.4. Quadrilatère complet 249
4.5. Théorème de Sylvester-Gallai 251
4.6. Aire d'un triangle 252
4.7. Point de Gergonne 254
4.8. Quadrilatère formé par les sommets et l'orthocentre d'un triangle 255
4.9. Caractérisation des triangles équilatéraux 256
4.10. Un triangle équilatéral 256
4.11. Théorème de Ptolémée 258
4.12. Puissance d'un point par rapport à un cercle 258
4.13. Suite de cercles 260
4.14. Condition de cocyclicité de quatre points 260
4.15. Problème angulaire 262
4.16. Intersection de deux disques 264
4.17. Ellipses semblables 265
4.18. Rayon de courbure d'une ellipse 266
4.19. Cercles tangents à une ellipse 268
4.20. Triangles d'aire maximale inscrits dans une ellipse 271
4.21. Rectangles d'aire maximale inscrits dans une ellipse 273
4.22. Problème de Pappus 274
4.23. Cocyclicité de quatre points d'une ellipse 275
4.24. Cercle orthoptique d'une ellipse 276
4.25. Sphère orthoptique d'un ellipsoïde 278
4.26. Centres des cercles circonscrits à une famille de triangles inscrits dans une parabole 280
4.27. Centres des triangles équilatéraux inscrits dans une parabole 281
4.28. Points équidistants de deux droites non coplanaires 283
4.29. Diamètre transfini 284
4.30. Deux problèmes d'extrema 286
4.31. Problème d'extremum 289
4.32. Triangle de périmètre minimal inscrit dans un triangle . . . 291
4.33. Billard convexe compact 293

4.34. Problème de Fermat 295
4.35. Polygone dont les milieux des côtés sont donnés 299
4.36. Triangles à côtés parallèles 300
4.37. Équivalence entre parties de $\mathbb{R}^2$ 302
4.38. Étude d'une suite d'applications affines 304
4.39. Applications du plan conservant l'orthogonalité 305
4.40. Sous-groupes finis du groupe affine réel 309
4.41. Une partie bornée de $\mathbb{R}^2$ n'est pas dédoublable 309
4.42. Application conservant la distance unité 310
4.43. Groupe des isométries du simplexe régulier 313
4.44. Théorème de Carathéodory (1907) 316
4.45. Théorème de Jung (1901) 317
4.46. Points extrémaux d'un ensemble convexe 320
4.47. Rectangle circonscrit à un convexe compact 322
4.48. Théorème de Pick (1899) 326
4.49. Droites à une distance strictement positive de $\mathbb{Z}^2$ 328
4.50. Polygone ayant trois sommets à coordonnées entières 329
4.51. Recouvrement par des droites 331
4.52. Rotations de $\mathbb{R}^3$ laissant un réseau invariant 333
4.53. Un problème de Pólya (1918) 334
4.54. Théorème de Minkowski 336
4.55. Formule d'Euler 338
4.56. Rectangles semi-entiers 340
4.57. Pavage d'un rectangle par des carrés 341

Index **351**

Introduction

Cet ouvrage est le troisième tome d'algèbre d'un recueil d'exercices de mathématiques destiné à la préparation des oraux des concours d'entrée aux Écoles normales supérieures et à l'École polytechnique. Trois volumes sont consacrés à l'algèbre et trois à l'analyse.

La vocation première des Écoles normales est de former des chercheurs ou des enseignants-chercheurs. Le concours d'entrée vise donc à détecter les qualités scientifiques du candidat, son aptitude à la recherche. À l'oral, on jugera avant tout la capacité de prendre des initiatives, d'utiliser une indication, de mener à bien une démarche. On ne sera pas surpris que les exercices posés aient un contenu mathématique riche, qu'ils soient très éloignés du simple exercice technique, d'application du cours, qu'ils soient souvent difficiles. Ils visent la plupart du temps à la démonstration d'un résultat mathématique significatif. Ils pourraient apparaître excessivement difficiles, si on perdait de vue le déroulement concret de l'épreuve. L'oral des ENS est un long dialogue (l'épreuve dure environ cinquante minutes, comme d'ailleurs à l'École polytechnique) entre le candidat et l'examinateur, qui tout au long de l'épreuve fournit des indications, quand c'est nécessaire, pour relancer la réflexion du candidat et tester ses réactions. Il est d'ailleurs impossible de rendre pleinement compte dans un recueil d'exercices du caractère oral de l'épreuve.

L'École polytechnique, quant à elle, est plus généraliste. Les exercices posés au concours sont de facture plus classique et, en règle générale, l'examinateur intervient moins. C'est au candidat de montrer sa maîtrise du programme dans la résolution d'un exercice dont la difficulté est cependant très variable. Certains sont proches des exercices d'ENS. Les énoncés circulent d'ailleurs d'un concours à l'autre, ou peuvent même être repris d'exercices d'Olympiades.

Les énoncés qui figurent dans ce recueil, ont été donnés entre 1996 et 2006. Ils sont extraits pour l'essentiel des listes publiées chaque année par la RMS (*Revue des mathématiques de l'enseignement supérieur* aux éditions Vuibert jusqu'en 2003 et désormais *Revue de la filière Mathématiques* aux éditions e.net) dont nous remercions les auteurs pour l'aide précieuse qu'ils apportent ainsi aux élèves et aux professeurs des classes préparatoires. Il s'agit de versions communiquées par les étudiants, reflétant la compréhension que ceux-ci ont eue de l'exercice et le déroulement conjoncturel de leur oral, comme le montrent les variations d'une année à l'autre pour un même exercice. Nous n'avons pas hésité à les modifier, pour rectifier des erreurs, compléter un énoncé

quand manifestement l'exercice s'est arrêté avant que le résultat que l'examinateur avait en vue ne soit atteint, ou ajouter des indications.

Nous avons choisi de laisser quelques énoncés «bruts», ceux pour lesquels nous estimons qu'une démarche naturelle (qui peut être longue et ardue) permet de conduire à la solution. Pour d'autres exercices, nous avons pris la liberté de rajouter des questions intermédiaires, qui auraient pu être celles posées par l'examinateur. Quitte à perdre en concision, nous avons tenu à rédiger les solutions les plus pédagogiques possible, essayant d'exposer clairement les idées et démarches de raisonnement, sans escamoter les détails ou calculs qui peuvent paraître évidents. On évite autant que possible l'introduction d'une astuce ou d'un objet *ad hoc* permettant d'atteindre rapidement la solution. S'il n'y a pas moyen d'expliquer l'origine de cette astuce, c'est que l'exercice est peu intéressant et que l'étudiant en tirera peu de profit.

À l'intérieur de chaque chapitre, les exercices ont été regroupés thématiquement et à l'intérieur de chaque thème, souvent par ordre de difficulté croissante. Ainsi regroupés, ils apparaîtront plus accessibles, car plongés dans leur contexte mathématique, éclairés par d'autres exercices voisins. Les introductions historiques qui ouvrent chaque chapitre, outre leur intérêt propre, visent au même but. Enfin, nous avons agrémenté les énoncés de quelques remarques préliminaires. Sans faire de rappels de cours systématiques, nous avons énoncé, voire redémontré, certains résultats : lemmes classiques, intervenant dans la résolution d'un grand nombre d'exercices, ou résultats au contraire à la lisière du programme, mais utiles, pour lesquels des éclaircissements étaient nécessaires. On trouvera aussi des remarques de synthèse ou des généralisations qui, nous l'espérons, pourront amener le candidat curieux à approfondir ses connaissances. Les quelques indications bibliographiques ont le même objectif.

Le lecteur ne tirera profit de ce livre que s'il cherche des solutions personnelles des exercices avant d'en étudier les corrigés. Une bonne connaissance du cours est indispensable. En effet, les théorèmes du programme fournissent bon nombre de schémas de démonstration. Rappelons aussi quelques démarches générales qui peuvent faciliter l'appréhension des exercices difficiles :

▷ l'étude de cas particuliers permet souvent d'entrevoir les idées qu'il faudra mettre en œuvre dans la résolution du cas général. Ce peut être l'étude du problème pour les petites dimensions, dans un exercice de géométrie euclidienne, ...

▷ le renforcement des hypothèses peut aboutir à un problème plus simple : pour un exercice sur les matrices symétriques positives on peut commencer par le cas où la matrice est définie positive et chercher à

prolonger ensuite au cas général, par exemple à l'aide d'un passage à la limite.

▷ dans l'étude des problèmes géométriques on attachera une grande importance au choix des coordonnées (cartésiennes ou barycentriques) et on préservera le plus possible les symétries du problème.

Au-delà des étudiants en classe préparatoire, ces ouvrages intéresseront aussi les candidats au CAPES et à l'Agrégation, qui y trouveront matière à réviser les principales notions du programme, ainsi que des exemples pour nourrir un développement pour leur oral.

Ce troisième tome est consacré à la géométrie. Le premier chapitre concerne les aspects géométriques des espaces euclidiens et il pourra être abordé en grande partie dès la première année en classe préparatoire. Le second est consacré aux endomorphismes symétriques, thème très riche comme le montre le nombre d'exercices. Ces deux premiers chapitres se terminent à chaque fois par quelques exercices plus anciens consacrés aux espaces hermitiens. En effet, le programme des classes préparatoires a été nettement allégé sur ce point lors de la dernière réforme. Nous avons toutefois tenu à garder ces énoncés pour nos nombreux lecteurs agrégatifs et en espérant peut-être un retour en grâce futur des espaces hermitiens (cela est souvent souhaité dans les rapports de concours des Écoles normales supérieures). Le chapitre trois est consacré aux formes quadratiques, une section également assez réduite dans les programmes actuels. Nous avons rédigé les solutions en tenant compte des connaissances au programme (par exemple en ce qui concerne la signature) et indiqué clairement les quelques exercices qui dépassent les résultats exigibles des candidats. Enfin le dernier chapitre est dédié à la géométrie affine euclidienne dans toute sa richesse (géométrie du triangle, coniques, problèmes d'extrema, isométries, convexité,...). Nous avons mis un grand nombre de figures pour éclairer les solutions. Les énoncés sont tous formulés dans un espace réel, voire dans $\mathbb{R}^n$, ce qui n'amène pas vraiment une perte de généralité et simplifie souvent la rédaction des solutions. Nos lecteurs connaissant la théorie des espaces affines n'auront aucun mal à placer chaque énoncé dans le contexte naturel qui est le sien.

Si vous souhaitez nous contacter pour nous faire part de vos remarques, vous pouvez envoyer un courriel à l'adresse *fgn.cassini@free.fr*.

Chapitre 1

Espaces euclidiens. Espaces hermitiens

Nous ne pourrions commencer ce chapitre sans y évoquer l'important héritage que constitue la géométrie élémentaire. Développée depuis les Grecs, elle a laissé dans ce domaine des Mathématiques des notions fondamentales telles que distance, orthogonalité, angle... et le vocabulaire associé qui dépasse le cadre des espaces de dimension 2 et 3 pour s'appliquer à des espaces de dimension n, et même de dimension infinie dans le cas des espaces fonctionnels. On retiendra que la traduction d'un problème abstrait sur les espaces euclidiens en termes de géométrie classique, à l'aide d'une figure par exemple, est souvent une source puissante d'inspiration.

Bien évidemment, la conscience du produit scalaire sous-jacent à la structure euclidienne de $\mathbb{R}^n$ est très récente puisqu'elle s'appuie sur le concept de vecteurs apparu véritablement à la fin du XIX^e *siècle. C'est plutôt à travers l'étude des formes quadratiques, vues pendant des siècles comme polynômes homogènes de degré 2, que vont se faire les plus grandes avancées, façonnant les outils de l'algèbre linéaire moderne : déterminant, valeurs propres, rang, forme bilinéaire... Le point de vue analytique s'appuyant sur un système de coordonnées fut dominant depuis Descartes et Fermat jusqu'à ce que la théorie des espaces de Hilbert au* XX^e *siècle fasse faire un pas de plus dans l'abstraction, nous donnant la notion générale de forme sesquilinéaire sur un espace de dimension quelconque.*

La première série d'exercices porte sur des calculs algébriques de base avec la norme et le produit scalaire. Ce dernier sera presque toujours noté $(x, y) \longmapsto \langle x, y \rangle$.

1.1. Bornes d'une fonction où intervient le produit scalaire

Soit E un espace euclidien et $(a, b) \in \mathrm{E}^2$. Déterminer les bornes supérieures et inférieures sur $\mathrm{E}\backslash\{0\}$ de

$$f : x \longmapsto \frac{\langle x, a\rangle\langle x, b\rangle}{\|x\|^2}.$$

(École normale supérieure)

▷ **Solution.**

Si a ou b est nul, f est identiquement nulle et les bornes valent 0. Supposons a et b non nuls et considérons un plan P contenant a et b. On a $f(x) = f\left(\frac{x}{\|x\|}\right)$ et donc

$$\sup_{x\in E\setminus\{0\}} f(x) = \sup_{\|x\|=1} f(x)$$

et il en va de même pour la borne inférieure. On note p la projection orthogonale sur P. On a $\langle x, a\rangle = \langle p(x), a\rangle$ et l'égalité analogue pour b. Ainsi, pour x vecteur unitaire, on a $f(x) = \langle p(x), a\rangle\langle p(x), b\rangle$ et donc

$$\sup_{x\in E\setminus\{0\}} f(x) = \sup_{\substack{e\in P\\ 0<\|e\|\leqslant 1}} \langle e, a\rangle\langle e, b\rangle,$$

puisque $\|p(x)\| \leqslant \|x\|$ et que chaque $e \in$ P non nul de norme plus petite que 1 est l'image par p d'un vecteur unitaire. Il en va de même pour la borne inférieure.

Munissons P d'une orientation quelconque et notons θ l'angle orienté entre a et b. Pour e non nul dans P, on a donc en notant φ l'angle orienté entre a et e,

$$\begin{aligned}\langle e, a\rangle\langle e, b\rangle &= \|a\|\|e\|\cos\varphi\|b\|\|e\|\cos(\theta-\varphi)\\ &= \frac{\|a\|\|b\|\|e\|^2}{2}\left(\cos\theta+\cos(2\varphi-\theta)\right).\end{aligned}$$

Lorsque e parcourt le disque unité de P privé de 0, la quantité $\|e\|$ décrit $]0, 1]$ et φ décrit tous les réels modulo 2π. On en déduit que

$$\boxed{\sup_{x\in E\setminus\{0\}} f(x) = \frac{\|a\|\|b\|(\cos\theta+1)}{2} = \frac{\langle a, b\rangle + \|a\|\|b\|}{2}}$$

et

$$\boxed{\inf_{x\in E\setminus\{0\}} f(x) = \frac{\|a\|\|b\|(\cos\theta-1)}{2} = \frac{\langle a, b\rangle - \|a\|\|b\|}{2}}. \quad \triangleleft$$

L'énoncé suivant peut s'interpréter de la manière suivante : étant donné n boules ouvertes de rayon 1 deux à deux disjointes dans un espace euclidien, que dire du rayon d'une boule qui les contient toutes ?

1.2. Boule contenant n points à distances mutuelles supérieures à 2

Soit E un espace euclidien et $(x_1, \ldots, x_n) \in \mathrm{E}^n$, vérifiant :

$$\forall (i,j) \in [\![1,n]\!]^2, \quad i \neq j \Longrightarrow \|x_i - x_j\| \geqslant 2.$$

Soit B une boule fermée de rayon R contenant tous les x_i. Montrer que

$$\mathrm{R} \geqslant \sqrt{\frac{2(n-1)}{n}}.$$

(École polytechnique)

▷ **Solution.**

En remplaçant éventuellement les vecteurs $x_1, \ldots, x_n$ par $x_1 - x_0, \ldots, x_n - x_0$, où x_0 est le centre de la boule B, on peut supposer que celui-ci est 0. On a alors $\|x_i\| \leqslant \mathrm{R}$ pour tout $1 \leqslant i \leqslant n$.

Pour minorer R, il nous faut tenir compte de toutes les conditions d'écartement entre les x_i. La quantité la plus simple qui tienne compte de tous les écarts est

$$\mathrm{S} = \sum_{1 \leqslant i,j \leqslant n} \|x_i - x_j\|^2.$$

Nous prenons les normes au carré sinon il serait impossible de simplifier la somme. Développons les carrés scalaires de S :

$$\mathrm{S} = \sum_{1 \leqslant i,j \leqslant n} \|x_i\|^2 + \|x_j\|^2 - 2\langle x_i, x_j \rangle,$$

ce qui donne

$$\mathrm{S} = 2n \sum_{i=1}^{n} \|x_i\|^2 - 2\left\| \sum_{i=1}^{n} x_i \right\|^2 \leqslant 2n \sum_{i=1}^{n} \|x_i\|^2.$$

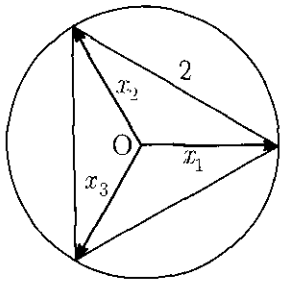
On a donc $\mathrm{S} \leqslant 2n^2\mathrm{R}^2$. Mais par hypothèse, on a

$$\mathrm{S} \geqslant \sum_{i \neq j} 4 = 4(n^2 - n) = 4n(n-1).$$

On en déduit que $2n\mathrm{R}^2 \geqslant 4n(n-1)$ ce qui conduit bien à

$$\boxed{\mathrm{R} \geqslant \sqrt{\frac{2(n-1)}{n}}}. \quad \triangleleft$$

Regardons si la minoration obtenue est optimale. Pour placer n points vérifiant $\|x_i - x_j\| \geqslant 2$ (pour $i \neq j$) dans une boule de centre 0 et de rayon $\mathrm{R}_n = \sqrt{\dfrac{2(n-1)}{n}}$ il est nécessaire, d'après la solution de l'exercice, que tous les x_i aient pour norme R_n et que $\|x_i - x_j\| = 2$ pour tout couple (i,j) avec $i \neq j$. Les vecteurs $x_1, \ldots, x_n$ forment alors ce que l'on appelle un n-simplexe régulier. C'est un objet qui n'existe que si la dimension de E *est supérieure ou égale à $n-1$ (on pourra notamment se reporter aux exercices 1.20 et 1.27). Par exemple pour former un triangle équilatéral (cas $n = 3$) il faut au moins que* $\dim \mathrm{E} \geqslant 2$. *Inversement, si par exemple* $\dim \mathrm{E} = 2$, *et si les vecteurs x_1, x_2, x_3 forment un triangle équilatéral de côté 2, il est facile de vérifier que le rayon du cercle circonscrit à ce triangle équilatéral est* $\mathrm{R}_0 = \dfrac{2}{\sqrt{3}}$.



Le lecteur pourra montrer plus généralement que le rayon de la sphère circonscrite à un n-simplexe régulier de côté 2 est bien R_n. *Autrement dit, la minoration de l'exercice est optimale dès que la dimension de l'espace euclidien dépasse $n-1$.*

L'énoncé suivant est très classique et prouve que l'on peut construire en dimension n des familles de vecteurs de cardinal $n+1$ faisant entre eux un angle obtus.

1.3. Familles obtusangles

Soit $(u_1, \ldots, u_p)$ une famille de vecteurs de $\mathbb{R}^n$ vérifiant $\langle u_i, u_j \rangle < 0$ pour $i \neq j$.

1. Montrer que $p-1$ vecteurs parmi eux forment toujours une famille libre de $\mathbb{R}^n$.

2. Montrer que l'on ne peut trouver plus de $n+1$ vecteurs réunissant ces conditions.

3. Montrer que l'on peut en trouver $n+1$.
(École polytechnique)

Solution.

1. Par symétrie on peut se contenter de montrer que la famille $(u_1, \dots, u_{p-1})$ est libre. Soit $(\lambda_1, \dots, \lambda_{p-1}) \in \mathbb{R}^{p-1}$ tel que $\sum_{i=1}^{p-1} \lambda_i u_i = 0$. On sépare dans cette somme les coefficients positifs et négatifs. On pose $\mathrm{I} = \{i \in [\![1, p-1]\!], \lambda_i > 0\}$ et $\mathrm{J} = [\![1, p-1]\!] \setminus \mathrm{I}$, ces ensembles pouvant être vides. On a alors $\sum_{i \in \mathrm{I}} \lambda_i u_i = -\sum_{j \in \mathrm{J}} \lambda_j u_j$ et donc

$$\left\| \sum_{i \in \mathrm{I}} \lambda_i u_i \right\|^2 = -\left\langle \sum_{i \in \mathrm{I}} \lambda_i u_i, \sum_{j \in \mathrm{J}} \lambda_j u_j \right\rangle = -\sum_{i \in \mathrm{I}, j \in \mathrm{J}} \lambda_i \lambda_j \langle u_i, u_j \rangle \leqslant 0.$$

On en déduit que $\sum_{i \in \mathrm{I}} \lambda_i u_i = \sum_{j \in \mathrm{J}} \lambda_j u_j = 0$. On obtient ensuite, en multipliant scalairement par u_p,

$$0 = \left\langle \sum_{i \in \mathrm{I}} \lambda_i u_i, u_p \right\rangle = \sum_{i \in \mathrm{I}} \lambda_i \langle u_i, u_p \rangle.$$

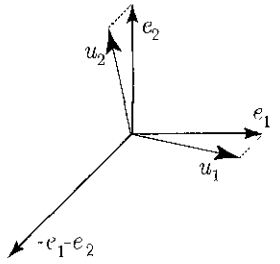
Or tous les termes de la somme sont négatifs. On en déduit que, pour tout $i \in \mathrm{I}$, $\lambda_i \langle u_i, u_p \rangle = 0$ et donc $\lambda_i = 0$, puisque $\langle u_i, u_p \rangle < 0$. On montre de même que $\lambda_j = 0$ pour tout $j \in \mathrm{J}$. La famille $(u_1, \dots, u_{p-1})$ est donc libre.

2. Si l'on peut trouver p vecteurs réunissant ces conditions, la question 1 montre que $p-1$ d'entre eux forment une famille libre et ainsi $p - 1 \leqslant n$, soit $p \leqslant n+1$. On ne peut pas trouver plus de $n+1$ vecteurs réunissant ces conditions.

3. Soit $(e_1, \dots, e_n)$ une base orthonormée de $\mathbb{R}^n$ et λ un réel. On pose $u_{n+1} = -\sum_{i=1}^{n} e_i$. Les produits scalaires $\langle e_k, u_{n+1} \rangle$ valent tous -1 et pour $i \neq j$ les produits scalaires $\langle e_i, e_j \rangle$ sont nuls. On va un peu déformer les vecteurs e_k en leur ajoutant un multiple de u_{n+1} pour obtenir des produits strictement négatifs (voir la figure ci-après en dimension 2).

Pour tout $i \in [\![1, n]\!]$, on pose $u_i = e_i + \lambda u_{n+1}$ où λ est un réel à choisir. On a, pour $1 \leqslant i < j \leqslant n$,

$$\langle u_i, u_j \rangle = \lambda^2 \|u_{n+1}\|^2 + \lambda \langle u_{n+1}, e_i \rangle + \lambda \langle u_{n+1}, e_j \rangle = n\lambda^2 - 2\lambda = \lambda(n\lambda - 2)$$



et

$$\langle u_i, u_{n+1}\rangle = \lambda\|u_{n+1}\|^2 + \langle e_i, u_{n+1}\rangle = \lambda n - 1.$$

Il suffit donc de prendre $0 < \lambda < \dfrac{1}{n}$ pour que la famille $\langle u_i\rangle_{1\leqslant i\leqslant n+1}$ convienne. $\lhd$

L'exercice suivant, qui regroupe des questions posées à plusieurs oraux, donne différentes caractérisations des normes euclidiennes. L'une des plus importantes est l'identité du parallélogramme : si x, y sont deux vecteurs d'un espace euclidien E *on a facilement*

$$\|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$$

en développant les carrés scalaires de gauche.

1.4. Caractérisation des normes euclidiennes

Soit $(\mathrm{E}, \|\ \|)$ un espace vectoriel normé réel. Montrer l'équivalence des propriétés suivantes :

(i) la norme $\|\ \|$ est euclidienne *i.e.* provient d'un produit scalaire ;

(ii) $\forall (x,y) \in \mathrm{E}^2,\ \|x+y\|^2 + \|x-y\|^2 = 2(\|x\|^2 + \|y\|^2)$;

(iii) pour tout $(x,y) \in \mathrm{E}^2$, l'application $t \longmapsto \|x+ty\|^2$ est un polynôme de degré inférieur ou égal à 2 ;

(iv) la restriction de $\|\ \|$ à tout plan P inclus dans E est euclidienne :

(v) $\forall x \in \mathrm{E},\ \mathrm{H}(x) = \{y \in \mathrm{E},\ \|x+y\|^2 = \|x\|^2 + \|y\|^2\}$ est un ensemble stable par homothétie.

(École polytechnique, École normale supérieure)

$\triangleright$ **Solution.**

Notons tout d'abord que le point (i) implique facilement tous les autres. Le point (ii) est l'identité du parallélogramme vérifiée par toute norme euclidienne. L'implication $(i) \Longrightarrow (iii)$ découle du fait que si $\|\ \|$

provient du produit scalaire $\langle\ ,\ \rangle$, alors la quantité $\|x+ty\|^2$ se développe en $\|x\|^2 + 2t\langle x, y\rangle + t^2\|y\|^2$. L'implication $(i) \Longrightarrow (iv)$ est triviale. Enfin $(i) \Longrightarrow (v)$ résulte du fait que $\mathrm{H}(x)$ n'est autre que le sous-espace $x^\perp$. On s'intéresse donc surtout aux réciproques.

• Montrons tout d'abord que $(ii) \Longrightarrow (i)$. On va voir que c'est l'implication la plus importante, car il est assez facile d'obtenir l'identité du parallélogramme à partir des conditions (iii), (iv) et (v).

Si la norme $\|\ \|$ provient d'un produit scalaire, celui-ci est nécessairement défini par

$$\forall (x, y) \in \mathrm{E}^2, \quad \langle x, y\rangle = \left\|\frac{x+y}{2}\right\|^2 - \left\|\frac{x-y}{2}\right\|^2.$$

On définit donc $\langle\ ,\ \rangle$ par l'expression ci-dessus et nous allons montrer qu'il s'agit d'un produit scalaire sur E. On se donne x, x', y des vecteurs de E.

- Il est déjà clair que $\langle x, y\rangle = \langle y, x\rangle$.

- On va montrer que $\langle\ ,\ \rangle$ est bilinéaire. La symétrie permet de se limiter à la linéarité à gauche. On va commencer par regarder $\langle 2x, y\rangle$. Par définition on a

$$\langle 2x, y\rangle = \left\|x + \frac{y}{2}\right\|^2 - \left\|x - \frac{y}{2}\right\|^2.$$

Or, avec l'hypothèse (ii), on peut écrire

$$\left\|x + \frac{y}{2}\right\|^2 = \left\|\frac{x}{2} + \frac{x+y}{2}\right\|^2 = 2\left\|\frac{x}{2}\right\|^2 + 2\left\|\frac{x+y}{2}\right\|^2 - \left\|\frac{y}{2}\right\|^2$$

et

$$\left\|x - \frac{y}{2}\right\|^2 = \left\|\frac{x}{2} + \frac{x-y}{2}\right\|^2 = 2\left\|\frac{x}{2}\right\|^2 + 2\left\|\frac{x-y}{2}\right\|^2 - \left\|\frac{y}{2}\right\|^2.$$

En soustrayant les deux expressions, il vient $\langle 2x, y\rangle = 2\langle x, y\rangle$.

- Calculons maintenant $\langle x, y\rangle + \langle x', y\rangle$:

$$\langle x, y\rangle + \langle x', y\rangle = \left\|\frac{x+y}{2}\right\|^2 - \left\|\frac{x-y}{2}\right\|^2 + \left\|\frac{x'+y}{2}\right\|^2 - \left\|\frac{x'-y}{2}\right\|^2$$

ce qui donne avec l'hypothèse,

$$\frac{1}{2}\left(\left\|\frac{x+x'+2y}{2}\right\|^2 + \left\|\frac{x-x'}{2}\right\|^2 - \left\|\frac{x+x'-2y}{2}\right\|^2 - \left\|\frac{x-x'}{2}\right\|^2\right),$$

et par conséquent $\langle x, y\rangle + \langle x', y\rangle = \frac{1}{2}\langle x + x', 2y\rangle = \langle x + x', y\rangle$, le point précédent étant utilisé dans la dernière égalité.

- Posons alors $f(t) = \langle tx, y\rangle - t\langle x, y\rangle$ pour tout réel t. D'après ce qui précède on a pour tout $(t, t') \in \mathbb{R}^2$, $f(t + t') = f(t) + f(t')$ et $f(0) =$

$f(1) = f(2) = 0$. De plus, f est continue (la norme est continue). Il est alors classique de montrer que f est une homothétie. Comme $f(1) = 0$ on a donc $f = 0$.

Ainsi, $\langle\ ,\ \rangle$ est une forme bilinéaire symétrique. Comme $\langle x, x\rangle = \|x\|^2$, il s'agit bien d'un produit scalaire dont la norme associée est $\|\ \|$.

Les assertions (i) et (ii) sont donc équivalentes : les normes vérifiant l'identité du parallélogramme sont les normes euclidiennes (ce résultat constitue le théorème de Fréchet-Von Neumann-Jordan).

• Montrons que $(iii) \Longrightarrow (ii)$. Soit $(x, y) \in \mathrm{E}^2$ avec y non nul. Par hypothèse, on sait que $\mathrm{Q}(t) = \|x + ty\|^2 + \|x - ty\|^2$ est un polynôme de degré inférieur à 2. Son terme constant est $\mathrm{Q}(0) = 2\|x\|^2$. Comme Q est pair, il n'a pas de terme en t. Enfin, comme y est non nul on a $\mathrm{Q}(t) \sim 2\|y\|^2 t^2$ en $+\infty$ et le coefficient dominant de Q vaut $2\|y\|^2$. Ainsi, on a $\mathrm{Q}(t) = 2\|x\|^2 + 2t^2\|y\|^2$ pour tout réel t. On obtient l'identité du parallélogramme pour x et y en prenant $t = 1$. Bien entendu, cette identité reste trivialement vérifiée lorsque y est nul.

• Il est tout à fait évident que $(iv) \Longrightarrow (ii)$.

• Pour finir, montrons que $(v) \Longrightarrow (ii)$. Considérons à nouveau deux vecteurs x et y de E. On peut supposer la famille (x, y) libre, car sinon l'identité du parallélogramme est évidente. Notons P le plan $\mathrm{Vect}(x, y)$. Observons que si $y \in \mathrm{H}(x)$ alors $x \in \mathrm{H}(y)$. Si cela est réalisé, on dira (en anticipant le résultat !) que les deux vecteurs x et y sont orthogonaux. Par hypothèse, on a dans ce cas,

$$\|\lambda x + \mu y\|^2 = \lambda^2\|x\|^2 + \mu^2\|y\|^2$$

pour tout couple de réels (λ, μ). Autrement dit, λx et μy sont aussi orthogonaux. Pour avoir le résultat, il suffit en fait d'avoir une base (e_1, e_2) de P formée de deux vecteurs orthogonaux (que l'on peut supposer unitaires quitte à les diviser par leur norme). En effet, si l'on pose $x = x_1e_1 + x_2e_2$ et $y = y_1e_1 + y_2e_2$ on a alors

$$\|x + y\|^2 = \|(x_1 + y_1)e_1 + (x_2 + y_2)e_2\|^2 = (x_1 + y_1)^2 + (x_2 + y_2)^2$$

$$\|x - y\|^2 = \|(x_1 - y_1)e_1 + (x_2 - y_2)e_2\|^2 = (x_1 - y_1)^2 + (x_2 - y_2)^2$$

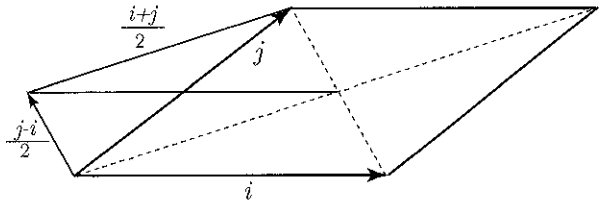
et aussi $\|x\|^2 = x_1^2 + x_2^2$ et $\|y\|^2 = y_1^2 + y_2^2$. L'identité du parallélogramme découle de ce que $(x_i + y_i)^2 + (x_i - y_i)^2 = 2(x_i^2 + y_i^2)$ pour $i = 1, 2$.

Notons B l'ensemble des bases de P et f la fonction qui à $(i, j) \in \mathrm{B}$ associe $f(i, j) = \|i+j\|^2 - \|i\|^2 - \|j\|^2$. Il nous faut démontrer que f s'annule en un couple (i, j) de B. On pense à un argument de type théorème des valeurs intermédiaires. L'ensemble B s'identifie à $\mathrm{GL}_2(\mathbb{R})$ dès lors que l'on choisit une base particulière de P. Mais cet ensemble n'est pas connexe. C'est en revanche le cas de B^+, l'ensemble des bases directes (la plan P étant orienté par le choix de la base particulière), qui s'identifie

lui au sous-groupe de $\mathrm{GL}_2(\mathbb{R})$ formé des matrices de déterminant > 0. Comme $f(j,i) = f(i,j)$, on peut se restreindre à B^+ sans problème. La fonction f est bien continue. Supposons qu'elle ne s'annule pas sur B^+. Elle y garde donc un signe constant, par exemple strictement positif. Pour toute base (i,j), on a alors $f(i,j) = \|i+j\|^2 - \|i\|^2 - \|j\|^2 > 0$ et aussi $f(i,-j) = \|i-j\|^2 - \|i\|^2 - \|j\|^2 > 0$. En sommant les deux, on a

$$g(i,j) = \|i+j\|^2 + \|i-j\|^2 - 2(\|i\|^2 + \|j\|^2) > 0.$$

Or il est impossible d'avoir cela pour toute base : en effet, si l'on prend la base $\left(\dfrac{i+j}{2}, \dfrac{j-i}{2}\right)$,



on voit que

$$g\left(\frac{i+j}{2}, \frac{j-i}{2}\right) = \|i\|^2 + \|j\|^2 - 2\left(\left\|\frac{i+j}{2}\right\|^2 + \left\|\frac{j-i}{2}\right\|^2\right) = -\frac{1}{2}g(i,j) < 0.$$

D'où l'existence d'une base orthogonale et le résultat. ◁

En constatant qu'une norme donnée ne vérifie pas l'identité du parallélogramme, on prouve que cette norme ne dérive pas d'un produit scalaire : c'est le cas pour la norme de la convergence uniforme sur l'espace $C([0,1],\mathbb{R})$. L'identité du parallélogramme tombe en défaut avec les fonctions $x \longmapsto x$ et $x \longmapsto 1-x$.

L'exercice suivant utilise une identité de polarisation.

1.5. Norme d'un endomorphisme symétrique

Soit H un espace préhilbertien réel muni d'un produit scalaire et T un endomorphisme continu de H vérifiant

$$\forall x \in \mathrm{H},\ \ \forall y \in \mathrm{H},\ \ (\mathrm{T}x, y) = (x, \mathrm{T}y).$$

Montrer que $\|\mathrm{T}\| = \sup_{\|x\|=1} |(\mathrm{T}x, x)|$.

(École polytechnique)

▷ **Solution.**

D'après l'inégalité de Cauchy-Schwarz et la définition de $\|\mathrm{T}\|$, on a pour tout vecteur x de norme 1

$$|\langle \mathrm{T}x, x\rangle| \leqslant \|\mathrm{T}x\| \leqslant \|\mathrm{T}\|.$$

On en déduit, en notant $\mathrm{K} = \sup_{\|x\|=1} |\langle \mathrm{T}x, x\rangle|$, que $\mathrm{K} \leqslant \|\mathrm{T}\|$.

Si x et y sont des vecteurs de norme **1**, on peut écrire

$$\langle \mathrm{T}(x+y), x+y\rangle - \langle \mathrm{T}(x-y), x-y\rangle = 2\langle \mathrm{T}x, y\rangle + 2\langle \mathrm{T}y, x\rangle = 4\langle \mathrm{T}x, y\rangle,$$

car T est symétrique. On en déduit que

$$\begin{aligned}\langle \mathrm{T}x, y\rangle &= \frac{1}{4}(\langle \mathrm{T}(x+y), x+y\rangle - \langle \mathrm{T}(x-y), x-y\rangle\rangle \\ &\leqslant \frac{1}{4}(\mathrm{K}\|x+y\|^2 + \mathrm{K}\|x-y\|^2) \leqslant \frac{1}{2}\mathrm{K}(\|x\|^2 + \|y\|^2\rangle \leqslant \mathrm{K}.\end{aligned}$$

Si $\mathrm{T}x \neq 0$, on obtient, en prenant $y = \dfrac{1}{\|\mathrm{T}x\|}\mathrm{T}x$, $\|\mathrm{T}x\| \leqslant \mathrm{K}$, inégalité qui reste vérifiée si $\mathrm{T}x = 0$. On en déduit que $\|\mathrm{T}\| = \sup_{\|x\|=1} \|\mathrm{T}x\| \leqslant \mathrm{K}$, puis l'égalité voulue. ◁

La propriété est en particulier vraie si H *est un espace euclidien et* T *un endomorphisme symétrique. Dans ce cas* $\|\mathrm{T}\|$ *est égal au rayon spectral* $\rho(\mathrm{T}) = \max_{\lambda \in \mathrm{Sp}\,\mathrm{T}} |\lambda|$. *En effet, si* $(e_1, \ldots, e_n)$ *est une base orthonormée de* H *formée de vecteurs propres de* T *on a, pour* $x = \sum_{i=1}^{n} x_i e_i$ *unitaire,*

$$|\langle \mathrm{T}x, x\rangle| = \left|\sum_{i=1}^{n} \lambda_i x_i^2\right| \leqslant \sum_{i=1}^{n} |\lambda_i| x_i^2 \leqslant \rho(\mathrm{T}),$$

où λ_i *est la valeur propre associée à* e_i. *On conclut en observant qu'il y a égalité pour* $x = e_j$ *avec* $|\lambda_j| = \rho(\mathrm{T})$.

Dans un espace muni d'un produit scalaire, le célèbre théorème de Pythagore affirme que les vecteurs x, y *sont orthogonaux si, et seulement si* $\|x+y\|^2 = \|x\|^2 + \|y\|^2$. *On relie ainsi métrique et orthogonalité. La version géométrique élémentaire était connue des Babyloniens. Les Grecs en avait la démonstration et de nombreuses conséquences (Livre I des* Éléments *d'Euclide).*

L'exercice suivant se propose de déterminer toutes les fonctions f *du plan euclidien* $\mathbb{R}^2$ *à valeurs réelles qui vérifient* $f(x+y) = f(x) + f(y)$ *lorsque* $x \perp y$, *avec une hypothèse de régularité en plus.*

1.6. Fonctions additives sur deux vecteurs orthogonaux

On munit $\mathbb{R}^2$ de sa structure euclidienne canonique.

1. Trouver toutes les applications $f : \mathbb{R}^2 \longrightarrow \mathbb{R}_+$ telles que pour tous vecteurs u et v orthogonaux, $f(u+v) = f(u)+f(v)$. On pourra supposer, pour commencer, que f est paire.

2. Trouver toutes les applications $f : \mathbb{R}^2 \longrightarrow \mathbb{R}$ continues telles que pour tous vecteurs u et v orthogonaux, $f(u+v) = f(u) + f(v)$.

(**École normale supérieure, École polytechnique**)

Solution.

1. • Comme le conseille l'énoncé, on suppose d'abord que f est paire. Si x et y sont deux vecteurs de même norme on constate que $\dfrac{x+y}{2}$ et $\dfrac{x-y}{2}$ sont orthogonaux et on a donc

$$f(x) = f\left(\frac{x+y}{2}\right) + f\left(\frac{x-y}{2}\right)$$

et

$$f(y) = f\left(\frac{x+y}{2}\right) + f\left(\frac{y-x}{2}\right).$$

On en déduit, puisque f est paire, que $f(x) = f(y)$. Ainsi, $f(x)$ ne dépend que de la norme de x. On choisit un vecteur quelconque e de norme 1 et on pose, pour tout $\lambda \in \mathbb{R}_+$, $\varphi(\lambda) = f(\lambda e)$. On a, pour tout $x \in \mathbb{R}^2$,

$$f(x) = f(\|x\|e) = \varphi(\|x\|).$$

On va traduire l'hypothèse en une équation fonctionnelle vérifiée par l'application φ. Soit λ et μ deux réels positifs, u et v deux vecteurs orthogonaux de $\mathbb{R}^2$, de norme respective λ et μ. Comme $u \perp v$, on a $\|u+v\| = \sqrt{\lambda^2+\mu^2}$. L'hypothèse $f(u+v) = f(u) + f(v)$, c'est-à-dire $\varphi(\|u+v\|) = \varphi(\|u\|) + \varphi(\|v\|)$, se traduit par

$$\varphi(\sqrt{\lambda^2+\mu^2}) = \varphi(\lambda) + \varphi(\mu).$$

En posant $\psi(\lambda) = \varphi\left(\sqrt{\lambda}\right)$ pour $\lambda \in \mathbb{R}_+$, on a donc $\psi(\lambda+\mu) = \psi(\lambda)+\psi(\mu)$ pour tout $(\lambda,\mu) \in \mathbb{R}_+^2$. En étendant ψ à $\mathbb{R}$ par imparité, on obtient un endomorphisme du groupe $(\mathbb{R},+)$. Il est bien connu qu'un tel morphisme est de la forme $x \longmapsto ax$ $(a \in \mathbb{R})$ pour peu que l'on ajoute l'hypothèse ψ continue ou ψ croissante[1]. Montrons que φ est croissante ; on en déduira la croissance de ψ. Soit $(\lambda,\mu) \in \mathbb{R}_+^2$, avec $\lambda \geqslant \mu$. On a alors

1. Si ψ est un morphisme de groupes de $\mathbb{R}$ dans $\mathbb{R}$ supposé croissant, on a pour tout

$$\varphi(\lambda) = \varphi(\mu) + \varphi\left(\sqrt{\lambda^2 - \mu^2}\right) \geqslant \varphi(\mu),$$

car φ, tout comme f, est à valeurs dans $\mathbb{R}_+$. Ainsi ψ est croissante et il existe $a \in \mathbb{R}$ tel que, pour tout $\lambda \in \mathbb{R}_+$, $\psi(\lambda) = a\lambda$. On obtient, pour tout $u \in \mathbb{R}^2$, $f(u) = \varphi(\|u\|) = \psi(\|u\|^2) = a\|u\|^2$. On note que $a \geqslant 0$, car f prend des valeurs positives.

• Si f n'est plus supposée paire, on décompose f en somme de sa partie paire g et de sa partie impaire h. Elles sont définies pour $u \in \mathbb{R}^2$ par

$$g(u) = \frac{f(u) + f(-u)}{2} \quad \text{et} \quad h(u) = \frac{f(u) - f(-u)}{2}.$$

Si u et v sont deux vecteurs orthogonaux, on a

$$\begin{aligned} g(u+v) &= \frac{1}{2}(f(u+v) + f(-u-v)) \\ &= \frac{1}{2}(f(u) + f(v) + f(-u) + f(-v)) \\ &= g(u) + g(v). \end{aligned}$$

car $-u$ et $-v$ sont orthogonaux. On peut faire la même démonstration pour h. Ainsi g et h vérifient la même hypothèse que f. De plus g, tout comme f, est à valeurs positives. D'après ce qui précède, il existe $a \geqslant 0$ tel que, pour tout $u \in \mathbb{R}^2$, $g(u) = a\|u\|^2$.

Étudions maintenant h. Si u et v sont deux vecteurs orthogonaux et de même norme, on a $h(u+v) = h(u) + h(v)$, $h(u-v) = h(u) + h(-v) = h(u) - h(v)$ et $h(u+v) + h(u-v) = h(2u)$, car $(u+v) \perp (u-v)$. On en déduit que $h(2u) = 2h(u)$ puis que, pour tout $n \in \mathbb{N}$, $h(u) = 2^n h\left(\frac{u}{2^n}\right)$. Sachant que $h = f - g$ et que $f \geqslant 0$, on en déduit que, pour tout $n \in \mathbb{N}$,

$$h(u) = 2^n h\left(\frac{u}{2^n}\right) \geqslant -2^n g\left(\frac{1}{2^n}u\right) \geqslant -2^n a \frac{1}{2^{2n}}\|u\|^2 = -\frac{a\|u\|^2}{2^n}.$$

En faisant tendre n vers $+\infty$, on obtient $h(u) \geqslant 0$. La fonction h est impaire et ne prend que des valeurs positives : c'est la fonction nulle. Ainsi, on a $f = g$.

On vérifie réciproquement que toute fonction $u \longmapsto a\|u\|^2$ avec $a \geqslant 0$ a les propriétés voulues.

2. On reprend le cheminement de la première question en supposant d'abord f paire. Les fonctions φ et ψ définies de manière analogue

---

$n \in \mathbb{Z}$ et $x \in \mathbb{R}$, $\psi(nx) = n\psi(x)$. Si $p \in \mathbb{Z}$ et $q \in \mathbb{Z}^*$, comme $q\psi\left(\frac{p}{q}x\right) = \psi\left(q\frac{p}{q}x\right) = \psi(px) = p\psi(x)$, on obtient pour tout rationnel r, $\psi(rx) = r\psi(x)$. Notons $a = \psi(1)$. En prenant $(q_n)_{n\in\mathbb{N}}$ (resp. $(q'_n)_{n\in\mathbb{N}}$) une suite croissante (resp. décroissante) de rationnels convergeant vers x, l'inégalité $\psi(q_n) = aq_n \leqslant \psi(x) \leqslant \psi(q'_n) = aq'_n$ donne, en passant à la limite, $\psi(x) = ax$.

héritent de la continuité de f. Comme $\psi : \mathbb{R} \longrightarrow \mathbb{R}$ est additive, c'est une application linéaire de la forme $x \longmapsto ax$ avec a un réel donné[2] et finalement, on a $f : x \longmapsto a\|x\|^2$.

Pour f quelconque, on décompose f en somme de g fonction paire et h fonction impaire. De même que précédemment, g et h sont additives sur les vecteurs orthogonaux. D'autre part, g et h sont continues et g est donc de la forme $x \longmapsto a\|x\|^2$ avec a réel.

Montrons que h est une forme linéaire. Comme précédemment, prenons u et v deux vecteurs orthogonaux et de même norme. On a $h(u+v) = h(u) + h(v)$, $h(u-v) = h(u) + h(-v) = h(u) - h(v)$ et $h(u+v) + h(u-v) = h(2u)$, car $(u+v) \perp (u-v)$. On en déduit que $h(2u) = 2h(u)$ puis que, pour tout $n \in \mathbb{N}$, $h(u) = 2^n h\left(\frac{u}{2^n}\right)$ et $h\left(\frac{u}{2^n}\right) = \frac{h(u)}{2^n}$. Montrons par récurrence sur $k \geqslant 0$ que $h(ku) = kh(u)$ et $h(kv) = kh(v)$. C'est vu pour $k = 1$ ou 2. Supposons $k \geqslant 3$. Alors $(k-1)u + v$ et $u - (k-1)v$ sont orthogonaux. Il s'ensuit par hypothèse de récurrence et par imparité de h

$$h((k-1)u + v) = h((k-1)u) + h(v) = (k-1)h(u) + h(v),$$

$$h(u - (k-1)v) = h(u) + h(-(k-1)v) = h(u) - (k-1)h(v),$$

et en sommant ces deux égalités, on trouve

$$h(ku - (k-2)v) = kh(u) - (k-2)h(v).$$

D'autre part, comme ku et $(k-2)v$ sont orthogonaux, il vient avec l'hypothèse de récurrence

$$h(ku - (k-2)v) = h(ku) - h((k-2)v) = h(ku) - (k-2)h(v).$$

On en déduit que $h(ku) = kh(u)$ et en échangeant les rôles de u et v, $h(kv) = kh(v)$. La récurrence est achevée. Par imparité, ces relations restent valables pour $k \in \mathbb{Z}$.

Compte-tenu de ce qui précède, on a pour tout vecteur u, tout $n \geqslant 0$ et tout $k \in \mathbb{Z}$,

$$h\left(\frac{k}{2^n}u\right) = \frac{h(ku)}{2^n} = \frac{k}{2^n}h(u).$$

Comme les nombres dyadiques $\frac{k}{2^n}$ sont denses dans $\mathbb{R}$, la continuité de h donne que pour tout vecteur u et tout $\lambda \in \mathbb{R}$, $h(\lambda u) = \lambda h(u)$. Il s'ensuit

[2] Comme dans la note précédente, on montre que pour tout $x \in \mathbb{R}$ et tout $q \in \mathbb{Q}$, $\psi(qx) = q\psi(x)$. On a alors pour tout $x \in \mathbb{R}$, $\psi(x) = ax$ avec $a = \psi(1)$. Il suffit de prendre une suite de rationnels convergente vers x et passer à la limite en utilisant la continuité de ψ.

que h est linéaire puisqu'en prenant u et v orthogonaux de norme 1, h coïncide avec la forme linéaire qui à un vecteur $x = \lambda u + \mu v$ associe $\lambda h(u) + \mu h(v)$.

Réciproquement, toute fonction de la forme $x \longmapsto a\|x\|^2 + h(x)$ où a est un réel quelconque et h une forme linéaire sur $\mathbb{R}^2$ est continue et additive sur les vecteurs orthogonaux. $\triangleleft$

Si l'on remplace $\mathbb{R}^2$ par $\mathbb{R}^n$ avec $n \geqslant 2$, on obtient les mêmes résultats avec la même démonstration, à ceci près que dans la dernière partie, on vérifie que h est une forme linéaire sur tout plan $\mathrm{Vect}(u, v)$ *de $\mathbb{R}^n$, ce qui permet de dire ensuite qu'il s'agit d'une forme linéaire sur $\mathbb{R}^n$.*

L'énoncé suivant étudie le même problème, mais sur $\mathbb{Z}^2$ et sans aucune hypothèse de régularité sur la fonction.

1.7. Applications de $\mathbb{Z}^2$ dans $\mathbb{R}$ additives sur les vecteurs orthogonaux

On munit le plan $\mathbb{R}^2$ de sa structure euclidienne canonique. Trouver toutes les applications $f : \mathbb{Z}^2 \longrightarrow \mathbb{R}$ telles que

$$u \perp v \Longrightarrow f(u+v) = f(u) + f(v).$$

(École normale supérieure)

$\triangleright$ **Solution.**

Notons pour commencer que l'ensemble E des applications vérifiant la propriété est clairement stable par combinaison linéaire : c'est donc un sous-espace vectoriel de $\mathcal{F}(\mathbb{Z}^2, \mathbb{R})$. L'ensemble $\mathbb{Z}^2$ est un groupe additif engendré par $e_1 = (1, 0)$ et $e_2 = (0, 1)$. On ne peut pas espérer qu'une application f de E soit déterminée par les images de ces deux seuls points, car on ne peut utiliser l'additivité qu'avec des couples de vecteurs orthogonaux. On va en fait montrer qu'une application f de E est parfaitement déterminée par les images des quatre points e_1, e_2, $-e_1$ et $-e_2$. Autrement dit, on va prouver que l'application linéaire $\psi : \mathrm{E} \longrightarrow \mathbb{R}^4$ qui à f associe $(f(e_1), f(e_2), f(-e_1), f(-e_2))$ est injective. Cela permettra donc de majorer la dimension de E par 4.

Soit $f \in \operatorname{Ker} \psi$. On montre que f est nulle. En prenant $u = v = (0, 0)$ on a $f(0, 0) = 0$. On a $(1, 1) = e_1 + e_2$ et la somme est orthogonale, donc $f(1, 1) = 0$. On voit de même que f est nulle en $(-1, 1)$, $(-1, -1)$ et $(1, -1)$. Regardons maintenant f sur le carré $\max(|x|, |y|) = 2$. On a $(2, 0) = (1, 1) + (1, -1)$ et la somme est orthogonale, donc $f(2, 0) = 0$. Par suite $f(2, 1) = f(2, 0) + f(0, 1) = 0$ et $f(2, -1) = f(2, 0) + f(0, -1) = 0$. On a de même $f(0, 2) = f(0, -2) = 0$ puis $f(2, 2) = f(2, -2) = 0$. Ainsi,

$f(2,x)=0$ pour tout $x\in[\![-2,2]\!]$. Par symétrie, f est nulle sur les trois autres côtés du carré.

Cela nous invite à poursuivre par récurrence. Supposons f nulle en tout $(x,y)\in\mathbb{Z}^2$ avec $\max(|x|,|y|)\leqslant n$ et montrons que f s'annule alors sur le carré $\max(|x|,|y|)=n+1$. On commence par obtenir un point de ce carré en lequel f s'annule : par exemple, $(n+1,n-1)=(n,-1)+(1,n)$ les deux vecteurs étant orthogonaux. Comme $(n+1,n-1)=(n+1,0)+(0,n-1)$ on a, avec l'hypothèse de récurrence

$$0=f(n+1,n-1)=f(n+1,0)+f(0,n-1)=f(n+1,0).$$

Ainsi, $f(n+1,0)$ est nul. On en déduit de suite que $f(n+1,k)=0$ pour tout $k\in[\![-n,n]\!]$. Par symétrie, on a $f(-n-1,k)=f(k,n+1)=f(k,-n-1)=0$ pour tout $k\in[\![-n,n]\!]$. On termine enfin pour les coins du carré en écrivant $(n+1,n+1)=(n+1,0)+(0,n+1)$. Cela termine la récurrence. Ainsi f est nulle et ψ est injective.

Essayons maintenant de trouver des fonctions linéairement indépendantes dans E. Il y a bien entendu les deux formes linéaires coordonnées $e_1^*:(x,y)\longmapsto x$ et $e_2^*:(x,y)\longmapsto y$ qui sont additives pour tout couple de $\mathbb{Z}^2$. On a aussi la norme euclidienne au carré, $\mathrm{N}:(x,y)\longmapsto x^2+y^2$, par le théorème de Pythagore. On a donc $3\leqslant\dim\mathrm{E}\leqslant 4$.

Peut-on exhiber une quatrième fonction de E qui forme une famille libre avec les trois ci-dessus ? Le commentaire qui suit l'exercice précédent montre que l'on ne peut pas espérer que celle-ci provienne de la restriction à $\mathbb{Z}^2$ d'une solution continue sur $\mathbb{R}^2$. Cela pourrait laisser penser que la dimension de E est 3 mais si l'on se donne f sur seulement trois des points utilisés ci-dessus, par exemple $\pm e_1$ et e_2 on ne peut déterminer f que sur très peu de points. Nous allons prouver que la dimension de E est bien égale à 4. L'image par ψ de la solution e_1^* (resp. e_2^* et N) est le vecteur $(1,0,-1,0)$ (resp. $(0,1,0,-1)$ et $(1,1,1,1)$). Pour préserver au plus la symétrie en x et en y nous allons chercher un antécédent par ψ au vecteur $(1,-1,1,-1)$ qui complète les trois vecteurs précédents en une base de $\mathbb{R}^4$.

Supposons qu'un tel antécédent f existe (f est alors unique d'après ce qui précède). En calculant les images par f des points (x,y) pour des petites valeurs de x et y, on constate que $f(x,y)=0$ lorsque x et y ont la même parité, que $f(x,y)=1$ lorsque x est impair et y pair et que $f(x,y)=-1$ lorsque x est pair et y impair. Autrement dit, on conjecture que $f(x,y)=\dfrac{1}{2}\left((-1)^y-(-1)^x\right)$ pour tout couple $(x,y)\in\mathbb{Z}^2$. Vérifions que cette fonction est bien dans E.

Soit (x,y) et (x',y') deux points de $\mathbb{Z}^2$ avec $xx'+yy'=0$. En particulier, les entiers xx' et yy' ont la même parité. Distinguons quatre cas :

• Supposons x pair et y pair. Alors $f(x,y)=0$ et $x+x'$ (resp. $y+y'$) a la même parité que x' (resp. y'). Ainsi, $f(x+x',y+y')=f(x',y')=f(x',y')+f(x,y)$.

• Supposons x pair et y impair. Alors $f(x,y)=-1$. La relation $xx'+yy'=0$ implique que yy' est pair et donc y' est pair. Alors $x+x'$ a la parité de x' et $y+y'$ est impair. Que x' soit pair ou impair, on a toujours $f(x+x',y+y')=f(x',y')-1$.

• On traite de même les cas où x est impair.

Ainsi, la fonction f est dans E et fournit un antécédent par ψ du vecteur $(1,-1,1,-1)$.

Conclusion. L'espace vectoriel E est de dimension 4. C'est l'ensemble des fonctions de la forme $(x,y)\longmapsto \mathrm{A}(x^2+y^2)+\mathrm{B}x+\mathrm{C}y+\mathrm{D}((-1)^y-(-1)^x)$ avec $(\mathrm{A},\mathrm{B},\mathrm{C},\mathrm{D})\in\mathbb{R}^4$. ◁

Dans un espace euclidien E, *une forme linéaire ℓ s'écrit de manière unique $\ell : x\longmapsto \langle a,x\rangle$ où a est un vecteur de* E *: c'est le théorème de représentation de Riesz. L'application qui à $a\in\mathrm{E}$ associe la forme $x\longmapsto\langle a,x\rangle$ établit un isomorphisme entre* E *et son dual. L'exercice suivant montre comment s'écrivent les formes linéaires sur l'espace $\mathcal{L}(\mathrm{E})$ à l'aide du produit scalaire.*

1.8. Forme linéaire sur un sous-espace de $\mathcal{L}(\mathrm{E})$

Soit E un espace euclidien de dimension n, $\mathcal{F}$ un sous-espace de $\mathcal{L}(\mathrm{E})$ et φ une forme linéaire sur $\mathcal{F}$.

1. On suppose qu'il existe $x\in\mathrm{E}$ tel que $f(x)\neq 0$ pour tout $f\in\mathcal{F}\setminus\{0\}$. Montrer l'existence de $y\in\mathrm{E}$ tel que, pour tout $f\in\mathcal{F}$, $\varphi(f)=\langle f(x),y\rangle$.

2. Soit $(e_1,\dots,e_n)$ une base de E. Montrer l'existence de $(y_1,\dots,y_n)\in\mathrm{E}^n$ tel que, pour tout $f\in\mathcal{F}$, $\varphi(f)=\sum_{i=1}^{n}\langle f(e_i),y_i\rangle$.

(**École polytechnique**)

▷ **Solution.**

1. L'application linéaire $g : f\in\mathcal{F}\longmapsto f(x)\in\mathrm{E}$ est injective, car par hypothèse, son noyau est réduit à 0. Elle réalise donc un isomorphisme de $\mathcal{F}$ sur l'espace $\mathrm{E}'=\operatorname{Im} g$. Considérons l'application linéaire $\psi : \mathrm{E}\longrightarrow\mathcal{F}^*$ qui à $y\in\mathrm{E}$ associe $\psi(y) : f\in\mathcal{F}\longmapsto\langle f(x),y\rangle$. L'application ψ se révèle injective sur l'espace $\mathrm{E}'=\operatorname{Im} g$, car si $f\in\mathcal{F}\setminus\{0\}$, alors $g(f)=f(x)\neq 0$ et donc $\psi(g(f))(f)=\|f(x)\|^2\neq 0$ et $\psi(g(f))\neq 0$. Il s'ensuit que son image est de dimension supérieure ou égale à $\dim\mathrm{E}'=\dim\mathcal{F}$. Comme

$\operatorname{Im}\psi \subset \mathcal{F}^*$ et $\dim \mathcal{F}^* = \dim \mathcal{F}$, l'application ψ est surjective. Le résultat est démontré.

Le vecteur y n'est pas l'unique solution dans E *: on peut évidemment lui ajouter n'importe quel vecteur de l'orthogonal de* E′. *On pouvait également constater qu'en notant $g^{-1} : \mathrm{E}' \longrightarrow \mathcal{F}$ l'isomorphisme réciproque, l'application $\varphi \circ g^{-1}$ est une forme linéaire sur l'espace euclidien* E′. *Par suite, il existe un vecteur y de* E′ *(unique) tel que $(\varphi \circ g^{-1})(t) = (t, y)$ pour tout $t \in \mathrm{E}'$. En appliquant cela avec $t = g(f) = f(x)$ pour $f \in \mathcal{F}$ on a donc $\varphi(f) = \langle f(x), y\rangle$ pour tout $f \in \mathcal{F}$.*

2. L'application linéaire $h : f \in \mathcal{F} \longmapsto (f(e_1), \ldots, f(e_n)) \in \mathrm{E}^n$ est injective, car si une application linéaire est nulle sur la base $(e_1, \ldots, e_n)$, elle est identiquement nulle. Donc h réalise un isomorphisme de $\mathcal{F}$ sur $\operatorname{Im} h \subset \mathrm{E}^n$. Introduisons $\Psi : \mathrm{E}^n \longrightarrow \mathcal{F}^*$, l'application linéaire qui au n-uplet $(y_1, \ldots, y_n)$ associe la forme linéaire $f \longmapsto \sum_{i=1}^{n} \langle f(e_i), y_i\rangle$. L'application Ψ est injective sur $\operatorname{Im} h$: en effet, si f est un élément non nul de $\mathcal{F}$, $\Psi(h(f))(f) = \sum_{i=1}^{n} \langle f(e_i), f(e_i)\rangle > 0$ et $\Psi(h(f)) \neq 0$. Ainsi, l'image de Ψ est de dimension au moins égale à $\dim \operatorname{Im} h = \dim \mathcal{F} = \dim \mathcal{F}^*$. Il s'ensuit que Ψ est surjective. Il existe donc $(y_1, \ldots, y_n) \in \mathrm{E}^n$ tel que $\varphi(f) = \sum_{i=1}^{n} \langle f(e_i), y_i\rangle$. ◁

Par exemple, si $\mathcal{F} = \mathcal{L}(\mathrm{E})$ et $(e_1, \ldots, e_n)$ est une base orthonormale, pour $\varphi = \mathrm{Tr}$, le n-uplet $(y_1, \ldots, y_n)$ est égal à $(e_1, \ldots, e_n)$.

On aborde maintenant quelques exercices sur les projecteurs orthogonaux. Rappelons que si p est un projecteur d'un espace euclidien E *alors p est orthogonal si, et seulement si, il est auto-adjoint. Le premier exercice donne une autre caractérisation utile des projecteurs orthogonaux : ce sont les seuls qui diminuent les longueurs, autrement dit, qui ont une norme triple inférieure ou égale à* 1.

1.9. Caractérisation des projecteurs orthogonaux (1)

Soit E un espace euclidien et p un projecteur de E. Montrer que p est orthogonal si, et seulement si, pour tout $x \in \mathrm{E}$, $\|p(x)\| \leqslant \|x\|$.
(École polytechnique)

Solution.

• Si p est un projecteur orthogonal, on a $p(x) \perp (x - p(x))$ pour tout $x \in \mathrm{E}$. Du théorème de Pythagore, on déduit que

$$\|x\|^2 = \|p(x)\|^2 + \|x - p(x)\|^2 \geqslant \|p(x)\|^2.$$

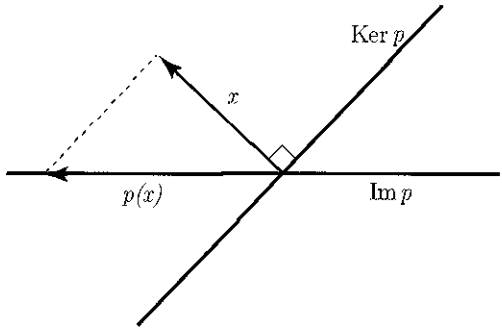
On a donc bien $\|p(x)\| \leqslant \|x\|$.

• Montrons la réciproque. On suppose donc $\|p\| \leqslant 1$. Soit $x \in \operatorname{Ker} p$ et $y \in \operatorname{Im} p$. On va montrer que $x \perp y$. Appliquons l'hypothèse au vecteur $x + \lambda y$ pour $\lambda \in \mathbb{R}$. On a $\|p(x + \lambda y)\|^2 \leqslant \|x + \lambda y\|^2$ et donc

$$\lambda^2 \|y\|^2 \leqslant \lambda^2 \|y\|^2 + \|x\|^2 + 2\lambda \langle x, y \rangle.$$

La fonction affine $\lambda \longmapsto \|x\|^2 + 2\lambda \langle x, y \rangle$ étant positive sur $\mathbb{R}$, elle est forcément constante et $\langle x, y \rangle = 0$. Donc $\operatorname{Ker} p \perp \operatorname{Im} p$ et p est un projecteur orthogonal. $\lhd$

Une autre manière de montrer la réciproque consiste à prouver la contraposée : si p n'est pas orthogonal il est facile de voir, à l'aide du théorème de Pythagore, que si x est un vecteur de $(\operatorname{Ker} p)^\perp$ qui n'est pas dans $\operatorname{Im} p$ (et il en existe) alors $\|p(x)\| > \|x\|$.



1.10. Caractérisation des projecteurs orthogonaux (2)

Soit $\mathcal{E}$ une partie de $\mathbb{R}^n$ et A un endomorphisme de $\mathbb{R}^n$ euclidien tels que,

$$\forall v \in \mathbb{R}^n, \quad \mathrm{A}v \in \mathcal{E} \text{ et } v - \mathrm{A}v \in \mathcal{E}^\perp.$$

Que dire de A et de $\mathcal{E}$?

(École polytechnique)

$\rhd$ **Solution.**

La première hypothèse signifie que $\operatorname{Im} \mathrm{A} \subset \mathcal{E}$. On va montrer qu'il y a égalité. Soit $x \in \mathcal{E}$. On a

$$\|x - \mathrm{A}x\|^2 = \langle x, x - \mathrm{A}x \rangle - \langle \mathrm{A}x, x - \mathrm{A}x \rangle = 0 - 0 = 0$$

d'après la seconde hypothèse. Donc $x = \mathrm{A}x \in \operatorname{Im} \mathrm{A}$.

On a donc montré que $\mathcal{E} = \operatorname{Im} \mathrm{A}$ et que tout vecteur de $\mathcal{E}$ est fixe par A. Par suite, $\mathcal{E}$ est un sous-espace vectoriel de E et A est une pro-

jection sur $\mathcal{E}$. Or, comme $x - Ax \in \mathcal{E}^\perp$ pour tout x, la projection A est orthogonale.

Conclusion. $\mathcal{E}$ est un sous-espace vectoriel de $\mathbb{R}^n$ et A est la projection orthogonale sur $\mathcal{E}$. ◁

1.11. Norme des projetés d'une base orthonormale

Soit E un espace vectoriel euclidien de dimension n et F un sous-espace vectoriel de E de dimension m. On note p la projection orthogonale de E sur F.

1. Calculer $\sum_{i=1}^{n} \|p(e_i)\|^2$ lorsque $(e_i)_{1\leqslant i\leqslant n}$ une base orthonormale de E.

2. Soit $(a_i)_{1\leqslant i\leqslant n}$ une famille de réels compris entre 0 et 1 dont la somme est m. Montrer l'existence d'une base orthonormale $(e_i)_{1\leqslant i\leqslant n}$ de E telle que, pour tout $i \in [\![1, n]\!]$, $\|p(e_i)\|^2 = a_i$.

(**École normale supérieure**)

Solution.

1. On a, pour tout $i \in [\![1, n]\!]$,

$$\|p(e_i\|^2 = (p(e_i), p(e_i)\rangle = \langle p(e_i), e_i\rangle,$$

car $p(e_i)$ est orthogonal à $e_i - p(e_i)$. On en déduit, puisque la base est orthonormale,

$$\sum_{i=1}^{n} \|p(e_i)\|^2 = \mathrm{Tr}(p) = m,$$

car la trace d'un projecteur est égale à son rang. On note que par ailleurs on a $0 \leqslant \|p(e_i)\|^2 \leqslant \|e_i\|^2 \leqslant 1$ pour tout $i \in [\![1, n]\!]$.

2. On observe que si $m = n$, tous les a_i sont nécessairement égaux à 1 et $p = \mathrm{Id_E}$: dans ce cas toute base orthonormale de E convient. Il en est de même lorsque $m = 0$ (tous les a_i sont égaux à 0 et $p = 0$).

On raisonne par récurrence sur n. Pour $n = 1$, $m = 0$ ou $m = 1$ et la remarque précédente permet de conclure.

On suppose que la propriété est vérifiée pour un espace vectoriel de dimension $n - 1$ ($n \geqslant 2$) et on la démontre pour un espace de dimension n. On peut supposer que $0 < m < n$. Soit e un vecteur unitaire de $\mathrm{F}^\perp$ et $\mathrm{E}' = e^\perp$. L'espace vectoriel E′ est de dimension $n - 1$ et contient F. On veut appliquer l'hypothèse de récurrence à la restriction p' de p à E′. Pour cela, il faudrait avoir $n - 1$ coefficients de somme m, ce qui peut s'obtenir en regroupant deux coefficients a_i. On souhaite donc trouver i et j dans $[\![1, n]\!]$ distincts tels que $a_i + a_j \leqslant 1$. Si une telle

paire n'existe pas, on a $\sum_{i<j}(a_i+a_j) > \mathrm{C}_n^2 = \dfrac{n(n-1)}{2}$, mais par ailleurs $\sum_{i<j}(a_i+a_j) = (n-1)\sum_{i=1}^{n} a_i = (n-1)m$. On obtient donc $m > \dfrac{n}{2}\cdot$ Distinguons deux cas.

• Si on a $m \leqslant \dfrac{n}{2}$, on peut trouver deux indices i et j distincts tels que $a_i + a_j \leqslant 1$. On peut supposer que $a_{n-1}+a_n \leqslant 1$, car la propriété ne dépend pas de l'ordre des a_i. Par hypothèse de récurrence, il existe une base orthonormale $\mathcal{B} = (e_1, \ldots, e_{n-2}, e')$ de E′ telle que, pour tout i de $[\![1, n-2]\!]$, $\|p(e_i)\|^2 = \|p'(e_i)\|^2 = a_i$ et $\|p(e')\|^2 = \|p'(e')\|^2 = a_{n-1}+a_n$.

Par ailleurs $p(e) = 0$, puisque $e \in \mathrm{F}^\perp$. Si $a_n = 0$, la base $(e_1, \ldots, e_{n-2}, e', e)$ convient. Sinon, on cherche e_{n-1} et e_n dans $\mathrm{Vect}(e', e)$ tels que $\|p(e_i)\|^2 = a_i$ pour $n-1 \leqslant i \leqslant n$. Pour $\theta \in \mathbb{R}$ posons $e_{n-1} = \cos\theta e' + \sin\theta e$ et $e_n = -\sin\theta e' + \cos\theta e$. On a fait tourner la base (e', e) d'un angle θ. Ces vecteurs sont de norme 1 et orthogonaux et on a

$$\|p(e_{n-1})\|^2 = \cos^2\theta \|p(e')\|^2 = \cos^2\theta(a_{n-1}+a_n)$$

et

$$\|p(e_n)\|^2 = \sin^2\theta(a_{n-1}+a_n).$$

Il suffit de choisir θ tel que $\cos^2\theta = \dfrac{a_{n-1}}{a_{n-1}+a_n}$ pour que la base $(e_1, \ldots, e_n)$ convienne.

• Si $m > \dfrac{n}{2}$, on a $n - m < \dfrac{n}{2}\cdot$ On considère la projection orthogonale q sur le sous-espace $\mathrm{F}^\perp$ de E. Les coefficients $1 - a_i$ appartiennent à $[0,1]$ et vérifient $\sum_{i=1}^{n}(1-a_i) = n - m$. Comme $\dim \mathrm{F}^\perp < \dfrac{n}{2}$, le raisonnement précédent s'applique à q et on peut trouver une base orthonormale $\mathcal{B}(e_1, \ldots, e_n)$ de E telle que, pour tout $i \in [\![1, n]\!]$, $\|q(e_i)\|^2 = 1 - a_i$. On a pour tout $i \in [\![1, n]\!]$, $\|p(e_i)\|^2 = \|e_i\|^2 - \|q(e_i)\|^2 = 1 - (1 - a_i) = a_i$, ce qui termine la démonstration par récurrence.

Nous avons donc démontré qu'étant donné un projecteur orthogonal p de rang m, il existe une base orthonormale $(e_1, \ldots, e_n)$ telle que $\|p(e_i)\|^2 = a_i$ pour tout $i \in [\![1, n]\!]$ si, et seulement si, les nombres a_i appartiennent à $[0,1]$ et ont pour somme m. ◁

Comme il a été montré dans la question 1 de l'exercice, les coefficients $\|p(e_i)\|^2$ *sont les coefficients diagonaux de la matrice* A *de* p *dans la base* $\mathcal{B}$. *Dans une base orthonormale de* E *formée d'une base de* F *et d'une base de* $\mathrm{F}^\perp$, *la matrice de* p *est* $\mathrm{J}_m = \mathrm{Diag}(\underbrace{1, \ldots, 1}_{m \text{ termes}}, \underbrace{0, \ldots, 0}_{n-m \text{ termes}})$. *Comme* $\mathrm{A} = \mathrm{U}\mathrm{J}_m\mathrm{U}^{-1}$, *le problème posé se traduit matriciellement : il existe une matrice orthogonale* U *telle que* $\mathrm{U}\mathrm{J}_m\mathrm{U}^{-1}$ *ait pour diagonale* $(a_1, \ldots, a_n)$ *si, et seulement si, les nombres* a_i *appartiennent à* $[0,1]$ *et ont pour*

somme m. La récurrence de la question 2 peut être faite en raisonnant matriciellement.

L'exercice suivant est à rapprocher de l'exercice 6.11 du tome 1 d'algèbre.

1.12. Pseudo-inverse

Soit E et E′ deux espaces euclidiens, $a \in \mathcal{L}(E, E')$. Montrer qu'il existe une unique application linéaire $b \in \mathcal{L}(E', E)$ telle que :

(*i*) $a \circ b$ et $b \circ a$ sont symétriques ;

(*ii*) $a \circ b \circ a = a$ et $b \circ a \circ b = b$.

(École polytechnique)

Solution.

Pour alléger la rédaction nous noterons la composée de deux applications linéaires par simple juxtaposition en omettant le symbole $\circ$.

• *Analyse.* Supposons que b soit une solution du problème. On a alors $abab = ab$ et $baba = ba$, donc ab et ba sont des projecteurs. Comme ce sont des endomorphismes symétriques, il s'agit de projecteurs orthogonaux. On va préciser les images de ces projecteurs.

Pour tout $x \in \operatorname{Ker} ba$, on a $a(x) = aba(x) = 0$, donc $\operatorname{Ker} ba \subset \operatorname{Ker} a$. Comme l'inclusion inverse est toujours vérifiée, on en déduit que $\operatorname{Ker} ba = \operatorname{Ker} a$. Donc ba est la projection orthogonale sur $(\operatorname{Ker} a)^\perp$. On a, pour tout $x \in E$, $ab(a(x)) = a(x)$. On en déduit que $\operatorname{Im} a \subset \operatorname{Im} ab$. Comme on a toujours l'inclusion inverse, on en déduit que $\operatorname{Im} ab = \operatorname{Im} a$, puis que $\operatorname{Ker} ab = (\operatorname{Im} a)^\perp$.

Essayons maintenant de caractériser b. Comme $\operatorname{Im} ba = (\operatorname{Ker} a)^\perp$ on a pour tout $x \in (\operatorname{Ker} a)^\perp$, $b(a(x)) = x$. Or la restriction a' de a à $(\operatorname{Ker} a)^\perp$ établit un isomorphisme entre $(\operatorname{Ker} a)^\perp$ et $\operatorname{Im} a$. On en déduit que, pour tout $x \in \operatorname{Im} a$,

$$b(x) = a'^{-1}(x).$$

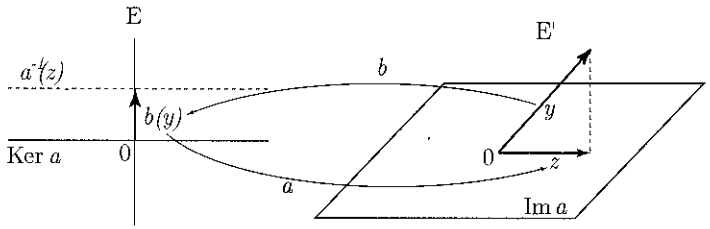
Si on prend $x \in (\operatorname{Im} a)^\perp$, on a $(ab)(x) = 0$ et donc $b(x) = (bab)(x) = 0$.

Cette analyse montre donc que si b existe il est unique et caractérisé par $b_{|\operatorname{Im} a} = a'^{-1}$ et $b_{|(\operatorname{Im} a)^\perp} = 0$.

• *Synthèse.* Il existe évidemment une application linéaire b de E′ dans E dont la restriction à $\operatorname{Im} a$ soit a'^{-1} et la restriction à $(\operatorname{Im} a)^\perp$ soit nulle. Montrons qu'elle a bien les propriétés voulues. On a $ba = a'^{-1}a$, donc ba est la projection orthogonale sur $(\operatorname{Ker} a)^\perp$. De même, $ab_{|(\operatorname{Im} a)^\perp} = 0$ et $ab_{|\operatorname{Im} a} = aa'^{-1} = \operatorname{Id}_{\operatorname{Im} a}$, donc ab est la projection orthogonale sur $\operatorname{Im} a$. On en déduit que (*i*) est vérifié, puisqu'une projection orthogonale est un endomorphisme symétrique.

L'égalité $aba = a$ résulte du fait que ab est la projection orthogonale sur $\operatorname{Im} a$; de même, on a $bab = b$, car ba est la projection orthogonale sur $(\operatorname{Ker} a)^\perp$ et $\operatorname{Im} b = (\operatorname{Ker} a)^\perp$. Ainsi (ii) est vérifié. $\lhd$

Donnons une description géométrique de ce pseudo-inverse b. L'application linéaire a est quelconque et a priori ni injective ni surjective. Une équation $a(x) = y$ d'inconnue $x \in \mathrm{E}$ n'a donc pas forcément de solution et si elle en a, il n'y a pas forcément unicité. Si $y \in \mathrm{E}'$ est fixé, l'idée est chercher l'élément de $\operatorname{Im} a$ le plus proche de y à savoir le projeté orthogonal de y sur $\operatorname{Im} a$. Notons-le z. Parmi tous les antécédents de z, il y en a un unique dans l'orthogonal de $\operatorname{Ker} a$. C'est cet antécédent qui est l'image de y par b. Ce choix a le mérite de préserver la symétrie : si b est le pseudo-inverse de a, alors a est le pseudo-inverse de b. Voir aussi l'exercice 2.24 pour une expression du pseudo-inverse.



Cette notion intervient dans de nombreux domaines : analyse fonctionnelle, traitement du signal, réseaux de neurones...

Dans l'exercice suivant on pourra utiliser la caractérisation des projecteurs orthogonaux vue en 1.9.

1.13. Condition pour que deux projecteurs orthogonaux commutent

Soit E un espace euclidien, p et q deux projecteurs orthogonaux de E. Montrer l'équivalence entre

(i) $p \circ q = q \circ p$;

(ii) $p \circ q$ est un projecteur ;

(iii) p et q possèdent une base commune de diagonalisation.

(**École polytechnique**)

$\rhd$ **Solution.**

• On a $(i) \Longrightarrow (iii)$. C'est un résultat général : deux endomorphismes diagonalisables qui commutent peuvent se diagonaliser dans une même base (voir l'exercice 2.19 du tome 2 d'algèbre). Notons que (i) implique aussi trivialement (ii), puisque si (i) est vérifié, on a

$$(p \circ q)^2 = p \circ q \circ p \circ q = p^2 \circ q^2 = p \circ q.$$

Cette implication est d'ailleurs vraie pour des projecteurs quelconques.

• On a $(iii) \Longrightarrow (i)$. Si p et q se diagonalisent dans une même base, ils commutent forcément.

• Il reste donc à prouver que $(ii) \Longrightarrow (i)$. Le fait que l'on travaille avec des projecteurs orthogonaux est ici fondamental. On verra plus loin que le résultat est faux avec des projecteurs quelconques. L'idée est d'utiliser la caractérisation de l'exercice 1.9. En effet, si $x \in \mathrm{E}$, on a, puisque p et q sont orthogonaux,

$$\|(p \circ q)(x)\| = \|p(q(x))\| \leqslant \|q(x)\| \leqslant \|x\|.$$

Ainsi, $p \circ q$ est orthogonal et donc auto-adjoint. Il vient alors

$$p \circ q = (p \circ q)^* = q^* \circ p^* = q \circ p.$$

D'où le résultat. $\lhd$

Si les projecteurs p et q ne sont pas orthogonaux, $p \circ q$ peut être un projecteur sans que p et q commutent, comme on le voit en considérant les projecteurs p et q de $\mathbb{R}^2$ de matrices respectives $\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *et* $\begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ *dans la base canonique. On peut en effet vérifier que $p \circ q = 0$ et $q \circ p \neq 0$.*

Le fait qu'un projecteur orthogonal diminue les longueurs intervient encore dans l'exercice suivant. Attirons également l'attention sur une propriété reliant image et noyau d'une application linéaire et de son adjoint : si $u : \mathrm{E} \longrightarrow \mathrm{F}$ est une application linéaire entre deux espaces euclidiens, on a $\operatorname{Im} u^* = (\operatorname{Ker} u)^\perp$ et $\operatorname{Ker} u^* = (\operatorname{Im} u)^\perp$.

1.14. Composition de projecteurs orthogonaux

Soit E un espace euclidien, $\mathrm{N}_1, \ldots, \mathrm{N}_k$ des sous-espaces vectoriels de E et $p_1, \ldots, p_k$ les projecteurs orthogonaux sur $\mathrm{N}_1, \ldots, \mathrm{N}_k$. On pose $q = p_k \circ \cdots \circ p_1$ et $\mathrm{N}_0 = \bigcap_{i=1}^{k} \mathrm{N}_i$.

1. Montrer que pour tout $x \in \mathrm{E}$, $\|q(x)\| = \|x\| \iff x \in \mathrm{N}_0$.

2. Montrer que $\operatorname{Ker}(q - \mathrm{Id}) = \mathrm{N}_0$.
En déduire que $\operatorname{Ker}(q^* - \mathrm{Id}) = \mathrm{N}_0$.

3. Montrer que $\mathrm{E} = \mathrm{N}_0 \overset{\perp}{\oplus} \operatorname{Im}(q - \mathrm{Id})$.

4. Soit $(x_n)_{n \in \mathbb{N}}$ une suite de E telle que pour tout n, $\|x_n\| \leqslant 1$ et $\lim_{n \to +\infty} \|q(x_n)\| = 1$. Montrer que $\lim_{n \to +\infty} \|(q - \mathrm{Id})(x_n)\| = 0$.

(École polytechnique)

▷ **Solution.**

1. Il est clair qu'un vecteur x de N_0 est laissé fixe par q, puisqu'il l'est par tous les projecteurs p_i. En particulier, on a $\|q(x)\| = \|x\|$. Réciproquement, soit x un vecteur qui n'est pas dans N_0. Notons i le plus petit indice tel que $x \notin \mathrm{N}_i$. On a $p_1(x) = \cdots = p_{i-1}(x)$ mais $p_i(x) \neq x$. On a alors $\|p_i(x)\| < \|x\|$: cela découle simplement du théorème de Pythagore, puisque $\|x\|^2 = \|p_i(x)\|^2 + \|x - p_i(x)\|^2$. Les projecteurs $p_{i+1}, \ldots, p_k$ ne pouvant pas augmenter la norme, on a donc $\|q(x)\| < \|x\|$.

On a donc $\|q(x)\| \leqslant \|x\|$ pour tout vecteur x de E et il y a égalité si et seulement si $x \in \mathrm{N}_0$.

2. L'égalité $\mathrm{Ker}(q - \mathrm{Id}) = \mathrm{N}_0$ découle directement de ce qui précède. On a $q^* = p_1^* \circ \cdots \circ p_k^* = p_1 \circ \cdots \circ p_k$: il s'agit de la composée des mêmes projecteurs mais dans un ordre différent. En appliquant le résultat précédent, on a donc aussi

$$\mathrm{Ker}(q^* - \mathrm{Id}) = \bigcap_{i=1}^{k} \mathrm{N}_i = \mathrm{N}_0.$$

3. On a $\mathrm{N}_0 = \mathrm{Ker}(q^* - \mathrm{Id}) = \mathrm{Ker}(q - \mathrm{Id})^* = \mathrm{Im}(q - \mathrm{Id})^{\perp}$. Donc N_0 est le supplémentaire orthogonal de $\mathrm{Im}(q - \mathrm{Id})$. On a bien $\mathrm{E} = \mathrm{N}_0 \overset{\perp}{\oplus} \mathrm{Im}(q - \mathrm{Id})$.

4. La suite $(\|(q - \mathrm{Id})(x_n)\|)$ est bornée par 2. Supposons par l'absurde qu'elle ne converge pas vers 0. Elle admet alors une valeur d'adhérence $\alpha > 0$. Soit $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ une extraction telle que $\|(q - \mathrm{Id})(x_{\varphi(n)})\|$ tende vers α. Quitte à encore prendre une sous-suite de celle-ci, on peut très bien supposer que la suite $(x_{\varphi(n)})$ converge aussi vers un vecteur a avec $\|a\| \leqslant 1$. Alors, la suite $\| q(x_{\varphi(n)})\|$ converge vers $\|q(a)\|$ et donc $\|q(a)\| = 1$. Et, toujours en passant à la limite, on a aussi, $\|q(a) - a\| = \alpha$. Or, on a $1 = \|q(a)\| \leqslant \|a\| \leqslant 1$ de sorte qu'il y a égalité $\|q(a)\| = \|a\| = 1$. D'où la contradiction, car d'après la question 1, $q(a) - a$ doit alors être nul et ne peut donc avoir une norme égale à α. ◁

La situation de l'exercice précédent est un cas particulier du suivant.

1.15. Contraction d'un espace euclidien

Soit E un espace euclidien et $u \in \mathcal{L}(\mathrm{E})$ tel que $\|u\| \leqslant 1$. Montrer que $\mathrm{E} = \mathrm{Ker}(u - \mathrm{Id}) \overset{\perp}{\oplus} \mathrm{Im}(u - \mathrm{Id})$ et que la suite $v_p = \dfrac{1}{p+1} \displaystyle\sum_{k=0}^{p} u^k$ converge vers le projecteur orthogonal sur $\mathrm{Ker}(u - \mathrm{Id})$.

(**École polytechnique**)

Solution.

• Le fait que $\mathrm{Ker}(u-\mathrm{Id})$ et $\mathrm{Im}(u-\mathrm{Id})$ sont en somme directe est très général et reste valable pour une contraction d'un espace vectoriel normé quelconque. En effet, soit x dans l'intersection des deux espaces. On a $u(x)=x$ et x s'écrit $x=u(y)-y$ pour un certain vecteur y. On a donc aussi $x=u(x)=u^2(y)-u(y)$ puis $x=u^3(y)-u^2(y)$ et ainsi de suite $x=u^{k+1}(y)-u^k(y)$ pour tout entier $k\geqslant 0$. Sommons ces égalités pour k variant entre 0 et un entier $p-1$. On obtient $px=u^p(y)-y$. On en déduit que

$$\|x\|=\frac{\|u^p(y)-y\|}{p}\leqslant\frac{\|u^p(y)\|+\|y\|}{p}\leqslant\frac{2\|y\|}{p}.$$

En faisant tendre p vers l'infini, on a donc $x=0$. La somme de $\mathrm{Ker}(u-\mathrm{Id})$ et $\mathrm{Im}(u-\mathrm{Id})$ est donc directe et, comme E est de dimension finie, on a bien $\mathrm{E}=\mathrm{Ker}(u-\mathrm{Id})\oplus\mathrm{Im}(u-\mathrm{Id})$ par le théorème du rang.

• On va montrer maintenant que la suite (v_p) converge vers la projection sur $\mathrm{Ker}(u-\mathrm{Id})$ parallèlement à $\mathrm{Im}(u-\mathrm{Id})$. Notons tout d'abord que si $x\in\mathrm{Ker}(u-\mathrm{Id})$, on a $u(x)=x$ et donc $v_p(x)=x$ pour tout p. Prenons maintenant x dans $\mathrm{Im}(u-\mathrm{Id})$ que l'on écrit comme avant $x=u(y)-y$. On a $u^k(x)=u^{k+1}(y)-u^k(y)$ pour tout $k\geqslant 0$ de sorte que $v_p(x)=\dfrac{1}{p+1}\,(u^{p+1}(y)-y)$ et cela tend vers 0 comme précédemment. Ainsi, (v_p) converge bien vers p, le projecteur sur $\mathrm{Ker}(u-\mathrm{Id})$ parallèlement à $\mathrm{Im}(u-\mathrm{Id})$.

Pour finir, prouvons que ce projecteur est orthogonal. On a, pour tout k, $|||u^k|||\leqslant|||u|||^k\leqslant 1$ (par sous-multiplicativité de la norme triple) et donc, par inégalité triangulaire, on a $|||v_p|||\leqslant 1$ pour tout entier p. En passant à la limite, on a donc $|||p|||\leqslant 1$. Il est classique (voir l'exercice I.9) que cela permet de dire que p est un projecteur orthogonal et donc que $\mathrm{Ker}(u-\mathrm{Id})\perp\mathrm{Im}(u-\mathrm{Id})$. ◁

On peut donner une autre solution pour l'orthogonalité de l'image et du noyau à l'aide de l'adjoint. En effet, comme

$$|||u|||=\sup_{\substack{\|x\|\leqslant 1\\ \|y\|\leqslant 1}}\langle u(x),y\rangle=\sup_{\substack{\|x\|\leqslant 1\\ \|y\|\leqslant 1}}\langle x,u^*(y)\rangle=|||u^*|||,$$

on a $|||u^*|||=|||u|||\leqslant 1$. *Soit alors* $x\in\mathrm{Ker}(u-\mathrm{Id})$. *On a*

$$\|u^*(x)-x\|^2=\|u^*(x)\|^2-2\langle x,u^*(x)\rangle+\|x\|^2=\|u^*(x)\|^2-2\langle u(x),x\rangle+\|x\|^2$$

et donc $\|u^*(x)-x\|^2\leqslant\|x\|^2-2\|x\|^2+\|x\|^2=0$ *et donc* $u^*(x)=x$. *Par symétrie, on a donc* $\mathrm{Ker}(u-\mathrm{Id})=\mathrm{Ker}(u^*-\mathrm{Id})=\mathrm{Im}(u-\mathrm{Id})^\perp$. *Cette approche à été proposée dans un autre exercice d'oral posé à l'École polytechnique.*

Lorsqu'on dispose d'un sous-espace de dimension finie F *dans un espace réel* E *muni d'un produit scalaire, nous savons que* $\mathrm{E} = \mathrm{F} \overset{\perp}{\oplus} \mathrm{F}^{\perp}$. *Il en résulte notamment que pour tout* $x \in \mathrm{E}$, *on peut écrire* $x = x_{\mathrm{F}} + y$ *avec* $x_{\mathrm{F}} \in \mathrm{F}$ *et* $y \in \mathrm{F}^{\perp}$. *Par définition,* x_{F} *est le projeté orthogonal de* x *sur* F. *On vérifie facilement à l'aide du théorème de Pythagore que*

$$\|x - x_{\mathrm{F}}\| = d(x, \mathrm{F}) = \inf_{f \in \mathrm{F}} \|x - f\|$$

et que x_{F} *est l'unique point de* F *pour lequel la distance à* F *est atteinte.*

Un calcul de borne inférieure peut souvent s'interpréter en termes de distance à un sous-espace de dimension finie dans un espace préhilbertien. En voici un exemple.

1.16. Distance à un sous-espace de $\mathbb{R}[\mathrm{X}]$

Évaluer

$$\inf_{(a_1,\dots,a_n)\in\mathbb{R}^n} \int_0^{+\infty} e^{-x}(1 + a_1 x + \cdots + a_n x^n)^2 \,\mathrm{d}x.$$

(École polytechnique)

▷ **Solution.**

On définit un produit scalaire sur $\mathbb{R}[\mathrm{X}]$, en posant pour P et Q dans $\mathbb{R}[\mathrm{X}]$,

$$\langle \mathrm{P}, \mathrm{Q} \rangle = \int_0^{+\infty} e^{-x} \mathrm{P}(x)\mathrm{Q}(x) \,\mathrm{d}x.$$

L'expression que l'on cherche à minimiser est égale à $\|1 + \mathrm{P}\|^2$ où $\mathrm{P} = \sum_{i=1}^{n} a_i \mathrm{X}^i$: c'est le carré de la distance entre P et -1. Quand les a_i varient dans $\mathbb{R}$, P décrit le sous-espace $\mathrm{E} = \mathrm{Vect}(\mathrm{X}, \mathrm{X}^2, \dots, \mathrm{X}^n)$ de $\mathbb{R}[\mathrm{X}]$ qui est de dimension finie. Il en résulte que la distance entre P et -1 possède un minimum, atteint pour un unique polynôme, à savoir le projeté orthogonal de -1 sur E. Ce projeté, que nous appellerons toujours P, est caractérisé par le fait que, $\langle \mathrm{P} + 1, \mathrm{X}^k \rangle = 0$ pour tout $k \in [\![1, n]\!]$. Calculons ces produits scalaires

$$\langle \mathrm{P} + 1, \mathrm{X}^k \rangle = \int_0^{+\infty} e^{-x} \Big(x^k + \sum_{i=1}^{n} a_i x^{k+i} \Big) \mathrm{d}x.$$

On vérifie aisément que, pour tout entier n, $\int_0^{+\infty} e^{-x} x^n \,\mathrm{d}x = n!$ (en utilisant ses connaissances sur la fonction Gamma, ou plus prosaïquement, par récurrence à l'aide d'une intégration par parties). On a donc, pour

tout $k \in [\![1,n]\!]$, $k! + \sum_{i=1}^{n} a_i(k+i)! = 0$, ce qui peut s'écrire en divisant par $k!$,

$$1 + \sum_{i=1}^{n} a_i(k+1)(k+2)\dots(k+i) = 0.$$

On va voir qu'il n'est pas indispensable d'expliciter P, c'est-à-dire de résoudre ce système linéaire, pour trouver la distance de P à -1. En effet, le minimum cherché vaut

$$\begin{aligned}\|\mathrm{P}+1\|^2 &= \langle \mathrm{P}+1, \mathrm{P}+1\rangle = \langle \mathrm{P}+1, 1\rangle \\ &= \int_0^{+\infty} e^{-x}(1 + a_1x + \dots + a_nx^n)\,\mathrm{d}x = 1 + \sum_{i=1}^{n} a_i i!,\end{aligned}$$

car $\mathrm{P}+1$ est orthogonal à P.

Considérons alors le polynôme $\mathrm{Q} = 1 + \sum_{i=1}^{n} a_i(\mathrm{X}+1)\dots(\mathrm{X}+i)$. Il est de degré inférieur ou égal à n, le coefficient de X^n étant a_n. Il suffit de calculer $\mathrm{Q}(0)$. Or les relations d'orthogonalité qui définissent P s'écrivent $\mathrm{Q}(1) = \dots = \mathrm{Q}(n) = 0$. On a donc $\mathrm{Q} = a_n(\mathrm{X}-1)\dots(\mathrm{X}-n)$ et $\mathrm{Q}(0) = a_n(-1)^n\,n!$. On remarque de plus que $\mathrm{Q}(-1) = 1$, c'est-à-dire que $a_n(-1)^n(n+1)! = 1$. Ainsi $(-1)^n a_n = \dfrac{1}{(n+1)!}$ et finalement $\mathrm{Q}(0) = \dfrac{n!}{(n+1)!} = \dfrac{1}{n+1}$. Le minimum cherché vaut $\boxed{\dfrac{1}{n+1}}$. ◁

Il est important de retenir que le produit scalaire canonique de $\mathcal{M}_n(\mathbb{R})$ peut s'écrire à l'aide de la trace :

$$(\mathrm{M},\mathrm{N}) = \sum_{1\leqslant i,j\leqslant n} m_{ij}n_{ij} = \mathrm{Tr}\,({}^t\mathrm{MN})$$

pour $\mathrm{M} = (m_{ij})_{1\leqslant i,j\leqslant n}$ et $\mathrm{N} = (n_{ij})_{1\leqslant i,j\leqslant n}$. La norme euclidienne associée est

$$\|\mathrm{M}\| = \sqrt{\sum_{1\leqslant i,j\leqslant n} m_{ij}^2} = \sqrt{\mathrm{Tr}({}^t\mathrm{MM})}.$$

L'exercice qui vient constitue une présentation générale de ce produit scalaire et le suivant s'inscrit dans le thème des calculs des distances.

1.17. Norme euclidienne canonique sur $\mathcal{M}_n(\mathbb{R})$

On pose $\mathrm{N}(\mathrm{A}) = \sqrt{\mathrm{Tr}({}^t\mathrm{AA})}$ pour $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$.

1. Montrer que N est une norme.

2. Montrer que $|\operatorname{Tr}(A)| \leqslant \sqrt{n}\,N(A)$.
3. Montrer que si $O \in O_n(\mathbb{R})$, on a $N(AO) = N(OA) = N(A)$.
4. Montrer que $N(AB) \leqslant N(A)N(B)$.
5. Comparer N avec d'autres normes de $\mathcal{M}_n(\mathbb{R})$.
(**École polytechnique**)

▷ **Solution.**

1. Il suffit d'observer que l'application $\Phi : (A, B) \longmapsto \operatorname{Tr}({}^tAB)$ est un produit scalaire sur $\mathcal{M}_n(\mathbb{R})$ (N est tout simplement la norme euclidienne associée à ce produit scalaire). En effet, la bilinéarité de Φ est une conséquence de la linéarité de la trace et de la transposition. Si A et B sont dans $\mathcal{M}_n(\mathbb{R})$, on a

$$\Phi(A, B) = \operatorname{Tr}({}^tAB) = \operatorname{Tr}({}^t({}^tAB)) = \operatorname{Tr}({}^tB\,{}^t({}^tA)) = \operatorname{Tr}({}^tBA) = \Phi(B, A).$$

Quant au caractère défini positif, il provient du fait que si $A = (a_{ij})_{1\leqslant i,j\leqslant n}$, alors $\Phi(A, A) = \sum_{1\leqslant i,j\leqslant n} a_{ij}^2$ qui est strictement positif pour A non nulle.

2. On utilise l'inégalité de Cauchy-Schwarz pour le produit scalaire ci-dessus :

$$|\operatorname{Tr}(A)| = |\operatorname{Tr}(IA)| \leqslant N(I)N(A) = \sqrt{n}\,N(A).$$

Comme il y a égalité pour $A = I_n$, on en déduit que la triple norme de la forme linéaire $\operatorname{Tr} : \mathcal{M}_n(\mathbb{R}) \longrightarrow \mathbb{R}$ est égale à $\sqrt{n}$.

3. Pour $O \in O_n(\mathbb{R})$ et $A \in \mathcal{M}_n(\mathbb{R})$, on a

$$\operatorname{Tr}({}^t(OA)(OA)) = \operatorname{Tr}({}^tA\,{}^tOOA) = \operatorname{Tr}({}^tAA),$$

et donc $N(OA) = N(A)$. De même, on a

$$\operatorname{Tr}({}^t(AO)(AO)) = \operatorname{Tr}({}^tO\,{}^tAAO) = \operatorname{Tr}({}^tAAO\,{}^tO) = \operatorname{Tr}({}^tAA)$$

et donc $N(AO) = N(A)$.

4. Posons $A = (a_{ij})_{1\leqslant i,j\leqslant n}$, $B = (b_{ij})_{1\leqslant i,j\leqslant n}$ et $C = AB$ avec $C = (c_{ij})_{1\leqslant i,j\leqslant n}$. On a $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$ pour tout $1 \leqslant i, j \leqslant n$ et il vient

$$\begin{aligned} N(AB)^2 &= \sum_{1\leqslant i,j\leqslant n} c_{ij}^2 = \sum_{i,j} \left(\sum_{k=1}^{n} a_{ik}b_{kj}\right)^2 \\ &\leqslant \sum_{1\leqslant i,j\leqslant n} \left(\sum_{k=1}^{n} a_{ik}^2 \sum_{l=1}^{n} b_{lj}^2\right) \quad \text{(inégalité de Cauchy-Schwarz)} \end{aligned}$$

$$\leqslant \left(\sum_{1\leqslant i,k\leqslant n} a_{ik}^2\right)\left(\sum_{1\leqslant j,l\leqslant n} b_{lj}^2\right) = \mathrm{N}(\mathrm{A})^2\mathrm{N}(\mathrm{B})^2.$$

La norme N *est donc une norme d'algèbre.*

5. Pour $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n}$, posons $\|\mathrm{A}\|_1 = \sum_{1\leqslant i,j\leqslant n} |a_{ij}|$ ainsi que $\|\mathrm{A}\|_\infty = \max_{1\leqslant i,j\leqslant n} |a_{ij}|$. On définit ainsi deux normes sur $\mathcal{M}_n(\mathbb{C})$. On a

$$\|\mathrm{A}\|_\infty \leqslant \mathrm{N}(\mathrm{A}) \leqslant n\|\mathrm{A}\|_\infty$$

et les constantes sont optimales. On a aussi

$$\mathrm{N}(\mathrm{A}) \leqslant \|\mathrm{A}\|_1 \leqslant n\mathrm{N}(\mathrm{A})$$

avec des constantes également optimales. ◁

1.18. Distance d'une matrice à l'espace des matrices symétriques

Soit $\mathcal{S}$ l'espace des matrices symétriques réelles de taille n et $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{R})$. Déterminer

$$\inf_{\mathrm{M}=(m_{ij})\in\mathcal{S}} \sum_{1\leqslant i,j\leqslant n} (a_{ij} - m_{ij})^2.$$

(École polytechnique)

Solution.

Il s'agit de déterminer $\inf_{\mathrm{M}\in\mathcal{S}} \|\mathrm{A} - \mathrm{M}\|^2$ pour la structure euclidienne canonique de $\mathcal{M}_n(\mathbb{R})$, autrement dit le carré de la distance de la matrice A au sous-espace $\mathcal{S}$. Nous savons que si H désigne le projeté orthogonal de A sur $\mathcal{S}$, alors on a

$$\|\mathrm{A} - \mathrm{H}\|^2 = \inf_{\mathrm{M}\in\mathcal{S}} \|\mathrm{A} - \mathrm{M}\|^2.$$

Cette borne inférieure, que nous noterons $d(\mathrm{A}, \mathcal{S})^2$, est atteinte en H (et c'est d'ailleurs le seul point où elle est atteinte).

Nous savons que si $\mathcal{A}$ désigne le sous-espace des matrices antisymétriques réelles de taille n, alors $\mathcal{S} \oplus \mathcal{A} = \mathcal{M}_n(\mathbb{R})$. En fait, $\mathcal{S}$ et $\mathcal{A}$ sont des supplémentaires orthogonaux, car si $\mathrm{M} \in \mathcal{S}$ et $\mathrm{N} \in \mathcal{A}$, on a

$$\begin{aligned}\langle \mathrm{M}, \mathrm{N}\rangle &= \mathrm{Tr}({}^t\mathrm{MN}) = \mathrm{Tr}(\mathrm{MN}) = \mathrm{Tr}(\mathrm{M}(-{}^t\mathrm{N})) = -\mathrm{Tr}(\mathrm{M}\,{}^t\mathrm{N}) \\ &= -\mathrm{Tr}({}^t\mathrm{NM}) = -\langle \mathrm{N}, \mathrm{M}\rangle = -\langle \mathrm{M}, \mathrm{N}\rangle\end{aligned}$$

d'où $\langle M, N\rangle = 0$. Ainsi, on a bien $\mathcal{M}_n(\mathbb{R}) = \mathcal{S} \overset{\perp}{\oplus} \mathcal{A}$ et donc $H = \dfrac{A + {}^tA}{2}$, et $A - H = \dfrac{A - {}^tA}{2}$. Il en résulte que

$$\boxed{d(A, \mathcal{S})^2 = \frac{1}{4} \sum_{1 \leqslant i,j \leqslant n} (a_{ij} - a_{ji})^2}. \quad \triangleleft$$

Notons qu'il est possible de résoudre cet exercice très élémentairement sans introduire le formalisme ci-dessus puisqu'on se ramène simplement à minimiser le trinôme $(x - a_{ij})^2 + (x - a_{ji})^2$ pour $i \neq j$.

On retrouve le produit scalaire canonique sur un espace de matrices dans l'exercice suivant, mais il s'agit cette fois de matrices rectangulaires. Le problème consiste encore à calculer des distances mais cette fois à des sous-espaces affines.

1.19. Problèmes de minimisation dans $\mathcal{M}_{p,q}(\mathbb{R})$

Soit $\Phi : \mathcal{M}_{p,q}(\mathbb{R})^2 \longrightarrow \mathbb{R}$ définie par $\Phi(A, B) = \mathrm{Tr}({}^tAB)$.

1. Établir que Φ est un produit scalaire. On notera $\| \ \|$ la norme associée.

2. Soit $a \in \mathbb{R}^q$, $a \neq 0$ et $A \in \mathcal{M}_{p,q}(\mathbb{R})$ fixés dans la suite. Pour $b \in \mathbb{R}^p$, on note $\Omega(b) = \{Y \in \mathcal{M}_{p,q}(\mathbb{R}),\ (A + Y)a = b\}$. Montrer que $\Omega(b)$ est non vide et qu'il contient un élément de norme minimum. Cet élément sera noté $f(b)$. Expliciter $f(b)$.

3. Soit $c \in \mathbb{R}^p$ et

$$\begin{array}{rccc} & \mathbb{R}^p & \longrightarrow & \mathcal{M}_{p,q+1}(\mathbb{R}) \\ \phi : & x & \longmapsto & (f(c + x), x). \end{array}$$

Trouver x qui minimise $\|\phi(x)\|$.

(École polytechnique)

▷ **Solution.**

1. Φ est bien définie et bilinéaire, car la trace est linéaire ; elle est symétrique car pour $(A, B) \in \mathcal{M}_{p,q}(\mathbb{R})^2$,

$$\Phi(A, B) = \mathrm{Tr}({}^tAB) = \mathrm{Tr}({}^t({}^tAB)) = \mathrm{Tr}({}^tBA) = \Phi(B, A),$$

étant donné qu'une matrice carrée et sa transposée ont la même trace. Enfin, si $A = (a_{ij})_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant q}} \in \mathcal{M}_{p,q}(\mathbb{R})$, on a

$$\Phi(\mathrm{A},\mathrm{A}) = \mathrm{Tr}({}^t\mathrm{AA}) = \sum_{\substack{1\leqslant i\leqslant p\\ 1\leqslant j\leqslant q}} a_{ij}^2 \geqslant 0,$$

avec inégalité stricte pour A non nulle, ce qui montre que Φ est définie positive.

Ainsi, Φ est un produit scalaire. On note que sur $\mathbb{R}^p$, que nous identifierons à $\mathcal{M}_{p,1}(\mathbb{R})$, il coïncide avec le produit scalaire canonique.

2. Les éléments $a \in \mathbb{R}^q \setminus \{0\}$ et $\mathrm{A} \in \mathcal{M}_{p,q}(\mathbb{R})$ étant fixés, considérons l'application linéaire

$$\varphi : \mathrm{Y} \in \mathcal{M}_{p,q}(\mathbb{R}) \longmapsto \mathrm{Y}a \in \mathbb{R}^p.$$

On observe que

$$\mathrm{Y} \in \Omega(b) \Longleftrightarrow \varphi(\mathrm{Y}) = b - \mathrm{A}a.$$

Pour démontrer que $\Omega(b)$ n'est pas vide, il suffit de démontrer que φ est surjective. L'ensemble $\Omega(b)$ sera alors un sous-espace affine de $\mathcal{M}_{p,q}(\mathbb{R})$ dirigé par $\mathrm{Ker}\,\varphi$.

Soit $x \in \mathbb{R}^p$. Puisque a est non nul, il est évident qu'il existe une application linéaire $l \in \mathcal{L}(\mathbb{R}^q, \mathbb{R}^p)$ telle que $l(a) = x$ (pour en définir une il suffit par exemple de compléter a en une base de $\mathbb{R}^q$). Si on note Y la matrice de l dans les bases canoniques de $\mathbb{R}^q$ et $\mathbb{R}^p$, on a bien $\mathrm{Y}a = x$. Cela démontre que φ est surjective et que $\Omega(b)$ est non vide.

Il est alors clair que $\Omega(b)$ admet un élément de norme minimale : c'est tout simplement le projeté orthogonal de la matrice nulle sur $\Omega(b)$ qui réalise la distance de cette matrice nulle au sous-espace affine $\Omega(b)$. Notons-le $f(b)$. C'est le seul élément de $\Omega(b)$ qui se trouve dans $(\mathrm{Ker}\,\varphi)^\perp$. C'est ce dernier point qui va nous permettre d'expliciter $f(b)$.

Nous savons que $(\mathrm{Ker}\,\varphi)^\perp = \mathrm{Im}\,\varphi^*$, où $\varphi^* : \mathbb{R}^p \longrightarrow \mathcal{M}_{p,q}(\mathbb{R})$ est l'adjoint de φ. Déterminons φ^*. On notera comme d'habitude $\langle\ ,\ \rangle$ le produit scalaire canonique sur $\mathbb{R}^p$ identifié à $\mathcal{M}_{p,1}(\mathbb{R})$. Pour $(\mathrm{Y},\mathrm{Z}) \in \mathcal{M}_{p,q} \times \mathcal{M}_{p,1}$, on a

$$\begin{aligned}\Phi(\mathrm{Y},\varphi^*(\mathrm{Z})) &= \langle\varphi(\mathrm{Y}),\mathrm{Z}\rangle = \langle \mathrm{Y}a,\mathrm{Z}\rangle = {}^t(\mathrm{Y}a)\mathrm{Z}\\ &= \mathrm{Tr}({}^t a\,{}^t\mathrm{YZ}) = \mathrm{Tr}({}^t\mathrm{YZ}\,{}^t a) = \Phi(\mathrm{Y},\mathrm{Z}\,{}^t a).\end{aligned}$$

On en déduit que φ^* est définie par $\varphi^*(\mathrm{Z}) = \mathrm{Z}\,{}^t a$ pour tout $\mathrm{Z} \in \mathbb{R}^p$ et que $(\mathrm{Ker}\,\varphi)^\perp$ est l'ensemble des matrices de la forme $\mathrm{Z}\,{}^t a$, avec $\mathrm{Z} \in \mathbb{R}^p$.

Cherchons alors $\mathrm{Z} \in \mathbb{R}^p$ tel que $f(b) = \mathrm{Z}\,{}^t a$. On a, par définition,

$$\varphi(f(b)) = f(b)a = \mathrm{Z}\,{}^t aa = \|a\|^2\mathrm{Z}.$$

Ainsi $\varphi(f(b))$ est égal à $b - \mathrm{A}a$ si, et seulement si, $\mathrm{Z} = \dfrac{1}{\|a\|^2}(b - \mathrm{A}a)$, ce qui nous donne

$$\boxed{f(b) = \frac{1}{\|a\|^2}(b - \mathrm{A}a)\,{}^t a}.$$

3. Par définition du produit scalaire, on a pour tout $x \in \mathbb{R}^p$,

$$\|\phi(x)\|^2 = \|f(x+c)\|^2 + \|x\|^2.$$

Calculons $\|f(x+c)\|^2$, c'est-à-dire $\mathrm{Tr}({}^t f(x+c) f(x+c))$. On obtient

$$\begin{aligned} \|f(x+c)\|^2 &= \mathrm{Tr}\left(\frac{1}{\|a\|^2}\, a\, {}^t(c+x-\mathrm{A}a)\, \frac{1}{\|a\|^2}\,(c+x-\mathrm{A}a)\, {}^t a \right) \\ &= \mathrm{Tr}\left(\frac{1}{\|a\|^4}\, {}^t(c+x-\mathrm{A}a)(c+x-\mathrm{A}a)\, {}^t a a \right) \\ &= \frac{1}{\|a\|^2}\, \mathrm{Tr}({}^t(c+x-\mathrm{A}a)(c+x-\mathrm{A}a)). \end{aligned}$$

On a finalement, pour tout $x \in \mathbb{R}^p$.

$$\|f(x+c)\|^2 = \frac{1}{\|a\|^2}\|c+x-\mathrm{A}a\|^2$$

et donc,

$$\|\phi(x)\|^2 = \frac{1}{\|a\|^2}\|c+x-\mathrm{A}a\|^2 + \|x\|^2.$$

Un simple développement du premier carré nous donne

$$\|\phi(x)\|^2 = \frac{1}{\|a\|^2}\left((1+\|a\|^2)\|x\|^2 + 2\Phi(c-\mathrm{A}a, x) + \|c-\mathrm{A}a\|^2 \right),$$

$$\|\phi(x)\|^2 = \frac{1+\|a\|^2}{\|a\|^2}\left\| x + \frac{1}{1+\|a\|^2}(c-\mathrm{A}a) \right\|^2 + \frac{\|c-\mathrm{A}a\|^2}{1+\|a\|^2}.$$

Cela montre que $\|\phi(x)\|$ est minimal pour $\boxed{x = -\frac{1}{1+\|a\|^2}(c-\mathrm{A}a)}$. ◁

Les exercices suivants abordent le procédé d'orthonormalisation de Gram-Schmidt qui, à partir d'une suite libre, propose la construction d'une suite de vecteurs orthonormée. Précisons cela dans le cas d'une famille finie. On considère un espace réel E *muni d'un produit scalaire et* $(a_1, \ldots, a_p)$ *une famille libre de* E. *Il existe alors un unique système orthonormé* $(e_1, \ldots, e_p)$ *de* E *vérifiant les conditions suivantes :*

(i) $\forall k \in [\![1, p]\!]$, $\mathrm{Vect}(a_1, \ldots, a_k) = \mathrm{Vect}(e_1, \ldots, e_k)$;

(ii) $\forall k \in]\!]1, p]\!]$, $\langle a_k, e_k \rangle > 0$.

On démontre que le vecteur e_k *est nécessairement égal à* $\dfrac{\varepsilon_k}{\|\varepsilon_k\|}$ *avec*

$$\varepsilon_k = a_k - p_k(a_k) = a_k - \langle a_k, e_{k-1}\rangle e_{k-1} - \cdots - \langle a_k, e_1\rangle e_1,$$

où p_k *désigne la projection orthogonale sur le sous-espace* $\mathrm{Vect}(a_1, \ldots, a_{k-1}) = \mathrm{Vect}(e_1, \ldots, e_{k-1})$. *Ces formules permettent le calcul effectif des*

vecteurs e_k par récurrence. Il est bon de garder à l'esprit que lorsqu'on exprime a_k en fonction des e_i par une combinaison linéaire, $a_k = \lambda_k e_k + \lambda_1 e_1$, le coefficient $\lambda_k = \langle a_k, e_k \rangle$ est strictement positif. De même, si l'on pose $e_k = \mu_k a_k + \cdots + \mu_1 a_1$ (les μ_i étant réels), on a $\mu_k = \dfrac{1}{\lambda_k} > 0$.

1.20. Simplexes réguliers

Soit E un espace euclidien de dimension $n \geqslant 1$.

Une famille $(v_i)_{1 \leqslant i \leqslant n}$ est dite régulière si les v_i sont de norme 1 et si pour $1 \leqslant i < j \leqslant n$ on a $\|v_i - v_j\| = 1$.

1. Montrer que toute famille régulière est libre et qu'il existe des familles régulières.

2. Soit $(e_1, e_2, \ldots, e_n)$ la base obtenue par orthonormalisation de Schmidt à partir d'une famille régulière $(v_i)_{1 \leqslant i \leqslant n}$.

Montrer qu'il existe des réels $a_1, a_2, \ldots, a_n, b_1, b_2, \ldots, b_{n-1}$ tels que pour tout $i \in [\![1, n]\!]$, $v_i = a_i e_i + \sum_{j=1}^{i-1} b_j e_j$.

Calculer a_i pour $1 \leqslant i \leqslant n$ et b_i pour $1 \leqslant i \leqslant n-1$.

(École normale supérieure)

Solution.

1. Notons qu'une famille régulière correspond tout simplement à un simplexe régulier dont l'un des sommets est l'origine en regardant E comme un espace affine. Par exemple dans le plan, deux vecteurs (v_1, v_2) forment une famille régulière si, et seulement si, le triangle $(0, v_1, v_2)$ est équilatéral (on parle de 3-simplexe). Cet exercice démontrera en particulier l'existence de $(n+1)$-simplexes dans un espace de dimension n (voir aussi l'exercice 1.27).

Algébriquement, une famille $(v_i)_{1 \leqslant i \leqslant n}$ est régulière si, et seulement si,

$$\begin{cases} (i) & \|v_i\| = 1 \text{ pour } 1 \leqslant i \leqslant n; \\ (ii) & (v_i, v_j) = \dfrac{1}{2} \text{ pour } 1 \leqslant i < j \leqslant n. \end{cases}$$

• Montrons pour commencer que toute famille régulière est libre.

Si $(v_1, \ldots, v_n)$ est une famille régulière et si $(\alpha_1, \ldots, \alpha_n)$ appartient à $\mathbb{R}^n$, alors

$$\left\| \sum_{i=1}^{n} \alpha_i v_i \right\|^2 = \sum_{i=1}^{n} \alpha_i^2 \|v_i\|^2 + 2 \sum_{1 \leqslant i < j \leqslant n} \alpha_i \alpha_j (v_i, v_j) = \sum_{i=1}^{n} \alpha_i^2 + \sum_{1 \leqslant i < j \leqslant n} \alpha_i \alpha_j$$

et donc

$$\left\|\sum_{i=1}^{n}\alpha_i v_i\right\|^2 = \frac{1}{2}\sum_{i=1}^{n}\alpha_i^2 + \frac{1}{2}\left(\sum_{i=1}^{n}\alpha_i\right)^2.$$

On en déduit que $\sum_{i=1}^{n}\alpha_i v_i = 0$ implique $\sum_{i=1}^{n}\alpha_i^2 = 0$ et donc $\alpha_i = 0$ pour $1 \leqslant i \leqslant n$: la famille $(v_1, \ldots, v_n)$ est libre. On remarque que l'espace E étant supposé de dimension n, la famille $(v_1, \ldots, v_n)$ est une base de E.

• Montrons qu'il existe des familles régulières. Plusieurs approches sont possibles. On pourrait prouver l'existence d'une famille $(v_1, \ldots, v_n)$ ayant pour matrice de Gram la matrice contenant des 1 sur la diagonale et dont les autres coefficients valent tous $\frac{1}{2}$. Cette approche est suivie dans l'exercice 1.27. Nous allons ici donner une construction géométrique en raisonnant par récurrence sur la dimension n de E.

Si $n = 1$, un vecteur quelconque de norme 1 fait l'affaire. Supposons la propriété vérifiée pour un espace de dimension $n-1$. Soit E un espace euclidien de dimension n. Considérons un sous-espace E′ de E, de dimension $n-1$ et $(v_1, \ldots, v_{n-1})$ une famille régulière de E′. Soit f un vecteur de $(\mathrm{E}')^{\perp}$, de norme 1. Cherchons v_n tel que $(v_1, \ldots, v_{n-1}, v_n)$ soit une famille régulière de E sous la forme

$$v_n = \sum_{i=1}^{n-1}\lambda_i v_i + \alpha f,$$

où $\lambda_1, \ldots, \lambda_{n-1}, \alpha$ sont à déterminer. On a, pour $1 \leqslant j \leqslant n-1$,

$$(v_n, v_j) = \sum_{i=1}^{n-1}\lambda_i(v_i, v_j) = \lambda_j + \sum_{i\neq j}\frac{\lambda_i}{2}.$$

On veut, pour $1 \leqslant j \leqslant n-1$, $\langle v_n, v_j\rangle = \frac{1}{2}$, c'est-à-dire

$$\lambda_j + \sum_{i\neq j}\frac{\lambda_i}{2} = \frac{1}{2},$$

soit encore $\sum_{i=1}^{n-1}\lambda_i + \lambda_j = 1$.

Ce système linéaire admet une solution unique $\lambda_1 = \cdots = \lambda_{n-1} = \frac{1}{n}$ et on a alors

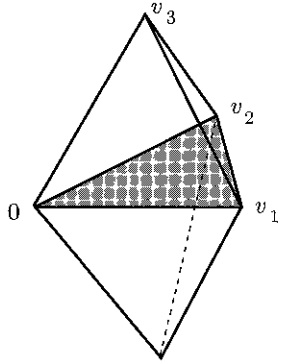
$$\|v_n\|^2 = \left\|\sum_{i=1}^{n-1}\lambda_i v_i\right\|^2 + \alpha^2 = \frac{1}{n^2}\left\|\sum_{i=1}^{n-1}v_i\right\|^2 + \alpha^2.$$

Sachant que la famille $(v_1, \ldots, v_{n-1})$ est régulière, on obtient

$$\left\|\sum_{i=1}^{n-1} v_i\right\|^2 = \sum_{i=1}^{n-1} \|v_i\|^2 + 2 \sum_{1\leqslant i<j\leqslant n-1} \langle v_i, v_j\rangle$$
$$= n-1+2\frac{(n-1)(n-2)}{2}\frac{1}{2} = \frac{n(n-1)}{2}.$$

La condition $\|v_n\| = 1$ équivaut donc à $\alpha^2 = \dfrac{n+1}{2n}$.

On peut donc déterminer $\lambda_1, \ldots, \lambda_{n-1}$ et α tels que $(v_1, \ldots, v_{n-1}, v_n)$ soit une famille régulière, ce qui termine la récurrence. On remarque que $(v_1, \ldots, v_{n-1})$ étant donné, $\lambda_1, \ldots, \lambda_{n-1}$ sont uniques et α est unique au signe près. Géométriquement, il y a seulement deux manières de compléter un n-simplexe régulier en un $(n+1)$-simplexe régulier (penser aux deux tétraèdres réguliers de $\mathbb{R}^3$ qu'on peut construire sur un triangle équilatéral donné du plan $\mathbb{R}^2$).



2. Soit $(e_1, \ldots, e_n)$ la base obtenue en appliquant le procédé d'orthonormalisation de Schmidt à $(v_1, \ldots, v_n)$.

• Par définition, il existe, pour tout i compris entre 1 et n, une famille de réels $(t_{ij})_{1\leqslant j\leqslant i}$ telle que $v_i = \sum_{j=1}^{i} t_{ij} e_j$ et $t_{ii} > 0$. Il s'agit de démontrer que t_{ij} à j fixé ne dépend pas de $i > j$.

Soient i et i' deux indices entre 1 et n, et k tel que $1 \leqslant k < i, i'$. On a, par définition d'une famille régulière, $\langle v_i, v_k\rangle = \langle v_{i'}, v_k\rangle = \dfrac{1}{2}$, et donc $(v_i - v_{i'}, v_k) = 0$. On en déduit que pour $1 \leqslant j < i, i'$, $v_i - v_{i'}$ appartient à $\mathrm{Vect}(v_1, \ldots, v_j)^\perp$. Or, par construction, $\mathrm{Vect}(v_1, \ldots, v_j)^\perp = \mathrm{Vect}(e_1, \ldots, e_j)^\perp$. On a donc $(v_i - v_{i'}, e_j) = 0$, $\langle v_i, e_j\rangle = \langle v_{i'}, e_j\rangle$, c'est-à-dire $t_{ij} = t_{i'j}$, si $i > j$ et $i' > j$.

Autrement dit, t_{ij}, pour $j < i$, est indépendant de i ; on peut le noter b_j si $1 \leqslant j < n$. En notant $a_i = t_{ii}$ pour $1 \leqslant i \leqslant n$, on obtient $2n-1$ réels $a_1, \ldots, a_n, b_1, \ldots, b_{n-1}$ tels que pour $1 \leqslant i \leqslant n$, $v_i = a_i e_i + \sum_{j=1}^{i-1} b_j e_j$.

• Calculons maintenant les a_i $(1 \leqslant i \leqslant n)$ et les b_i $(1 \leqslant i \leqslant n-1)$.

Pour $1 \leqslant i \leqslant n$, $\sum_{j=1}^{i-1} b_j e_j$ appartient à $\mathrm{Vect}(v_1, \ldots, v_{i-1})$ et e_i à $\mathrm{Vect}(v_1, \ldots, v_{i-1})^\perp$. Si l'on reprend la construction par récurrence de la question 1, on voit que a_i est égal au signe près au réel qui s'appelait α (en supposant ici que la dimension de E est i). On a donc $a_i^2 = \dfrac{i+1}{2i}$ et finalement, comme $a_i > 0$,

$$\boxed{a_i = \sqrt{\frac{i+1}{2i}}}.$$

Pour $1 \leqslant i \leqslant n-1$, on a

$$\langle v_i, v_{i+1} \rangle = \sum_{j=1}^{i-1} b_j^2 + a_i b_i = \frac{1}{2} \quad \text{et} \quad \|v_i\|^2 = \sum_{j=1}^{i-1} b_j^2 + a_i^2 = 1.$$

On en déduit que $a_i b_i - a_i^2 = -\dfrac{1}{2}$ et donc

$$b_i = \frac{a_i^2 - \frac{1}{2}}{a_i} = \frac{\frac{i+1}{2i} - \frac{1}{2}}{\sqrt{\frac{i+1}{2i}}} = \boxed{\frac{1}{\sqrt{2i(i+1)}}}. \quad \triangleleft$$

Une interprétation matricielle du procédé de Gram-Schmidt mène directement à la décomposition QR *d'une matrice inversible. Comme application directe, on obtient l'inégalité d'Hadamard qui majore le déterminant d'une matrice réelle. Tout cela est au programme de l'exercice suivant.*

1.21. Décomposition QR et inégalité d'Hadamard

1. *Décomposition* QR. Soit $A \in \mathrm{GL}_n(\mathbb{R})$. Montrer qu'il existe un unique couple (Q, R) avec $Q \in \mathrm{O}_n(\mathbb{R})$ et R triangulaire supérieure à coefficients diagonaux strictement positifs tel que $A = QR$.

2. Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ dont on note $C_1, \ldots, C_n$ les colonnes. Montrer que

$$|\det A| \leqslant \|C_1\| \ldots \|C_n\|$$

où $\|\ \|$ est la norme euclidienne de $\mathbb{R}^n$. Étudier le cas d'égalité.

3. En déduire, si l'on pose $\|A\|_\infty = \max_{1\leqslant i,j\leqslant n} |a_{ij}|$, une constante c, indépendante de n, telle que $|\det A| \leqslant c\|A\|_\infty^n n^{\frac{n}{2}}$ pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$.

(École polytechnique)

I - **Solution.**

1. Notons $\mathcal{B}_0 = (e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$ et $(C_1, \ldots, C_n)$ les colonnes de A, qui forment une base de $\mathbb{R}^n$ puisque A est inversible. Orthonormalisons la base $(C_1, \ldots C_n)$ au sens de Gram-Schmidt en une base orthonormale $(D_1, \ldots, D_n)$. Notons R la matrice de passage de $(D_1, \ldots, D_n)$ à $(C_1, \ldots, C_n)$. Par définition, cette matrice R est triangulaire supérieure, car $\mathrm{Vect}(C_1, \ldots, C_k) = \mathrm{Vect}(D_1, \ldots, D_k)$ pour tout k et de plus ses coefficients diagonaux sont strictement positifs, puisque le coefficient (k, k) vaut (C_k, D_k). Notons enfin Q la matrice de passage de $\mathcal{B}_0$ à $(D_1, \ldots, D_n)$. Comme ces deux dernières bases sont orthonormales, on a $Q \in O_n(\mathbb{R})$. Enfin A est simplement la matrice de passage de la base canonique $\mathcal{B}_0$ à la base $(C_1, \ldots, C_n)$ et on a donc la formule de changement de base $A = QR$. C'est le résultat voulu. L'unicité provient de l'unicité dans l'orthonormalisation de Gram-Schmidt, car si $A = Q'R'$ est une autre décomposition, il est facile de voir que les colonnes de Q' forment une base orthonormée de $\mathbb{R}^n$ qui est l'orthonormalisée de $(C_1, \ldots, C_n)$. Donc $Q' = Q$ puis $R' = R$.

Le lecteur trouvera dans le chapitre 2 une autre manière d'obtenir cette décomposition (cf. exercice 2.29). L'unicité peut aussi s'obtenir matriciellement : elle découle du fait que la seule matrice triangulaire supérieure avec des coefficients diagonaux positifs qui est orthogonale est l'identité.

2. L'inégalité est triviale si A n'est pas inversible. Supposons A inversible et considérons la décomposition de la question précédente $A = QR$. On a $|\det A| = |\det R| = \prod_{i=1}^{n} t_{ii}$ (les t_{ij} désignant les coefficients de R). Or, en reprenant les notations de la question 1, $t_{ii} = (C_i, D_i) \leqslant \|C_i\|$ par l'inégalité de Cauchy-Schwarz. Le résultat en découle.

Étudions le cas d'égalité. Si A n'est pas inversible il y a égalité si, et seulement si, l'une des colonnes de A est nulle. Sinon, il y a égalité si, et seulement si, C_i est colinéaire à D_i pour tout i. Cela signifie que la famille des colonnes de A est orthogonale.

Géométriquement, $|\det A|$ représente le volume dans $\mathbb{R}^n$ du parallélotope défini par les vecteurs colonnes de A. L'inégalité de Hadamard signifie que ce volume est plus petit que le produit des longueurs

des côtés et que, dans le cas où le parallélotope n'est pas aplati, il y a égalité si. et seulement si, le parallélotope est droit.

3. C'est une application directe de la question précédente. On a $\|C_k\| \leqslant \sqrt{n}\|A\|_\infty$ pour tout k et donc $|\det A| \leqslant n^{n/2}\|A\|_\infty^n$. C'est le résultat avec la constante $c = 1$.

En se plaçant dans $\mathbb{C}^n$ muni de sa structure hermitienne canonique, on montre par une argumentation tout à fait analogue que les majorations des questions 2 et 3 restent vraies pour les matrices de $\mathcal{M}_n(\mathbb{C})$.

Dans l'exercice suivant on applique l'orthonormalisation de Gram-Schmidt à une suite infinie : la base canonique de $\mathbb{R}[X]$.

1.22. Base orthonormale dans $\mathbb{R}[X]$

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels. Pour P et Q dans $\mathbb{R}[X]$, on pose

$$\langle P, Q\rangle = \sum_{n=0}^{+\infty} P^{(n)}(a_n)Q^{(n)}(a_n).$$

1. Montrer que l'on définit ainsi un produit scalaire. On note $(P_n)_{n\in\mathbb{N}}$ la base obtenue par orthonormalisation de Gram-Schmidt à partir de la base $(X^n)_{n\in\mathbb{N}}$.

2. Calculer $P_i^{(j)}(a_j)$ pour tout couple $(i, j) \in \mathbb{N}^2$.

3. Montrer que pour $x \in \mathbb{R}$ et $n \in \mathbb{N}^*$,

$$P_n(x) = \int_{a_0}^{x} dt_1 \int_{a_1}^{t_1} dt_2 \cdots \int_{a_{n-1}}^{t_{n-1}} dt_n.$$

4. Calculer P_n lorsque $a_n = n\alpha$ pour tout $n \in \mathbb{N}$ où $\alpha \in \mathbb{R}$ est fixé.

(**École polytechnique**)

▷ **Solution.**

1. Notons d'abord que si $(P, Q) \in \mathbb{R}[X]^2$, $\langle P, Q\rangle$ est bien défini puisque dans la série $\sum_{n=0}^{+\infty} P^{(n)}(a_n)Q^{(n)}(a_n)$, il n'y a qu'un nombre fini de termes non nuls (car pour n assez grand, $P^{(n)} = Q^{(n)} = 0$). La bilinéarité est une conséquence de la linéarité de la dérivation et la symétrie est claire. Enfin, si $P \in \mathbb{R}[X]$, on a

$$\langle P, P\rangle = \sum_{n=0}^{+\infty} P^{(n)}(a_n)^2 \geqslant 0.$$

De plus, si $P \neq 0$ et $p = \deg P$, en notant α son coefficient dominant on obtient

$$\langle P, P\rangle \geqslant P^{(p)}(a_p)^2 = \alpha^2(p!)^2 > 0.$$

Il s'agit donc d'un produit scalaire.

2. Comme pour $i \in \mathbb{N}$, $\mathrm{Vect}(P_0, \ldots, P_i) = \mathrm{Vect}(1, X, \ldots, X^i) = \mathbb{R}_i[X]$ et $\mathrm{Vect}(P_0, \ldots, P_{i-1}) = \mathbb{R}_{i-1}[X]$, le degré de P_i vaut nécessairement i. Ainsi, P_0 est une constante c ; on doit avoir $\langle P_0, P_0\rangle = c^2 = 1$ et $\langle P_0, 1\rangle = c > 0$, donc $P_0 = 1$. Soit maintenant $i \in \mathbb{N}^*$.

- Si $j > i$ on a, pour des raisons de degré, $P_i^{(j)} = 0$ et $P_i^{(j)}(a_j) = 0$.
- Montrons par récurrence sur $j \in [\![0, i-1]\!]$ que $P_i^{(j)}(a_j) = 0$.

On a $\langle P_i, 1\rangle = 0$ *i.e.* $P_i(a_0) \times 1 = 0$. C'est vérifié pour $j = 0$. Supposons le résultat vrai jusqu'au rang $j-1$ avec $1 \leqslant j \leqslant i-1$. On a par hypothèse de récurrence

$$0 = \langle P_i, X^j\rangle = \sum_{k=0}^{j} \underbrace{P_i^{(k)}(a_k)}_{\text{nul si } k<j} \frac{j!}{(j-k)!} a_k^{j-k} = P_i^{(j)}(a_j) j!.$$

On obtient donc bien $P_i^{(j)}(a_j) = 0$.

- Calculons enfin $P_i^{(i)}(a_i)$. Compte-tenu de ce qui précède, on a d'une part $0 < \langle P_i, X^i\rangle = P_i^{(i)}(a_i) i!$ et d'autre part

$$1 = \langle P_i, P_i\rangle = P_i^{(i)}(a_i)^2.$$

Comme $P_i^{(i)}(a_i) > 0$, il en résulte que $P_i^{(i)}(a_i) = 1$.

On conclut que pour tout $(i, j) \in \mathbb{N}^2$, $P_i^{(j)}(a_j) = 1$ si $i = j$ et 0 sinon.

3. Posons pour $x \in \mathbb{R}$ et $n \geqslant 1$,

$$Q_n(x) = \int_{a_0}^{x} dt_1 \int_{a_1}^{t_1} dt_2 \cdots \int_{a_{n-1}}^{t_{n-1}} dt_n.$$

Il est clair que Q_n est une fonction polynomiale de degré n (il s'agit d'une intégration successive de fonctions polynômes). Montrons par récurrence sur $n \geqslant 1$ que, pour toute suite $(a_j)_{j\in\mathbb{N}}$, et pour tout entier $j \in \mathbb{N}$, $Q_n^{(j)}(a_j) = \delta_{nj}$. Pour des raisons de degré, c'est déjà vrai pour tout n lorsque $j > n$. On se préoccupera donc seulement des indices $j \leqslant n$ dans la récurrence. On a $Q_1 : x \longmapsto \int_{a_0}^{x} dt_1 = x - a_0$. On a bien $Q_1(a_0) = 0$ et $Q_1'(a_1) = 1$. C'est valable pour $n = 1$.

Supposons $n \geqslant 2$ et le résultat vrai au rang $n-1$. On a déjà $Q_n(a_0) = 0$. En dérivant Q_n, on obtient

$$Q_n'(x) = \int_{a_1}^{x} dt_2 \cdots \int_{a_{n-1}}^{t_{n-1}} dt_n.$$

En appliquant l'hypothèse de récurrence à Q'_n, avec la suite $(a_{j+1})_{j\in\mathbb{N}}$, il vient

$$Q'_n(a_1) = Q''_n(a_2) = \cdots = Q'^{(n-2)}_n(a_{n-1}) = 0 \quad \text{et} \quad Q'^{(n-1)}_n(a_n) = 1,$$

ce qui s'écrit

$$Q'_n(a_1) = Q''_n(a_2) = \cdots = Q^{(n-1)}_n(a_{n-1}) = 0 \quad \text{et} \quad Q^{(n)}_n(a_n) = 1.$$

La récurrence est achevée. Comme pour tout $j \in \mathbb{N}$, $Q^{(j)}_n(a_j) = P^{(j)}_n(a_j)$, il s'ensuit que $(P_n - Q_n, P_n - Q_n\rangle = \sum_{j=0}^{+\infty} (P_n - Q_n)^{(j)}(a_j)^2 = 0$ et donc $P_n = Q_n$.

4. Retenons de la question précédente qu'un polynôme $P \in \mathbb{R}[X]$ est parfaitement déterminé par la suite $P^{(n)}(a_n)$. Il s'agit de trouver P_n de degré n vérifiant $P_n(0) = P'_n(\alpha) = \cdots = P^{(n-1)}_n((n-1)\alpha) = 0$ et $P^{(n)}_n(n\alpha) = 1$.

Soit $n \in \mathbb{N}$. Notons $Q = P'_n(X + \alpha)$. Dans ces conditions, on a

$$Q(0) = P'_n(\alpha) = 0,\ Q'(\alpha) = P''_n(2\alpha) = 0, \ldots$$

$$Q^{(n-2)}((n-2)\alpha) = P^{(n-1)}_n((n-1)\alpha) = 0,$$

$$Q^{(n-1)}((n-1)\alpha) = P^{(n)}_n(n\alpha) = 1,$$

et si $k \geqslant n$, $Q^{(k)}(k\alpha) = 0$, car $\deg Q = n-1$. Par conséquent, on obtient

$$Q = P'_n(X + \alpha) = P_{n-1}, \text{ ou encore } P'_n(X) = P_{n-1}(X - \alpha).$$

Cette relation permet un calcul rapide des premiers termes ($P_1 = X$, $P_2 = \dfrac{X(X-2\alpha)}{2}$ et $P_3 = \dfrac{X(X-3\alpha)^2}{6}$) qui nous invite à prouver par récurrence que

$$P_n = \frac{X(X - n\alpha)^{n-1}}{n!}.$$

On a bien $P_1 = X$. Supposons $n \geqslant 2$. Notons $P = \dfrac{X(X-n\alpha)^{n-1}}{n!}$. On a

$$\begin{aligned} P' &= \frac{X(X-n\alpha)^{n-2}}{n(n-2)!} + \frac{(X-n\alpha)^{n-1}}{n!} \\ &= \frac{(X-n\alpha)^{n-2}}{n!}[(n-1)X + (X-n\alpha)] \\ &= \frac{n(X-\alpha)(X-n\alpha)^{n-2}}{n!} = \frac{(X-\alpha)(X-n\alpha)^{n-2}}{(n-1)!}. \end{aligned}$$

Donc, on a $P'(X+\alpha) = \dfrac{X(X-(n-1)\alpha)^{n-2}}{(n-1)!} = P_{n-1}$ par hypothèse de récurrence. On en déduit $P' = P_{n-1}(X-\alpha) = P'_n$. Comme $P_n(0) = 0 = P(0)$, on conclut que $P_n = P = \dfrac{X(X-n\alpha)^{n-1}}{n!}$. ◁

La notion de polynômes orthogonaux apparaît à l'occasion de certains problèmes d'analyse fonctionnelle (équations intégrales, problème de Sturm-Liouville, équations aux dérivées partielles...). Ils constituent souvent des bases adaptées à l'étude d'opérateurs symétriques (ou hermitiens). La terminologie est réservée à une suite orthogonale de polynômes $(P_n)_{n\in\mathbb{N}}$ avec $\deg P_n = n$, $\mathbb{R}[X]$ étant vu comme partie d'un espace de fonctions $\mathcal{E}$ muni d'un produit scalaire pour lequel la multiplication par X est un opérateur symétrique. C'est notamment le cas pour les produits scalaires du type

$$(f, g) \longmapsto \langle f, g\rangle = \int_I f(t)g(t)\omega(t)\,dt,$$

où $\omega : I \longrightarrow \mathbb{R}$ est une fonction positive continue non nulle appelée « poids ». Bien évidemment, la suite $(P_n)_{n\in\mathbb{N}}$ dépend du produit scalaire. Pour $I = [-1, 1]$ et $\omega = 1$, on trouve les polynômes de Legendre présentés dans l'exercice 1.21 du tome 2 d'analyse. En prenant $I =]-1, 1[$ et $\omega : t \longmapsto \dfrac{1}{\sqrt{1-t^2}}$, on retrouve les polynômes de Tchebychev T_n vus dans l'exercice 5.36 du premier tome d'algèbre. En effet, si $n \neq m$, on a

$$\begin{aligned}\int_{-1}^{1} \frac{T_m(t)T_n(t)}{\sqrt{1-t^2}}\,dt &= \int_0^{\pi} T_m(\cos x)T_n(\cos x)\,dx \\ &= \int_0^{\pi} \cos(mx)\cos(nx)\,dx = 0,\end{aligned}$$

en ayant effectué le changement de variables $t = \cos x$. Les deux exercices suivants développent l'exemple des polynômes de Laguerre ($I = \mathbb{R}_+$ et $\omega : t \longmapsto e^{-t}$), puis les polynômes d'Hermite ($I = \mathbb{R}$ et $\omega : t \longmapsto e^{-t^2}$).

1.23. Polynômes de Laguerre

Soit $\mathcal{E}$ l'espace des fonctions continues $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$ telles que $x \longmapsto f(x)^2e^{-x}$ soit intégrable sur $\mathbb{R}_+$. On munit $\mathcal{E}$ du produit scalaire

$$\langle f, g\rangle = \int_{\mathbb{R}_+} f(x)g(x)e^{-x}\,dx \quad \text{pour } f, g \in \mathcal{E}.$$

1. Montrer l'existence d'une base $(L_n)_{n\in\mathbb{N}}$ de $\mathbb{R}[X]$ orthonormée pour $\langle\ ,\ \rangle$ telle que pour tout n, L_n est de degré n et de coefficient dominant strictement positif.

2. Soit $n \in \mathbb{N}^*$. Montrer l'existence de a_n, b_n, c_n dans $\mathbb{R}$ tels que

$$XL_n(X) = a_nL_{n+1}(X) + b_nL_n(X) + c_nL_{n-1}(X).$$

3. Montrer que si $(x, y) \in \mathbb{R}^2$ et $n \in \mathbb{N}$,

$$(x-y)\sum_{i=0}^{n} \mathrm{L}_i(x)\mathrm{L}_i(y) = a_n(\mathrm{L}_{n+1}(x)\mathrm{L}_n(y) - \mathrm{L}_{n+1}(y)\mathrm{L}_n(x)).$$

C'est la *formule de Darboux-Christoffel.*

4. Calculer $\inf_{(a,b,c)\in\mathbb{R}^3} \int_{\mathbb{R}_+} (x^3 - ax^2 - bx - c)^2 e^{-x}\mathrm{d}x.$

(École polytechnique)

▷ **Solution.**

Lors de l'oral, le candidat justifiera rapidement que $\mathcal{E}$ est un espace vectoriel et que l'on a bien défini un produit scalaire. En effet, pour $f, g \in \mathcal{E}$, l'application $x \longmapsto f(x)g(x)e^{-x}$ est intégrable sur $\mathbb{R}_+$, puisque si $x \geqslant 0$,

$$\begin{aligned} |f(x)g(x)e^{-x}| &\leqslant \frac{1}{2}\left(\left(f(x)e^{-x/2}\right)^2 + \left(g(x)e^{-x/2}\right)^2\right) \\ &\leqslant \frac{1}{2}\left(f(x)^2e^{-x} + g(x)^2e^{-x}\right). \end{aligned}$$

Comme $(f(x)+g(x))^2e^{-x} = f(x)^2e^{-x} + 2f(x)g(x)e^{-x} + g(x)^2e^{-x}$, l'application $f+g$ est dans $\mathcal{E}$. Puisqu'il en va de même pour λf et l'application nulle, $\mathcal{E}$ est un sous-espace de $\mathcal{C}(\mathbb{R}_+, \mathbb{R})$. D'après ce qui précède, $\langle\ ,\ \rangle$ est bien définie. C'est clairement une forme bilinéaire symétrique (par linéarité de l'intégrale). Comme pour f non nulle continue, on a $\int_{\mathbb{R}_+} f(x)^2e^{-x}\mathrm{d}x > 0$, nous avons bien défini un produit scalaire. Notons enfin que $\mathcal{E}$ contient les fonctions polynomiales car si $\mathrm{P} \in \mathbb{R}[\mathrm{X}]$, $\mathrm{P}(x)^2e^{-x} = o(e^{-x/2})$ pour x tendant vers $+\infty$ par croissance comparée et le théorème de comparaison des fonctions positives intégrables assure l'intégrabilité de $x \longmapsto \mathrm{P}(x)^2e^{-x}$ sur $\mathbb{R}_+$.

1. Considérons l'orthonormalisé au sens de Gram-Schmidt de la base $(1, \mathrm{X}, \ldots, \mathrm{X}^n, \ldots)$. On obtient alors un système orthonormé $(\mathrm{L}_0, \mathrm{L}_1, \ldots, \mathrm{L}_n, \ldots)$ qui vérifie pour $n \in \mathbb{N}^*$

$$\mathrm{Vect}(\mathrm{L}_0, \mathrm{L}_1, \ldots, \mathrm{L}_{n-1}) = \mathrm{Vect}(1, \mathrm{X}, \ldots, \mathrm{X}^{n-1}) = \mathbb{R}_{n-1}[\mathrm{X}]$$

et $\mathrm{Vect}(\mathrm{L}_0, \mathrm{L}_1, \ldots, \mathrm{L}_n) = \mathbb{R}_n[\mathrm{X}]$. Nécessairement, le polynôme L_n est de degré n. D'après les commentaires figurant page 37, le coefficient dominant de L_n est strictement positif. Comme L_n est de degré n, il est classique de conclure que la famille $(\mathrm{L}_n)_{n\in\mathbb{N}}$ est une base de $\mathbb{R}[\mathrm{X}]$.

2. Comme $\mathrm{XL}_n \in \mathbb{R}_{n+1}[\mathrm{X}]$, il existe des réels $\lambda_0, \lambda_1, \ldots, \lambda_{n+1}$ tels que

$$\mathrm{XL}_n = \lambda_0 \mathrm{L}_0 + \lambda_1 \mathrm{L}_1 + \cdots + \lambda_{n+1}\mathrm{L}_{n+1}.$$

Prenons $k \in [\![0, n-2]\!]$. On a $\langle \mathrm{XL}_n, \mathrm{L}_k\rangle = \lambda_k$ d'une part et d'autre part,

$$\langle \mathrm{XL}_n, \mathrm{L}_k\rangle = \int_{\mathbb{R}_+} x\mathrm{L}_n(x)\mathrm{L}_k\langle x\rangle e^{-x}\mathrm{d}x = \langle \mathrm{L}_n, \mathrm{XL}_k\rangle = 0,$$

car $\mathrm{XL}_k \in \mathbb{R}_{n-1}[\mathrm{X}] = \mathrm{Vect}\langle \mathrm{L}_0, \ldots, \mathrm{L}_{n-1}\rangle$ et L_n est orthogonal à chaque L_k avec $k < n$. Ainsi, $\lambda_0 = \lambda_1 = \cdots = \lambda_{n-2} = 0$ et le résultat est démontré.

Notons que la relation peut s'écrire pour $n = 0$ en posant $\mathrm{L}_{-1} = 0$ et c_0 quelconque. Retenons également que le fait que XL_n soit combinaison linéaire de L_{n-1}, L_n et L_{n+1}, est une propriété tout à fait générale dès que $\langle \mathrm{XP}, \mathrm{Q}\rangle = \langle \mathrm{P}, \mathrm{XQ}\rangle$, c'est-à-dire lorsque la multiplication par X est un endomorphisme symétrique de $\mathbb{R}[\mathrm{X}]$.

3. Soit $n \geqslant 1$ et $x, y \in \mathbb{R}$. Posons $\Delta_n = (x-y)\sum_{i=0}^{n} \mathrm{L}_i(x)\mathrm{L}_i(y)$. On va essayer de trouver une relation de récurrence. D'après la question précédente, la différence $\Delta_n - \Delta_{n-1} = x\mathrm{L}_n(x)\mathrm{L}_n\langle y) - y\mathrm{L}_n(x)\mathrm{L}_n(y)$ vaut

$$(a_n\mathrm{L}_{n+1}(x)+b_n\mathrm{L}_n(x)+c_n\mathrm{L}_{n-1}(x))\mathrm{L}_n(y)-(a_n\mathrm{L}_{n+1}(y)+b_n\mathrm{L}_n(y)+c_n\mathrm{L}_{n-1}(y))\mathrm{L}_n(x)$$

ce qui se simplifie en

$$a_n(\mathrm{L}_{n+1}(x)\mathrm{L}_n\langle y) - \mathrm{L}_{n+1}(y)\mathrm{L}_n(y)\rangle - c_n(\mathrm{L}_n(x)\mathrm{L}_{n-1}(y) - \mathrm{L}_{n-1}(x)\mathrm{L}_n(y)).$$

Si $c_n = a_{n-1}$, il sera facile de conclure puisqu'on a une somme télescopique. Vérifions cette conjecture. Considérons $\langle \mathrm{XL}_n, \mathrm{L}_{n-1}\rangle$. D'une part on a

$$\langle \mathrm{XL}_n, \mathrm{L}_{n-1}\rangle = \langle a_n\mathrm{L}_{n+1} + b_n\mathrm{L}_n + c_n\mathrm{L}_{n-1}, \mathrm{L}_{n-1}\rangle = c_n,$$

mais on a aussi $\langle \mathrm{XL}_n, \mathrm{L}_{n-1}\rangle = \langle \mathrm{L}_n, \mathrm{XL}_{n-1}\rangle$, ce qui donne,

$$\langle \mathrm{XL}_n, \mathrm{L}_{n-1}\rangle = (\mathrm{L}_n, a_{n-1}\mathrm{L}_n + b_{n-1}\mathrm{L}_{n-1} + c_{n-1}\mathrm{L}_{n-2}\rangle = a_{n-1}.$$

On obtient bien ce qu'on voulait : $a_{n-1} = c_n$. La formule est donc démontrée par récurrence dès lors qu'elle est valable pour $n = 0$. Il s'agit alors de montrer que

$$(x-y)\mathrm{L}_0(x)\mathrm{L}_0(y) = a_0(\mathrm{L}_1(x)\mathrm{L}_0(y) - \mathrm{L}_1(y)\mathrm{L}_0(x)\rangle$$

ou encore que $(x-y)\mathrm{L}_0 = a_0(\mathrm{L}_1(x) - \mathrm{L}_1(y))$, puisque L_0 est un polynôme constant non nul. Comme L_1 est affine et $\mathrm{XL}_0 = a_0\mathrm{L}_1 + b_0\mathrm{L}_0$, la relation est effectivement vérifiée : si α est le coefficient dominant de L_1, on a $a_0\alpha = \mathrm{L}_0$ et on a bien $(x-y)\mathrm{L}_0 = a_0\alpha(x-y)$.

4. Notons $f : x \longmapsto x^3$. La borne inférieure I à calculer n'est autre que le carré de la distance de f à $\mathbb{R}_2[X]$. Nous savons que (L_0, L_1, L_2) étant une base orthonormale de $\mathbb{R}_2[X]$, on a

$$I = \|f\|^2 - \langle f, L_0\rangle^2 - \langle f, L_1\rangle^2 - \langle f, L_2\rangle^2,$$

et ce, en vertu du théorème de Pythagore, puisque $\langle f, L_0\rangle L_0 + \langle f, L_1\rangle L_1 + \langle f, L_2\rangle L_2$ est le projeté orthogonal de f sur $\mathbb{R}_2[X]$.

Sachant que, pour tout $n \in \mathbb{N}$, $\displaystyle\int_{\mathbb{R}_+} x^n e^{-x} dx = n!$, calculons L_0, L_1 et L_2.

On a $\boxed{L_0 = 1}$ et si l'on écrit $L_1 = \alpha X + \beta$, la relation $\langle L_0, L_1\rangle = 0$ entraîne $\alpha + \beta = 0$. On a $L_1 = \alpha(X - 1)$ de norme 1 :

$$\|L_1\|^2 = \alpha^2 \int_{\mathbb{R}_+} (x-1)^2 e^{-x} dx = \alpha^2(2! - 2.1! + 1.0!) = \alpha^2.$$

Comme $\alpha > 0$, $\alpha = 1$ et on a $\boxed{L_1 = X - 1}$.

On écrit ensuite $L_2 = aX^2 + bL_1 + cL_0$ et on écrit les relations d'orthogonalité :

$$\langle L_2, L_0\rangle = a\langle X^2, L_0\rangle + c = 2a + c = 0,$$
$$\langle L_2, L_1\rangle = a\langle X^2, L_1\rangle + b = (6-2)a + b = 4a + b = 0.$$

Ainsi, $c = -2a$ et $b = -4a$. Enfin,

$$\|aX^2\|^2 = 24a^2 = 1 + b^2 + c^2 = 1 + 16a^2 + 4a^2$$

donne $a = \dfrac{1}{2}$ car $a > 0$. Ainsi, on a $\boxed{L_2 = \dfrac{1}{2}X^2 - 2X - 1}$. On a enfin

$$I = 6! - \left(\frac{1}{2}.5! - 2.4! - 3!\right)^2 - (4! - 3!)^2 - 3!^2 = 324. \lhd$$

Les polynômes qui portent le nom d'Hermite ont surtout été étudiés par Laplace lors de travaux en Probabilités. Hermite en a essentiellement donné une généralisation à plusieurs variables.

1.24. Polynômes d'Hermite

On pose pour x réel et $n \in \mathbb{N}$,

$$H_n(x) = (-1)^n e^{x^2} \frac{d^n e^{-x^2}}{dx^n}.$$

1. Montrer que H_n est une fonction polynomiale de degré n.
2. Montrer que $\mathrm{H}_{n+2} = 2\mathrm{X}\mathrm{H}_{n+1} - 2(n+1)\mathrm{H}_n$.
3. Montrer que $\mathrm{H}_n'' - 2\mathrm{X}\mathrm{H}_n' + 2n\mathrm{H}_n = 0$.
4. On pose $q_n(x) = \mathrm{H}_n(x)e^{-x^2/2}$ pour $x \in \mathbb{R}$. Montrer que

$$q_n'' + (2n + 1 - x^2)q_n = 0.$$

5. Calculer $\int_{\mathbb{R}} q_m(t)q_n(t)\mathrm{d}t$ pour $m, n \in \mathbb{N}$.

(École polytechnique)

Solution.

1. Vérifions cela par récurrence sur $n \in \mathbb{N}$. Pour $n = 0$, on a $\mathrm{H}_0 = 1$. Pour $n \geqslant 1$ et x réel, on peut écrire

$$\begin{aligned}
\mathrm{H}_n(x) &= (-1)^n \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\mathrm{d}^{n-1}e^{-x^2}}{\mathrm{d}x^{n-1}}\right)e^{x^2} \\
&= (-1)^n \frac{\mathrm{d}}{\mathrm{d}x}\left((-1)^{n-1}\mathrm{H}_{n-1}(x)e^{-x^2}\right)e^{x^2} \\
&= -\left(\mathrm{H}_{n-1}'(x)e^{-x^2} - 2x\mathrm{H}_{n-1}(x)e^{-x^2}\right)e^{x^2} \\
&= 2x\mathrm{H}_{n-1}(x) - \mathrm{H}_{n-1}'(x).
\end{aligned}$$

Si H_{n-1} est polynomiale de degré $n-1$ et de coefficient dominant 2^{n-1}, la relation que nous venons d'établir nous assure que H_n est polynomiale de degré n et de coefficient dominant 2^n. Cela termine la récurrence.

2. Nous avons obtenu la relation $\mathrm{H}_{n+1} = 2\mathrm{X}\mathrm{H}_n - \mathrm{H}_n'$ pour n entier naturel. Il s'agit de montrer que pour $n \geqslant 1$, $\mathrm{H}_{n+1} = 2\mathrm{X}\mathrm{H}_n - 2n\mathrm{H}_{n-1}$, autrement dit que $\mathrm{H}_n' = 2n\mathrm{H}_{n-1}$. Vérifions cela par récurrence. Pour $n = 1$, comme $\mathrm{H}_0 = 1$ et $\mathrm{H}_1 = 2\mathrm{X}$, l'identité est vérifiée. Supposons $n \geqslant 2$ et le résultat vrai au rang $n-1$. Comme $\mathrm{H}_n = 2\mathrm{X}\mathrm{H}_{n-1} - \mathrm{H}_{n-1}'$, on obtient en dérivant,

$$\begin{aligned}
\mathrm{H}_n' &= 2\mathrm{X}\mathrm{H}_{n-1}' + 2\mathrm{H}_{n-1} - \mathrm{H}_{n-1}'' \\
&= 4\mathrm{X}(n-1)\mathrm{H}_{n-2} + 2\mathrm{H}_{n-1} - 2(n-1)\mathrm{H}_{n-2}',
\end{aligned}$$

en ayant utilisé l'hypothèse de récurrence $\mathrm{H}_{n-1}' = 2(n-1)\mathrm{H}_{n-2}$ qui donne $\mathrm{H}_{n-1}'' = 2(n-1)\mathrm{H}_{n-2}'$. Ensuite, on écrit

$$\begin{aligned}
\mathrm{H}_n' &= 2(n-1)(2\mathrm{X}\mathrm{H}_{n-2} - \mathrm{H}_{n-2}') + 2\mathrm{H}_{n-1} \\
&= (2n-2)\mathrm{H}_{n-1} + 2\mathrm{H}_{n-1} = 2n\mathrm{H}_{n-1}.
\end{aligned}$$

On conclut donc que pour tout $n \geqslant 1$,

$$\boxed{\mathrm{H}_n' = 2n\mathrm{H}_{n-1}} \quad \text{et} \quad \boxed{\mathrm{H}_{n+1} = 2\mathrm{X}\mathrm{H}_n - 2n\mathrm{H}_{n-1}}.$$

3. Soit $n \geqslant 2$. D'après ce qui précède, on a

$$\begin{aligned} \mathrm{H}_n'' - 2\mathrm{X}\mathrm{H}_n' + 2n\mathrm{H}_n &= 2n\mathrm{H}_{n-1}' - 4\mathrm{X}n\mathrm{H}_{n-1} + 2n\mathrm{H}_n \\ &= 2n(2(n-1)\mathrm{H}_{n-2}) - 4n\mathrm{X}\mathrm{H}_{n-1} + 2n\mathrm{H}_n \\ &= 2n(\mathrm{H}_n - 2\mathrm{X}\mathrm{H}_{n-1} - 2(n-1)\mathrm{H}_{n-2}) = 0. \end{aligned}$$

Comme la relation est vraie pour $n = 0$ et $n = 1$, on a bien pour tout $n \in \mathbb{N}$,

$$\boxed{\mathrm{H}_n'' - 2\mathrm{X}\mathrm{H}_n' + 2n\mathrm{H}_n = 0}.$$

4. Soit $n \in \mathbb{N}$ et $x \in \mathbb{R}$. On a

$$\begin{aligned} q_n''(x) &= \mathrm{H}_n''(x)e^{-x^2/2} - 2xe^{-x^2/2}\mathrm{H}_n'(x) + \mathrm{H}_n(x)(x^2-1)e^{-x^2/2} \\ &= -2n\mathrm{H}_n(x)e^{-x^2/2} + \mathrm{H}_n(x)(x^2-1)e^{-x^2/2} \\ &= (x^2 - 1 - 2n)q_n(x). \end{aligned}$$

C'est ce qu'on voulait.

5. Il s'agit de calculer $\mathrm{I}_{m,n} = \int_{\mathbb{R}} e^{-t^2}\mathrm{H}_m(t)\mathrm{H}_n(t)\mathrm{d}t$. Remarquons avant toute chose que cette intégrale est bien définie, car la fonction $t \longmapsto e^{-t^2}\mathrm{H}_m(t)\mathrm{H}_n(t)$ est intégrable sur $\mathbb{R}$, puisque négligeable en $\pm\infty$ devant $t \longmapsto e^{-t^2/2}$. Supposons $n \leqslant m$. La fonction $t \longmapsto e^{-t^2}\mathrm{H}_m(t)$ admet comme primitive $t \longmapsto (-1)^m \dfrac{\mathrm{d}^{m-1}e^{-t^2}}{\mathrm{d}t^{m-1}} = -\mathrm{H}_{m-1}(t)e^{-t^2}$. Par intégration par parties, il en résulte

$$\int_{-x}^{x} e^{-t^2}\mathrm{H}_m(t)\mathrm{H}_n(t)\mathrm{d}t = \Big[-\mathrm{H}_{m-1}(t)e^{-t^2}\mathrm{H}_n(t)\Big]_{-x}^{x} + \int_{-x}^{x} e^{-t^2}\mathrm{H}_{m-1}(t)\mathrm{H}_n'(t)\mathrm{d}t,$$

ce qui donne en faisant tendre x vers $+\infty$,

$$\mathrm{I}_{m,n} = \int_{\mathbb{R}} e^{-t^2}\mathrm{H}_{m-1}(t)\mathrm{H}_n'(t)\mathrm{d}t.$$

En intégrant par parties m fois au total, il restera

$$\mathrm{I}_{m,n} = \int_{\mathbb{R}} e^{-t^2}\mathrm{H}_0(t)\mathrm{H}_n^{(m)}(t)\mathrm{d}t = \int_{\mathbb{R}} \mathrm{H}_n^{(m)}(t)e^{-t^2}\mathrm{d}t.$$

Si $m > n$, la dérivée d'ordre m de H_n est nulle et $\mathrm{I}_{m,n} = 0$. Si $m = n$, $\mathrm{H}_n^{(n)}(t) = 2^n n!$ et $\mathrm{I}_{n,n} = 2^n n!\sqrt{\pi}$. ◁

Si on considère $\mathcal{E}$ l'ensemble des fonctions f continues de $\mathbb{R}$ dans $\mathbb{R}$ telles que $x \longmapsto f(x)^2e^{-x^2}$ est intégrable sur $\mathbb{R}$, on reconnaît là encore une suite de polynômes orthogonaux pour le produit scalaire défini pour f, g dans $\mathcal{E}$ par

$$\langle f, g\rangle = \int_{\mathbb{R}} f(x)g(x)e^{-x^2}\,dx.$$

La relation de récurrence linéaire de la deuxième question est à rapprocher du commentaire de la solution de la seconde question de l'exercice 1.23.

1.25. Méthode de Gauss

On considère le produit scalaire sur $\mathbb{R}[X]$

$$(P, Q) \longmapsto \langle P, Q\rangle = \int_0^1 P(t)Q(t)dt.$$

1. Montrer qu'il existe une suite orthonormée $(P_n)_{n\in\mathbb{N}}$ d'éléments de $\mathbb{R}[X]$ telle que $\deg P_n = n$ pour tout n.

2. Montrer que P_n possède n racines simples situées toutes dans $]0, 1[$.

3. On fixe $n \in \mathbb{N}^*$. Soit $\alpha_1, \ldots, \alpha_n$ les racines de P_n. Montrer qu'il existe des réels $\lambda_1, \ldots, \lambda_n$ tels que, pour tout $P \in \mathbb{R}_{2n-1}[X]$, on ait

$$\int_0^1 P(t)dt = \sum_{i=1}^{n} \lambda_i P(\alpha_i).$$

Donner une expression de ces scalaires et leur signe.

(**École polytechnique**)

Solution.

1. Cette question classique se traite comme la première question de l'exercice 1.23.

2. Comme $\mathrm{Vect}(P_0, P_1, \ldots, P_{n-1}) = \mathbb{R}_{n-1}[X]$, P_n est orthogonal à tout polynôme Q de degré inférieur ou égal à $n-1$. Notons $\alpha_1 < \cdots < \alpha_p$ les racines distinctes de P_n se situant dans $]0, 1[$. Il s'agit de montrer que $p = n$. On note n_i l'ordre de α_i comme racine de P_n et on considère

$$A = \prod_{\substack{1\leqslant i\leqslant p \\ n_i \text{ impair}}} (X - \alpha_i).$$

Alors, le polynôme AP_n garde un signe constant au sens large sur $[0, 1]$ et comme il est continu et non nul, $\int_0^1 A(t)P_n(t)dt$ est non nulle. Il s'ensuit que $\deg A \geqslant n$ et nécessairement $p = n$. Donc, P_n est scindé à racines simples, toutes dans $]0, 1[$.

Il s'agit d'une propriété générale des polynômes orthogonaux. Cela est fait pour les polynômes de Legendre dans l'exercice 1.21 du tome 2 d'analyse. On peut aller plus loin et démontrer qu'entre deux racines de P_{n+1}, on trouve une racine de P_n. Dégageons les étapes de l'argumentation : comme cela a été fait à l'exercice 1.23, on montre que l'on peut écrire $XP_n = a_nP_{n+1} + b_nP_n + c_nP_{n-1}$ avec a_n, b_n, c_n dans $\mathbb{R}$, puis par récurrence que pour tout $(x, y) \in \mathbb{R}^2$,

$$(x-y)\sum_{i=0}^{n} \mathrm{P}_i(x)\mathrm{P}_i(y) = a_n(\mathrm{P}_{n+1}(x)\mathrm{P}_n(y) - \mathrm{P}_{n+1}(y)\mathrm{P}_n(x)).$$

En prenant $y = x+h$ avec h non nul, et en faisant tendre h vers 0 dans la relation

$$\sum_{i=0}^{n} \mathrm{P}_i(x)\mathrm{P}_i(x+h) = a_n \frac{(\mathrm{P}_{n+1}(x+h)\mathrm{P}_n(x) - \mathrm{P}_{n+1}(x)\mathrm{P}_n(x+h))}{h},$$

on obtient

$$\sum_{i=0}^{n} \mathrm{P}_i(x)^2 = a_n\left(\mathrm{P}'_{n+1}(x)\mathrm{P}_n(x) - \mathrm{P}_{n+1}(x)\mathrm{P}'_n(x)\right).$$

Soit $\alpha < \beta$ deux racines consécutives de P_{n+1}. Comme α et β sont simples, les dérivées de P_{n+1} en α et β sont non nulles, nécessairement de signes contraires. En testant en α et β la relation trouvée, comme P_0 est une constante non nulle, il vient

$$0 < \sum_{i=0}^{n} \mathrm{P}_i(\alpha)^2 = a_n\mathrm{P}'_{n+1}(\alpha)\mathrm{P}_n(\alpha) \;\; et \;\; 0 < \sum_{i=0}^{n} \mathrm{P}_i(\beta)^2 = a_n\mathrm{P}'_{n+1}(\beta)\mathrm{P}_n(\beta),$$

si bien que $\mathrm{P}_n(\alpha)\mathrm{P}_n(\beta) < 0$: le théorème des valeurs intermédiaires assure alors que P_n possède une racine dans l'intervalle $]\alpha,\beta[$.

3. Opérons la division euclidienne de $\mathrm{P} \in \mathbb{R}_{2n-1}[\mathrm{X}]$ par P_n. On écrit donc $\mathrm{P} = \mathrm{QP}_n + \mathrm{R}$ avec $\deg \mathrm{Q} < n$ et $\deg \mathrm{R} < n$. Pour tout i, on a $\mathrm{R}(\alpha_i) = \mathrm{P}(\alpha_i)$. Le polynôme P_n est orthogonal à Q et donc $\int_0^1 \mathrm{Q}(t)\mathrm{P}_n(t)\mathrm{d}t = 0$. Il s'ensuit que $\int_0^1 \mathrm{P}(t)\mathrm{d}t = \int_0^1 \mathrm{R}(t)\mathrm{d}t$. Le polynôme R peut s'écrire à l'aide des polynômes interpolateurs de Lagrange associés à $(\alpha_1,\ldots,\alpha_n)$, $\mathrm{L}_i = \prod_{j\neq i}\left(\frac{\mathrm{X}-\alpha_j}{\alpha_i-\alpha_j}\right)$ pour $1 \leqslant i \leqslant n$. Plus précisément, on a $\mathrm{R} = \sum_{i=1}^{n} \mathrm{R}(\alpha_i)\mathrm{L}_i = \sum_{i=1}^{n} \mathrm{P}(\alpha_i)\mathrm{L}_i$. Finalement, on obtient

$$\int_0^1 \mathrm{P}(t)\mathrm{d}t = \int_0^1 \mathrm{R}(t)\mathrm{d}t = \sum_{i=1}^{n} \mathrm{P}(\alpha_i)\int_0^1 \mathrm{L}_i(t)\mathrm{d}t.$$

Il suffit donc de poser

$$\boxed{\lambda_i = \int_0^1 \prod_{j\neq i}\left(\frac{t-\alpha_j}{\alpha_i-\alpha_j}\right)\mathrm{d}t}.$$

En testant en L_i^2 (qui est de degré $2n-2 < 2n-1$), on constate que $\lambda_i = \|\mathrm{L}_i\|^2 > 0$. ◁

Cette dernière question peut être traitée par des arguments de dualité comme on peut le lire dans l'exercice 1.21 du tome 2 d'Analyse. Elle fonde le principe du calcul approché des intégrales par la méthode de Gauss, la quantité $\sum_{i=1}^{n} \lambda_i f(\alpha_i)$ *fournissant une valeur approchée de* $\int_0^1 f(t)dt$ *(qui se trouve être exacte si* $f \in \mathbb{R}_{2n-1}[X]$*).*

Les matrices de Gram constituent la matière du thème suivant. Elles interviennent dans le calcul de la distance d'un vecteur à un sous-espace.

1.26. Matrices de Gram

1. Soit $x_1, \ldots, x_n$ des vecteurs d'un espace préhilbertien réel.

a. Montrer que la matrice $A = (\langle x_i, x_j \rangle)_{1 \leqslant i,j \leqslant n}$ est symétrique positive de même rang que la famille $(x_1, \ldots, x_n)$. *On dit que* A *est la matrice de Gram de la famille* $(x_1, \ldots, x_n)$. *On la note* $G(x_1, \ldots, x_n)$.

On suppose $(x_1, \ldots, x_n)$ libre et on pose $F = \text{Vect}(x_1, \ldots, x_n)$.

b. Soit $x \in E$. Démontrer que d, la distance de x à F, vérifie

$$d^2 = \frac{\det G(x, x_1, x_2, \ldots, x_n)}{\det G(x_1, \ldots, x_n)}.$$

c. On suppose F orienté. Montrer que

$$\det G(x_1, \ldots, x_n) = [x_1, \ldots, x_n]^2.$$

2. Soit $M = (m_{ij}) \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique positive. Montrer qu'il existe une famille $(v_1, \ldots, v_n)$ de n vecteurs de $\mathbb{R}^n$ telle que

$$\forall (i,j) \in [\![1, n]\!]^2, \quad m_{ij} = \langle v_i, v_j \rangle.$$

(École polytechnique)

Solution.

1.a. La matrice A est symétrique réelle et, pour tout $(\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n$, on a

$$\sum_{1 \leqslant i,j \leqslant n} \langle x_i, x_j \rangle \lambda_i \lambda_j = \left\| \sum_{i=1}^{n} \lambda_i x_i \right\|^2 \geqslant 0,$$

donc A est positive.

On note $F = \text{Vect}(x_1, \ldots, x_n)$. Introduisons l'application u qui à tout $x \in F$ associe le n-uplet $(\langle x, x_1 \rangle, \ldots, \langle x, x_n \rangle) \in \mathbb{R}^n$. Cette application est clairement linéaire. Elle est également injective : un vecteur du noyau

est orthogonal à tous les x_i et il est donc dans $\mathrm{F} \cap \mathrm{F}^\perp = \{0\}$. Il est nécessairement nul.

On a donc $\mathrm{rg}(x_1, \ldots x_n) = \mathrm{rg}(u(x_1), \ldots, u(x_n)) = \mathrm{rg}\,\mathrm{A}$.

b. On écrit $x = y + h$ avec $y \in \mathrm{F}$ et $h \in \mathrm{F}^\perp$. On a $d = \|h\|$, $\langle x, x\rangle = d^2 + \langle y, y\rangle$ et $\langle x, x_i\rangle = \langle y, x_i\rangle$ pour tout i. La première colonne de $\mathrm{G}(x, x_1, \ldots x_n)$ s'écrit

$$\begin{pmatrix} \|h\|^2 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \langle y, y\rangle \\ \langle x_1, y\rangle \\ \vdots \\ \langle x_n, y\rangle \end{pmatrix},$$

et la première ligne $(\|h\|^2 + \langle y, y\rangle, \langle y, x_1\rangle, \ldots, \langle y, x_n\rangle)$ si bien que par multilinéarité du déterminant, on obtient

$$\det \mathrm{G}(x. x_1, \ldots, x_n) = \det \begin{pmatrix} \|h\|^2 & \times \cdots \times \\ 0 & \\ \vdots & \boxed{(\langle x_i, x_j\rangle)} \\ 0 & \end{pmatrix} + \det \mathrm{G}(y, x_1, \ldots, x_n).$$

Comme la famille $(y, x_1, \ldots, x_n)$ est liée, d'après la première question, le rang de $\mathrm{G}(y, x_1, \ldots, x_n\rangle$ est n et son déterminant est nul. En développant le premier déterminant selon la première colonne, il vient

$$\det \mathrm{G}(x, x_1, \ldots, x_n) = d^2 \det \mathrm{G}(x_1, \ldots, x_n).$$

Comme $\mathrm{G}(x_1, \ldots, x_n)$ est de rang n, son déterminant est non nul et la relation est démontrée.

Cette formule présente l'intérêt de pouvoir calculer la distance à un sous-espace sans connaître une base orthonormée de ce sous-espace. Elle prouve également par récurrence que $\sqrt{\mathrm{G}(x_1, \ldots, x_n)}$ *correspond au volume dans* F *du parallélotope de côtés* $x_1, \ldots, x_n$ *(il s'agit de* $[0,1]x_1 + \cdots + [0,1]x_n$*) sur lequel on applique le principe que le volume vaut la mesure de la base multipliée par la hauteur.*

c. Pour cette question, prenons $(e_1, \ldots, e_n)$ une base orthonormale directe de F et notons $\mathrm{P} = (\langle e_i, x_j\rangle)_{1 \leqslant i,j \leqslant n}$ la matrice des x_i dans cette base. Comme pour tout $1 \leqslant i, j \leqslant n$,

$$\langle x_i, x_j\rangle = \sum_{k=1}^{n} \langle x_i, e_k\rangle . \langle x_j, e_k\rangle,$$

on en déduit que $\mathrm{G}(x_1, \ldots, x_n) = {}^t\mathrm{P}\mathrm{P}$ et en passant au déterminant

$$\det \mathrm{G}(x_1, \ldots, x_n) = \det {}^t\mathrm{P} \det \mathrm{P} = (\det \mathrm{P})^2 = [x_1, \ldots, x_n]^2.$$

Le produit mixte apparaît alors comme la mesure orientée du volume du parallélotope de côtés $x_1, \dots, x_n$.

2. La matrice symétrique réelle M se diagonalise dans une base orthonormée de $\mathbb{R}^n$. On peut donc trouver P orthogonale et $D = \mathrm{Diag}(d_1, \dots, d_n)$ à coefficients diagonaux positifs, telles que $M = {}^tPDP$. Posons $\Delta = \mathrm{Diag}(\sqrt{d_1}, \dots, \sqrt{d_n})$ et $B = \Delta P$. On a alors $M = {}^tBB$. Si l'on appelle $v_1, \dots, v_n$ les vecteurs colonnes de la matrice B, on obtient exactement $m_{ij} = \langle v_i, v_j \rangle$ pour tout couple (i, j). ◁

Notons qu'il est facile de voir (en rajoutant des 0) que M *est la matrice de Gram d'une famille de vecteurs de $\mathbb{R}^m$ pour tout $m \geqslant n$. En fait, si* M *est symétrique positive de taille n et de rang r, on peut s'intéresser à la dimension minimale p telle que* M *soit la matrice de Gram d'une famille de n vecteurs de $\mathbb{R}^p$. L'exercice a montré que $r \leqslant p \leqslant n$. En fait, on a l'égalité $p = r$. En effet, ayant trouvé la famille $(v_1, \dots, v_n)$, elle engendre un sous-espace* F *de dimension r de $\mathbb{R}^n$, il suffit ensuite de prendre $(e_1, \dots, e_r)$ une base orthonormale de* F *et si l'on considère les colonnes $X_1, \dots, X_n \in \mathbb{R}^r$ contenant les coordonnées des vecteurs v_i dans cette base, il est clair que $(X_i, X_j)_{\mathbb{R}^r} = (v_i, v_j)_{\mathbb{R}^n} = m_{ij}$.*

Cela est à la base de l'exercice suivant dans lequel on recherche des familles de vecteurs unitaires ayant deux à deux toujours le même produit scalaire.

1.27. Familles équiangulaires

Soit $m \geqslant 2$. On considère $\mathbb{R}^n$ muni de son produit scalaire canonique et $\lambda \in \mathbb{R}$. Pour quelles valeurs de n, existe-t-il $(x_1, \dots, x_m)$ dans $\mathbb{R}^n$ tel que les x_i soient unitaires et $\langle x_i, x_j \rangle = \lambda$ pour $i \neq j$?

(**École normale supérieure**)

▷ **Solution.**

Notons A la matrice de $\mathcal{M}_m(\mathbb{R})$ dont les coefficients diagonaux valent 1, les autres valant λ. On cherche en fait à savoir pour quelles valeurs de n, A est la matrice de Gram d'une famille de m vecteurs de $\mathbb{R}^n$. Comme nous venons de le mentionner, il faut et suffit que A soit positive et que $n \geqslant \mathrm{rg}\, A$. Cherchons donc les valeurs propres de A.

Si $\lambda = 0$, $A = I_m$ et la famille $(x_1, \dots, x_m)$ convient si, et seulement si, elle est orthonormée. On peut trouver une famille orthonormée de m vecteurs dans $\mathbb{R}^n$ si, et seulement si, $n \geqslant m$. On suppose dans la suite $\lambda \neq 0$.

Il est clair que $A - (1 - \lambda)I_m$ est de rang 1. Donc $1 - \lambda$ est dans le spectre de A et l'espace propre associé est de dimension $m-1$. Comme la trace de A est égale à m, la valeur propre restante de A est $1+(m-1)\lambda$. Il est alors facile de mener la discussion suivante :

• Si $\lambda > 1$ ou $\lambda < \dfrac{-1}{m-1}$, A n'est pas positive et il n'y a donc aucune solution, quel que soit n.

• Si $\lambda = 1$, A est positive de rang 1. Une famille $(x_1, \dots, x_m)$ existe dans $\mathbb{R}^n$ dès que $n \geqslant 1$: il suffit de prendre tous les x_i égaux à un même vecteur unitaire !

• Si $\dfrac{-1}{m-1} < \lambda < 1$, alors A est définie positive, donc de rang m. Une famille $(x_1, \dots, x_m)$ ayant A pour matrice de Gram existe si, et seulement si, $n \geqslant m$.

• Enfin lorsque $\lambda = \dfrac{-1}{m-1}$, A est positive et de rang $m-1$. Une famille $(x_1, \dots, x_m)$ ayant A pour matrice de Gram existe si, et seulement si, $n \geqslant m-1$. ◁

Dans le dernier cas étudié, les familles solutions sont les m-simplexes : cet exercice prouve donc que cet objet existe à partir de la dimension $m-1$ (voir aussi les exercices et 1.2 et 1.20).

1.28. Familles isométriques

Soit E un espace euclidien de dimension n et $(u_1, \dots, u_p)$, $(v_1, \dots, v_p)$ deux familles de p vecteurs de E.

Montrer que $\mathrm{Gram}(u_1, \dots, u_p) = \mathrm{Gram}(v_1, \dots, v_p)$ si, et seulement si, il existe $f \in \mathrm{O(E)}$ tel que $v_k = f(u_k)$ pour tout $k \in [\![1, p]\!]$.

(**École polytechnique**)

▷ **Solution.**

L'une des implications est triviale : en effet, s'il existe $f \in \mathrm{O(E)}$ tel que $v_k = f(u_k)$ pour tout $k \in [\![1.p]\!]$, on a $\langle u_i, u_j \rangle = \langle f(u_i), f(u_j) \rangle = \langle v_i, v_j \rangle$ pour tout couple (i, j), car f conserve le produit scalaire. Les matrices de Gram des deux familles sont donc les mêmes.

Supposons réciproquement que les deux familles aient la même matrice de Gram $\mathrm{G} \in \mathcal{M}_p(\mathbb{R})$. Nous pouvons démontrer que $\mathrm{rg}(u_1, \dots, u_p) = \mathrm{rg}\,\mathrm{G} = \mathrm{rg}(v_1, \dots, v_p)$ comme nous l'avons fait à l'exercice 1.26. Notons r ce rang. Le cas $r = 0$ étant trivial nous supposerons $r \geqslant 1$.

Supposons que la famille $(u_1, \dots, u_r)$ soit libre, afin d'alléger la rédaction, et notons $\mathrm{F}_u = \mathrm{Vect}(u_1, \dots, u_p) = \mathrm{Vect}(u_1, \dots, u_r)$ le sous-espace de dimension r engendré par les u_i. Comme $\mathrm{Gram}(u_1, \dots, u_r) = \mathrm{Gram}(v_1, \dots, v_r)$, la famille $(v_1, \dots, v_r)$ est aussi de rang r et engendre donc le sous-espace $\mathrm{F}_v = \mathrm{Vect}(v_1, \dots, v_p)$. Considérons l'unique application linéaire f de F_u dans F_v définie par $f(u_k) = v_k$ pour $1 \leqslant k \leqslant r$. Comme f conserve le produit scalaire sur une base de F_u. f est une isométrie de F_u sur F_v, ces sous-espaces étant munis de la structure euclidienne induite par celle de E. On a alors $f(u_k) = v_k$ pour tout

$k > r$: en effet l'hypothèse implique que le vecteur $f(u_k) - v_k \in \mathrm{F}_v$ est orthogonal à $v_1, \ldots, v_r$, puisque si $1 \leqslant i \leqslant r$, on a

$$\langle f(u_k) - v_k, v_i \rangle = \langle f(u_k), f(u_i) \rangle - \langle v_k, v_i \rangle = \langle u_k, u_i \rangle - \langle u_k, u_i \rangle = 0.$$

car f est une isométrie sur F_u. Ainsi, le vecteur $f(u_k) - v_k$ est dans $\mathrm{F}_v \cap \mathrm{F}_v^{\perp}$ et il est donc nul. On a $f(u_k) = v_k$. Il ne reste plus qu'à prouver que f se prolonge en un endomorphisme orthogonal de E. Pour cela il suffit de choisir une base orthonormale $(u'_{r+1}, \ldots, u'_n)$ de $\mathrm{F}_u^{\perp}$ et une base orthonormale $(v'_{r+1}, \ldots, v'_n)$ de $\mathrm{F}_v^{\perp}$ et de poser $f(u'_k) = v'_k$ pour tout k. Il est clair que f conserve le produit scalaire sur la base $(u_1, \ldots, u_r, u'_{r+1}, \ldots, u'_n)$ de E et donc conserve le produit scalaire par bilinéarité. D'où le résultat. ◁

Donnons une vision matricielle de ce résultat : soit A, B *deux matrices de* $\mathcal{M}_{n,p}(\mathbb{R})$. *On a* ${}^t\mathrm{AA} = {}^t\mathrm{BB}$ *si, et seulement si, il existe* $\mathrm{P} \in \mathrm{O}_n(\mathbb{R})$ *telle que* $\mathrm{B} = \mathrm{PA}$.

Le résultat que nous venons de prouver est au cœur de l'exercice suivant.

1.29. Image d'une base orthonormée par un projecteur orthogonal

Soit E un espace euclidien de dimension n et $(u_1, \ldots, u_n)$ une famille de E. On considère la matrice de Gram

$$\mathrm{G} = (\langle u_i, u_j \rangle)_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R}).$$

Montrer l'équivalence des deux conditions suivantes :

(*i*) il existe un projecteur orthogonal p et une base orthonormale $(e_1, \ldots, e_n)$ tels que pour tout i, $p(e_i) = u_i$;

(*ii*) $\mathrm{Sp}\,\mathrm{G} \subset \{0, 1\}$.

(École polytechnique)

Solution.

• Supposons d'abord (*i*). Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base orthonormale et p un projecteur orthogonal tels que $p(e_i) = u_i$ pour tout i. Notons A la matrice de p dans la base $\mathcal{B}$: c'est aussi la matrice de la famille $(u_1, \ldots, u_n)$ dans la base $\mathcal{B}$. Comme cette base est orthonormale, on a $\mathrm{G} = {}^t\mathrm{AA}$. Or, p étant un projecteur orthogonal, il est auto-adjoint et sa matrice A dans la base orthonormale $\mathcal{B}$ est symétrique. Ainsi, $\mathrm{G} = \mathrm{A}^2 = \mathrm{A}$ et le spectre de G est bien inclus dans $\{0, 1\}$.

• Supposons réciproquement que $\mathrm{Sp}\,\mathrm{G} \subset \{0, 1\}$. La matrice G est symétrique réelle donc diagonalisable et l'hypothèse implique que $\mathrm{G}^2 =$

G. Choisissons $\mathcal{B} = (e_1, \ldots, e_n)$ une base orthonormale quelconque de E. Notons $v_1, \ldots, v_n$ les vecteurs de E dont les coordonnées dans $\mathcal{B}$ sont les colonnes successives de G et notons p l'endomorphisme de E ayant G pour matrice dans la base $\mathcal{B}$. Comme $G^2 = G$, p est un projecteur et comme ${}^tG = G$, p est un projecteur orthogonal. De plus, on a $p(e_k) = v_k$ pour tout k. Or, la matrice de Gram de la famille $(v_1, \ldots, v_n)$ est ${}^tGG = G$. D'après l'exercice précédent, il existe donc $f \in O(E)$ tel que $f(v_k) = u_k$ pour tout k. Posons alors $e'_k = f(e_k)$ pour tout k. La famille $\mathcal{B}' = (e'_1, \ldots, e'_n)$ est une base orthonormale de E et on a pour tout k, $u_k = (f \circ p \circ f^{-1})(e'_k)$. Et il est aisé de constater que $p' = f \circ p \circ f^{-1}$ est encore un projecteur orthogonal de E. D'où le résultat. $\lhd$

Les exercices suivants concernent plus spécifiquement les endomorphismes orthogonaux : ce sont les opérateurs linéaires qui conservent le produit scalaire ou, ce qui revient au même, qui conservent la norme euclidienne.

1.30. Une matrice orthogonale

Soit $U = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix} \in \mathcal{M}_{n,1}(\mathbb{R})$ tel que $\sum_{i=1}^{n} u_i^2 = 1$.

Montrer que la matrice $A = I_n - 2U\,{}^tU$ de $\mathcal{M}_n(\mathbb{R})$ est orthogonale. Identifier l'automorphisme de $\mathbb{R}^n$ qu'elle définit.

(École polytechnique)

$\rhd$ **Solution.**

L'hypothèse s'écrit ${}^tUU = 1$. La matrice A est symétrique donc

$$A\,{}^tA = A^2 = I_n + 4U\,{}^tUU\,{}^tU - 4U\,{}^tU = I_n + 4U({}^tUU)\,{}^tU - 4U\,{}^tU = I_n,$$

d'après la remarque initiale. Ainsi, A est orthogonale. Comme $A^2 = I_n$, c'est la matrice d'une symétrie orthogonale. Si X appartient à $\mathcal{M}_{n,1}(\mathbb{R})$ et est orthogonal à U, alors ${}^tUX = 0$ et il vient $AX = X$; d'autre part, on a $AU = U - 2U({}^tUU) = -U$. La matrice A est donc la matrice de la symétrie orthogonale par rapport à l'hyperplan orthogonal à U. $\lhd$

Notons que $-A = 2U\,{}^tU - I_n$ *est la matrice de la symétrie orthogonale par rapport à la droite* $\mathbb{R}U$. *Quant à* $U\,{}^tU$, *c'est la matrice du projecteur orthogonal sur la droite* $\mathbb{R}U$.

L'exercice suivant est un cas particulier de l'exercice 1.15.

1.31. Convergence en moyenne des puissances d'une matrice orthogonale

Soit $M \in O_p(\mathbb{R})$ une matrice orthogonale réelle. Déterminer

$$\lim_{n\to+\infty} \frac{1}{n} \sum_{k=0}^{n-1} M^k.$$

(École polytechnique)

Solution.

Posons $A_n = \frac{1}{n} \sum_{k=0}^{n-1} M^k$, pour tout entier naturel non nul n. Remarquons qu'un point fixe de M est encore fixe pour A_n. Ainsi, si $X \in \mathrm{Ker}(M - I_p)$, alors $MX = X$ et, pour tout $n \in \mathbb{N}$, $A_nX = X$. La suite (A_nX) converge vers X.

Montrons que $\mathbb{R}^p = \mathrm{Ker}(M - I_p) \overset{\perp}{\oplus} \mathrm{Im}(M - I_p)$. Il suffit de démontrer que les deux sous-espaces sont orthogonaux. Pour $X \in \mathrm{Ker}(M - I_p)$ et $Y = MZ - Z$, on a

$${}^tXY = {}^tXMZ - {}^tXZ = {}^t(MX)MZ - {}^tXZ = {}^tX\,{}^tMMZ - {}^tXZ = 0.$$

Regardons ce qui se passe sur $\mathrm{Im}(M - I_p)$. Si $X \in \mathrm{Im}(M - I_p)$, il existe $Y \in \mathbb{R}^p$ tel que $X = MY - Y$. Alors, pour tout entier $n \geqslant 1$,

$$A_nX = \frac{1}{n} \sum_{k=0}^{n-1} M^kX = \frac{1}{n} \sum_{k=0}^{n-1} (M^{k+1}Y - M^kY) = \frac{1}{n}(M^nY - Y).$$

On en déduit que $\|A_nX\| \leqslant \frac{1}{n}(\|M^nY\| + \|Y\|) \leqslant \frac{2}{n}\|Y\|$, car M est orthogonale. La suite (A_nX) converge vers 0.

Pour tout $X \in \mathbb{R}^p$ qui s'écrit $Y + Z$, avec $Y \in \mathrm{Ker}(M - I_p)$ et $Z \in \mathrm{Im}(M - I_p)$, la suite (A_nZ) converge vers Y. Ainsi la suite (A_n) converge vers la matrice de la projection orthogonale sur $\mathrm{Ker}(M - I_p)$. ◁

Le fait d'avoir comme hypothèse M *orthogonale permet de démontrer simplement que* $\mathbb{R}^p = \mathrm{Ker}(M - I_p) \overset{\perp}{\oplus} \mathrm{Im}(M - I_p)$. *C'est plus délicat avec seulement* $\|M\| \leqslant 1$ *comme hypothèse (voir exercice 1.15).*

1.32. Orbites sous l'action d'un endomorphisme orthogonal

Soit E un espace euclidien et $(x_n)_{n\in\mathbb{Z}}$ une famille génératrice de E.

Montrer l'équivalence des conditions suivantes :

(i) il existe $u \in \mathrm{O(E)}$ tel que pour tout $n \in \mathbb{Z}$, $x_{n+1} = u(x_n)$;

(ii) il existe une suite $(a_k)_{k\in\mathbb{Z}}$ de $\mathbb{R}$ telle que pour tout $n \in \mathbb{Z}$ et tout $k \in \mathbb{Z}$, $(x_n, x_{n+k}) = a_k$.

(École polytechnique)

▷ **Solution.**

• Montrons que $(i) \Longrightarrow (ii)$. On a pour tout $n \in \mathbb{Z}$, $x_n = u^n(x_0)$. Ainsi, si n et k sont des entiers,

$$\langle x_n, x_{n+k}\rangle = \langle u^n(x_0), u^{n+k}(x_0)\rangle = \langle x_0, u^k(x_0)\rangle,$$

car $u^n \in \mathrm{O(E)}$ conserve le produit scalaire. Il suffit donc de poser $a_k = \langle x_0, u^k(x_0)\rangle$ pour tout k.

Notons que $a_{-k} = a_k$ pour tout $k \in \mathbb{Z}$: le produit scalaire entre deux vecteurs de l'orbite ne dépend que de l'écart en valeur absolue entre les indices de ces vecteurs.

• Montrons que $(ii) \Longrightarrow (i)$. La suite $(x_n)_{n\in\mathbb{Z}}$ étant génératrice de E, on peut en extraire une base $(x_{k_1}, \ldots, x_{k_p})$ où $p = \dim \mathrm{E}$ et $k_1 < k_2 < \cdots < k_p$. Considérons l'unique endomorphisme u de E défini par $u(x_{k_i}) = x_{k_i+1}$ pour tout $i \in [\![1, p]\!]$. On va montrer que u est dans O(E) puis que $u(x_n) = x_{n+1}$ pour tout entier $n \in \mathbb{Z}$.

Le premier point provient de ce que u conserve le produit scalaire pour deux vecteurs quelconques de la base choisie. En effet, pour i, j dans $[\![1, p]\!]$,

$$\langle u(x_{k_i}), u(x_{k_j})\rangle = \langle x_{k_i+1}, x_{k_j+1}\rangle = a_{k_j-k_i} = \langle x_{k_i}, x_{k_j}\rangle.$$

Il en découle par bilinéarité que u conserve le produit scalaire de E. C'est donc bien un élément de O(E).

Soit n un entier quelconque. On décompose x_n dans la base choisie : $x_n = \sum_{i=1}^{p} \lambda_i x_{k_i}$. On a alors $u(x_n) = \sum_{i=1}^{p} \lambda_i x_{k_i+1}$. Pour tout entier m, on a donc

$$\begin{aligned}\langle u(x_n), x_m\rangle = \sum_{i=1}^{p} \lambda_i \langle x_{k_i+1}, x_m\rangle &= \sum_{i=1}^{p} \lambda_i \langle x_{k_i}, x_{m-1}\rangle \\ &= \langle x_n, x_{m-1}\rangle = \langle x_{n+1}, x_m\rangle.\end{aligned}$$

Comme cela vaut en particulier en prenant pour m les entiers k_i, on a forcément $u(x_n) = x_{n+1}$. ◁

On peut enlever l'hypothèse $(x_n)_{n\in\mathbb{Z}}$ génératrice : dans la réciproque, il suffit de se placer sur le sous-espace F *engendré par la suite puis de prolonger ensuite l'endomorphisme orthogonal de* F *construit en un endomorphisme orthogonal de* E*, ce qui est facile.*

1.33. Générateurs de O(E)

Soit E un espace euclidien.

1. Montrer que tout $u \in \mathrm{O(E)}$ peut s'écrire comme produit d'au plus r réflexions où $r = \mathrm{rg}(u - \mathrm{Id_E})$. En déduire que les réflexions engendrent O(E).

2. Montrer que les retournements engendrent SO(E) lorsque $\dim \mathrm{E} = 3$.

(École polytechnique)

Solution.

1. Procédons par récurrence sur le rang r de $u - \mathrm{Id_E}$.

Si $r = 0$, $u = \mathrm{Id_E}$: il n'y a rien à vérifier.

Supposons $r \geqslant 1$ et le résultat vrai pour $\mathrm{rg}(u - \mathrm{Id_E}) < r$. Alors u n'est pas l'identité et il existe e de norme 1 tel que $u(e) \neq e$. De $\|u(e)\| = \|e\|$, on déduit $(u(e) - e) \perp (u(e) + e)$. Ainsi la réflexion s par rapport à l'orthogonal de $u(e) - e$ vérifie $s(u(e) + e) = u(e) + e$ et $s(u(e) - e) = -(u(e) - e)$, donc échange e et $u(e)$.

De plus, $\mathrm{Ker}(u - \mathrm{Id_E})$ est contenu dans $\mathrm{Ker}(s \circ u - \mathrm{Id_E})$. En effet, prenons $x \in \mathrm{Ker}(u - \mathrm{Id_E})$. Montrons que $(s \circ u)(x) = x$ *i.e.* $s(x) = x$. Pour cela, il suffit de vérifier que x est orthogonal à $u(e) - e$. Comme $u(x) = x$ et comme u est orthogonal, on a

$$(x, u(e) - e) = (u(x), u(e)) - \langle x, e\rangle = \langle x, e\rangle - \langle x, e\rangle = 0.$$

Ainsi, $\mathrm{Ker}(u - \mathrm{Id_E}) \subset \mathrm{Ker}(s \circ u - \mathrm{Id_E})$ et cette inclusion est stricte, puisque $e \notin \mathrm{Ker}(u - \mathrm{Id_E})$ et $s(u(e)) = e$. On peut donc appliquer l'hypothèse de récurrence à $s \circ u$, car $\mathrm{rg}(s \circ u - \mathrm{Id_E}) < \mathrm{rg}(u - \mathrm{Id_E})$. Il existe donc des réflexions $s_1, \ldots, s_p$ telles que $s \circ u = s_1 \circ \cdots \circ s_p$ avec $p \leqslant \mathrm{rg}(s \circ u - \mathrm{Id_E})$. Finalement, on a l'écriture $u = s \circ s_1 \circ \cdots \circ s_p$ où figurent $p + 1$ réflexions avec $p + 1 \leqslant r$.

En fait, $r = \mathrm{rg}(u - \mathrm{I})$ est le nombre minimal de réflexions s_i que l'on doit avoir dans une décomposition $u = s_1 \circ \cdots \circ s_p$. En effet, si u s'écrit $u = s_1 \circ \ldots \circ s_p$, où s_i est la réflexion par rapport à l'hyperplan H_i, il est clair que

$$\mathrm{H}_1 \cap \mathrm{H}_2 \cap \cdots \cap \mathrm{H}_p \subset \mathrm{Ker}(u - \mathrm{Id_E})$$

Comme la dimension de $\mathrm{H}_1 \cap \mathrm{H}_2 \cap \cdots \cap \mathrm{H}_p$ est supérieure à $\dim \mathrm{E} - p$, le rang de $u - \mathrm{Id_E}$ est inférieur ou égal à p.

2. Supposons $\dim \mathrm{E} = 3$. Soit u une rotation de E que l'on suppose distincte de l'identité. D'après le cours on a $\mathrm{rg}(u - \mathrm{Id}_\mathrm{E}) = 2$ puisque $\mathrm{Ker}(u - \mathrm{Id}_\mathrm{E})$ est une droite (l'axe de u). D'après la question précédente u peut s'écrire comme un produit de deux réflexions : $u = s_1 \circ s_2$. Or, si H est un plan de E et s la réflexion par rapport à H, alors $-s$ est le retournement par rapport à la droite $\mathrm{D} = \mathrm{H}^\perp$. La relation $u = (-s_1) \circ (-s_2)$ montre que u est la composée de deux retournements. On a prouvé que les retournements engendrent $\mathrm{SO(E)}$. $\triangleleft$

Le résultat de cette seconde question reste vrai pour toute dimension $n \geqslant 3$.

Le théorème de Maschke, élément fondamental de la théorie des représentations des groupes finis a déjà été rencontré dans le tome 1 d'algèbre (exercice 22 du chapitre 6). La preuve différente qui inspire l'énoncé suivant est valable lorsque le corps de base est $\mathbb{R}$ *(ou* $\mathbb{C}$ *avec un produit hermitien) et consiste à réaliser le groupe fini étudié comme un sous-groupe du groupe orthogonal relatif à un produit scalaire* ad hoc.

1.34. Théorème de Maschke (1898)

Soit E un espace euclidien et G un sous-groupe fini de $\mathrm{GL(E)}$.

1. Soit $f \in \mathrm{O(E)}$. Montrer que tout sous-espace stable par f admet un supplémentaire stable par f.

2. Montrer que $(x, y) \longmapsto (x|y) = \sum_{g \in \mathrm{G}} \langle g(x), g(y) \rangle$ définit un produit scalaire sur E.

3. Soit F un sous-espace stable par tous les éléments de G. Démontrer que F possède un supplémentaire stable par tous les éléments de G.

(**École polytechnique**)

$\triangleright$ **Solution.**

1. Soit F un sous-espace stable par f. Montrons que $\mathrm{F}^\perp$ est stable par f. Soit $x \in \mathrm{F}^\perp$. Montrons que pour tout $y \in \mathrm{F}$, $\langle f(x), y \rangle = 0$. Soit $y \in \mathrm{F}$. Comme f est bijective, $f(\mathrm{F}) = \mathrm{F}$. Il existe $t \in \mathrm{F}$ avec $y = f(t)$ et

$$\langle f(x), y \rangle = \langle f(x), f(t) \rangle = (x, t\rangle = 0.$$

Ainsi $f(x)$ appartient à $\mathrm{F}^\perp$.

2. Notons déjà que $(x, y) \longmapsto (x|y)$ est bien défini et symétrique. Comme les éléments g de G sont linéaires et $\langle\ .\ \rangle$ est bilinéaire, on vérifie sans difficulté que cette application est linéaire en x. Soit $x \in \mathrm{E}$ non nul. On a $(x|x) = \sum_{g \in \mathrm{G}} \|g(x)\|^2 > 0$, car chaque $g(x)$ est non nul, g

étant un isomorphisme. On conclut que l'on a bien affaire à un nouveau produit scalaire sur E.

3. L'intérêt de ce nouveau produit scalaire est qu'il fait des éléments de G des isomorphismes orthogonaux. En effet, soit $f \in \mathrm{G}$ et $x \in \mathrm{E}$. Comme G est un groupe, quand g décrit G, $g \circ f$ décrit G également, si bien que

$$(f(x)|f(x)) = \sum_{g\in\mathrm{G}} \langle g(f(x)), g(f(x))\rangle = \sum_{g\in\mathrm{G}} \langle g\circ f(x), g\circ f(x)\rangle$$
$$= \sum_{h\in\mathrm{G}} \langle h(x), h(x)\rangle = (x|x).$$

Si $f \in \mathrm{G}$, comme F est stable par f, l'orthogonal de F pour $(\ |\ )$ est aussi stable par f. ◁

La première question de l'exercice précédent traduit la semi-simplicité des endomorphismes orthogonaux. Les endomorphismes diagonalisables à polynôme caractéristique scindé vérifient également cette propriété (voir dans le tome 2 d'algèbre, l'exercice 37 du second chapitre). À partir de la semi-simplicité, l'énoncé suivant élucide complètement la structure des éléments du groupe orthogonal en toute dimension. Matriciellement, cela revient à donner un représentant simple de chaque classe de conjugaison du groupe $\mathrm{O}_n(\mathbb{R})$.

1.35. Réduction des matrices orthogonales

Soit E un espace vectoriel euclidien de dimension $n \geqslant 1$ et f un élément de $\mathrm{O(E)}$.

1. Montrer que tout sous-espace vectoriel stable par f admet un supplémentaire stable par f.

2. Que dire des valeurs propres réelles de f ? Montrer que si f n'a pas de valeur propre réelle, alors il existe un plan stable par f dans lequel f induit une rotation.

3. En déduire qu'il existe une base orthonormée dans laquelle la matrice de f est diagonale par blocs avec pour blocs I_p, $-\mathrm{I}_q$, $\mathrm{M}(\theta_1), \ldots, \mathrm{M}(\theta_r)$ où p, q, r sont des entiers naturels, où $\mathrm{M}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$ pour tout réel θ et où les θ_i sont des réels non nuls modulo π. À quelle condition f est-elle une rotation ?

4. En déduire que $\mathrm{SO}_n(\mathbb{R})$ est connexe par arcs.

(École polytechnique)

▷ **Solution.**

1. Cela a été fait dans la première question de l'exercice précédent.

2. Comme f conserve la norme, les seules valeurs propres réelles possibles sont 1 et -1. Notons par ailleurs, ce qui nous servira plus loin, que les espaces propres associés à ces deux valeurs propres sont orthogonaux : si $u(x) = x$ et $u(y) = -y$, on peut écrire

$$\langle x, y\rangle = \langle u(x), u(y)\rangle = \langle x, -y\rangle = -\langle x, y\rangle,$$

ce qui donne $\langle x, y\rangle = 0$.

Supposons maintenant que f n'ait pas de valeur propre réelle. Voici deux façons, toutes les deux intéressantes, pour obtenir un plan P stable par f.

Première solution. L'endomorphisme $v = u + u^*$ est symétrique donc diagonalisable. Soit $\lambda \in \mathbb{R}$ une valeur propre de v et x un vecteur propre associé. Comme $u(x)$ n'est pas colinéaire à x (car u n'a pas de valeur propre réelle), $\mathrm{P} = \mathrm{Vect}(x, u(x))$ est un plan vectoriel. Montrons qu'il est stable par u. Pour cela, il suffit de prouver que $u^2(x) \in \mathrm{P}$. Or on a $u(x) + u^{-1}(x) = u(x) + u^*(x) = \lambda x$ donc $u^2(x) = \lambda u(x) - x \in \mathrm{P}$. D'où le résultat.

Seconde solution. Soit z une racine complexe de χ_f. Alors $\overline{z}$ est aussi racine de χ_f et le polynôme à coefficients réels $\mathrm{Q} = (\mathrm{X}-z)(\mathrm{X}-\overline{z})$ divise le polynôme minimal μ_f de f. Si $\mathrm{Q}(f)$ était inversible, le quotient $\frac{\mu_f}{\mathrm{Q}}$ serait un polynôme annulateur de f, ce qui est impossible par définition de μ_f. Il existe donc $x \in \operatorname{Ker} \mathrm{Q}(f)$, non nul. La famille $(x, f(x))$ est forcément libre et en posant $\mathrm{Q} = \mathrm{X}^2 + a\mathrm{X} + b$, on obtient $f^2(x) = -af(x) - bx$, ce qui montre que le plan $\mathrm{P} = \mathrm{Vect}(x, f(x))$ est stable par f.

Remarquons que cette seconde solution n'utilise pas le fait que f est un endomorphisme orthogonal.

Pour conclure notre question, on note que la restriction de f au plan stable P est un endomorphisme orthogonal du plan sans valeur propre réelle, c'est-à-dire une rotation plane.

3. On raisonne par récurrence sur n. Le cas $n = 1$ est trivial, le cas $n = 2$ est étudié dans le cours. Supposons $n \geqslant 3$ et le résultat prouvé pour les dimensions strictement inférieures à n. On distingue deux cas :

• Si 1 ou -1 est valeur propre de f, on pose $\mathrm{F} = \operatorname{Ker}(f - \mathrm{Id}) \overset{\perp}{\oplus} \operatorname{Ker}(f + \mathrm{Id})$ (l'un des deux espaces pouvant être nul). On choisit une base orthonormale B_1 de F obtenue en juxtaposant des bases orthonormales de chacun des deux espaces propres (qui rappelons-le, sont orthogonaux). Si $\mathrm{F} = \mathrm{E}$ on a terminé : f est diagonalisable et dans la base B_1 sa matrice est diagonale. Sinon, on note g la restriction de f à $\mathrm{F}^\perp$. On applique l'hypothèse de récurrence à g qui est dans $\mathrm{O}(\mathrm{F}^\perp)$. Il existe une base orthonormale B_2 de $\mathrm{F}^\perp$ dans laquelle la matrice de g a la forme

indiquée. En fait, comme g n'a aucune valeur propre réelle, il n'y a que des blocs (2, 2) contenant des matrices $M(\theta)$. Dans la base orthonormale $B = (B_1, B_2)$, la matrice de f a la forme souhaitée.

• Si f n'a aucune valeur propre réelle, on a vu en 2 qu'il existe un plan P stable par f dans lequel f induit une rotation. La matrice de la restriction de f à P est une matrice $M(\theta_1)$ avec θ_1 non nul modulo π. Pour conclure, on applique comme avant l'hypothèse de récurrence à la restriction g de f à $P^\perp$.

En prenant le déterminant de la matrice obtenue, il est clair que f est une rotation si, et seulement si, la dimension du sous-espace propre relatif à la valeur propre -1 est paire.

4. Matriciellement, on vient de prouver que pour toute matrice A de $SO_n(\mathbb{R})$, il existe $P \in O_n(\mathbb{R})$ telle que

$$P^{-1}AP = \mathrm{Diag}(I_p, M(\theta_1), \ldots, M(\theta_r))$$

où les θ_i sont non nuls modulo 2π (on a regroupé les -1 éventuels deux par deux pour former des matrices $M(\pi)$). Si l'on pose pour $t \in [0, 1]$

$$A_t = P\,\mathrm{Diag}(I_p, M(t\theta_1), \ldots, M(t\theta_r))P^{-1},$$

alors $t \longmapsto A_t$ est un chemin continu joignant $A_0 = I_n$ à $A_1 = A$. Cela prouve que $SO_n(\mathbb{R})$ est connexe par arcs. ◁

Le complémentaire de $SO_n(\mathbb{R})$ dans $O_n(\mathbb{R})$ est également connexe par arcs car, étant donné une matrice de réflexion S de $\mathbb{R}^n$, c'est l'image de $SO_n(\mathbb{R})$ par l'application continue $M \longmapsto SM$. Ainsi, $O_n(\mathbb{R})$ possède deux composantes connexes par arcs (cette notion est développée au début de l'exercice 1.37) qui sont $SO_n(\mathbb{R})$ et son complémentaire.

Donnons maintenant une application de ce théorème de réduction des matrices orthogonales.

1.36. Exponentielle de matrices antisymétriques réelles

On note $\mathcal{A}_n(\mathbb{R})$ l'ensemble des matrices réelles antisymétriques de taille n. Montrer que $\exp : \mathcal{A}_n(\mathbb{R}) \longrightarrow SO_n(\mathbb{R})$ est surjective.

(École normale supérieure)

Solution.

• Si $A \in \mathcal{A}_n(\mathbb{R})$, on a

$${}^t\exp(A) = \exp({}^tA) = \exp(-A) = (\exp(A))^{-1}$$

et $\exp(A) \in O_n(\mathbb{R})$. Calculons le déterminant :

$$\det(\exp(A)) = \exp(\mathrm{Tr}(A)) = \exp(0) = 1$$

et on conclut $\exp(A) \in \mathrm{SO}_n(\mathbb{R})$. Ainsi exp est une application de $\mathcal{A}_n(\mathbb{R})$ dans $\mathrm{SO}_n(\mathbb{R})$.

• Montrons que cette application est surjective. Soit U une matrice orthogonale. Alors il existe une matrice orthogonale P telle que $U = PMP^{-1}$, où M est de la forme

$$M = \begin{pmatrix} 1 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 1 & & & & 0 & & \\ & & & -1 & & & & & \\ & & & & \ddots & & & & \\ & & & & & -1 & & & \\ & & 0 & & & & R_{\theta_1} & & \\ & & & & & & & \ddots & \\ & & & & & & & & R_{\theta_r} \end{pmatrix},$$

où $R_{\theta_j} = \begin{pmatrix} \cos\theta_j & -\sin\theta_j \\ \sin\theta_j & \cos\theta_j \end{pmatrix}$. Si $U \in \mathrm{SO}_n(\mathbb{R})$ les -1 sont en nombre pair, on peut les regrouper par deux. Puisque $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = R_\pi$, on peut écrire

$$M = \begin{pmatrix} 1 & & & & & \\ & \ddots & & & & \\ & & 1 & & 0 & \\ & & & R_{\theta_1} & & \\ & & 0 & & \ddots & \\ & & & & & R_{\theta_s} \end{pmatrix}.$$

On détermine d'abord $B \in \mathcal{A}_n(\mathbb{R})$ telle que $M = \exp B$. On pose $J = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. On obtient $J^2 = -I_2$, d'où l'on déduit que, pour tout entier n, $J^{2n} = (-1)^n I_2$ et $J^{2n+1} = (-1)^n J$, puis que

$$\exp(\theta J) = \sum_{n=0}^{+\infty} \frac{(-1)^n \theta^{2n}}{(2n)!} I_2 + \sum_{n=0}^{+\infty} \frac{(-1)^n \theta^{2n+1}}{(2n+1)!} J = \cos\theta I_2 + \sin\theta J = R_\theta.$$

Considérons la matrice $B = \begin{pmatrix} 0 & & & & \\ & \ddots & & & \\ & & 0 & & \\ & & & \theta_1 J & \\ & & & & \ddots \\ & & & & & \theta_s J \end{pmatrix}$.

Du calcul précédent, on déduit que $\exp B = M$, puis que

$$\exp(PBP^{-1}) = P\exp(B)P^{-1} = PMP^{-1} = U.$$

Par construction, B est antisymétrique. On pose $A = PBP^{-1}$. On a alors ${}^tA = {}^tP^{-1}\,{}^tB\,{}^tP = -PBP^{-1} = -A$, car P est orthogonale. Ainsi $A \in \mathcal{A}_n(\mathbb{R})$ et vérifie $\exp A = U$. ◁

Voici maintenant la preuve d'un résultat algébrique abstrait sur le groupe des rotations de $\mathbb{R}^3$ qui s'inscrit dans la classification des groupes classiques : le groupe SO_3 est simple i.e. *il ne possède que deux sous-groupes distingués, à savoir lui-même et le sous-groupe réduit à l'identité. La nature géométrique des éléments y joue un rôle central.*

1.37. Simplicité de SO_3

Soit G un sous-groupe de SO_3, le groupe des rotations de l'espace euclidien canonique $\mathbb{R}^3$, et G_0 la composante connexe par arcs de Id dans G.

1. Montrer que G_0 est un sous-groupe de G.

2. Montrer que si G est distingué dans SO_3 (ce qui signifie que pour tout $h \in SO_3$ et tout $g \in G$, on a $hgh^{-1} \in G$), il en est de même de G_0.

3. On suppose ici que G est connexe par arcs, distingué et non réduit à $\{Id\}$. Montrer que G contient une rotation d'angle π. En déduire que $G = SO_3$.

4. On suppose que G est distingué. Montrer que $G = \{Id\}$ ou $G = SO_3$.

(École normale supérieure)

Solution.

Commençons par quelques précisions sur les notions abordées dans cet exercice. On notera sous forme multiplicative la composée de deux éléments de G. Le groupe SO_3 est une partie de l'espace vectoriel $\mathcal{L}(\mathbb{R}^3)$ muni de sa topologie d'espace normé. On rappelle qu'un chemin de G est

une application $\gamma : [0,1] \longrightarrow G$, continue ; $\gamma(0)$ est l'origine du chemin et $\gamma(1)$ son extrémité.

On considère sur G la relation $\mathcal{R}$ définie par : $g\mathcal{R}h$ s'il existe un chemin de G d'origine g et d'extrémité h. Cette relation est une relation d'équivalence. En effet, si $g \in G$, $g\mathcal{R}g$, comme on le voit en considérant $\gamma : t \longmapsto g$; si $g\mathcal{R}h$ et si γ est un chemin de G d'origine g et d'extrémité h, l'application $\gamma' : t \longmapsto \gamma(1-t)$ est un chemin d'origine h et d'extrémité g ; enfin si $g\mathcal{R}h$ et $h\mathcal{R}k$ et si γ (resp. γ') est un chemin de G d'origine g et d'extrémité h (resp. d'origine h et d'extrémité k) l'application $\gamma'' : [0,1] \longrightarrow G$ définie par $\gamma''(t) = \gamma(2t)$ si $t \in \left[0, \frac{1}{2}\right]$ et $\gamma''(t) = \gamma'(2t-1)$ si $t \in \left[\frac{1}{2}, 1\right]$ est un chemin d'origine g et d'extrémité k. Les classes d'équivalence pour cette relation sont les composantes connexes par arcs de G.

1. Par définition, G_0 contient Id. Soit g et h deux éléments de G_0, γ et γ' des chemins de G_0 reliant Id à g et h respectivement. Considérons l'application $\gamma'' : t \longmapsto \gamma(t)(\gamma'(t))^{-1}$. Pour tout $t \in [0,1]$, $\gamma(t)$ et $\gamma'(t)$ appartiennent à G donc $\gamma(t)(\gamma'(t))^{-1}$ appartient à G, car G est un sous-groupe de SO_3. Enfin l'application $g \longmapsto g^{-1}$ est continue sur SO_3, car si l'on identifie un élément de SO_3 à sa matrice dans la base canonique, les coefficients de g^{-1} dépendent polynomialement des coefficients de g. Enfin $\gamma''(0) = \mathrm{Id} \cdot \mathrm{Id} = \mathrm{Id}$ et $\gamma''(1) = gh^{-1}$. Ainsi, γ'' est un chemin de Id à gh^{-1} et gh^{-1} appartient à G_0. On a montré que G_0 est un sous-groupe de G.

2. Soit $g \in G_0$, γ un chemin de G de Id à g et $h \in SO_3$. Considérons l'application $\gamma' : t \in [0,1] \longmapsto h\gamma(t)h^{-1} \in SO_3$. Pour tout $t \in [0,1]$, $\gamma(t) \in G$ et G étant distingué, $h\gamma(t)h^{-1} \in G$. L'application γ est continue de même que la multiplication à droite ou à gauche par un élément de SO_3, donc γ' est continue. Enfin, on a $\gamma'(0) = h\,\mathrm{Id}\,h^{-1} = \mathrm{Id}$ et $\gamma'(1) = hgh^{-1}$. L'application γ' est un chemin de Id à hgh^{-1} et hgh^{-1} appartient à G_0 : G_0 est distingué dans SO_3.

3. Si θ est l'angle d'une rotation g de $\mathbb{R}^3$ (notons que θ est défini au signe près, puisque si l'on change l'orientation de l'axe de la rotation, l'angle est changé en son opposé), nous savons qu'il existe une base orthonormale dans laquelle sa matrice est

$$\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

si bien que $\mathrm{Tr}\, g = 2\cos\theta + 1$ et donc l'application $g \in SO_3 \longmapsto \cos\theta = \frac{\mathrm{Tr}\, g - 1}{2}$ est une application continue. Il suffit de montrer que cette application prend la valeur -1 pour avoir une rotation $g \in G$ d'angle π.

Mais, grâce à l'argument de connexité, il va s'avérer plus facile de trouver un élément de G d'angle θ avec $\cos\theta = 0$, autrement dit, nous allons prouver que G contient une rotation r d'angle $\pm\frac{\pi}{2}$. Alors r^2 sera un élément de G d'angle π. Montrons pour commencer que G possède un élément s tel que $\cos\theta \leqslant 0$. On appliquera ensuite le théorème des valeurs intermédiaires. Par hypothèse, G possède un élément g différent de Id. Quitte à considérer g^{-1}, on peut supposer qu'une mesure θ de son angle appartient à $]0,\pi]$. Si $\cos\theta \leqslant 0$, on prend $s = g$, sinon on pose $N = E\left(\frac{\pi}{2\theta}\right)$. On a alors $\theta \in \left]0, \frac{\pi}{2}\right[$ et $N\theta \leqslant \frac{\pi}{2} < (N+1)\theta < 2\frac{\pi}{2}$; g^{N+1} est une rotation d'angle $(N+1)\theta \in \left[\frac{\pi}{2}, \pi\right[$; on pose $s = g^{N+1}$.

Puisque G est connexe par arcs, il existe un chemin γ de Id à s. L'application $\varphi : t \in [0,1] \longmapsto \frac{1}{2}(\mathrm{Tr}(\gamma(t)) - 1)$ est continue, puisque Tr et γ le sont. Puisque $\varphi(0) = \cos 0 = 1$ et $\varphi(1) = \frac{1}{2}(\mathrm{Tr}\, s - 1) \leqslant 0$, il existe d'après le théorème des valeurs intermédiaires, $t_0 \in [0,1]$ tel que $\varphi(t_0) = 0$. La rotation $r = \gamma(t_0)$ a un angle de $\pm\frac{\pi}{2}$ et $R = r^2$ un angle de π ; R est un retournement.

Pour tout $g \in SO_3$, gRg^{-1} appartient à G et $\mathrm{Tr}(gRg^{-1}) = \mathrm{Tr}(R)$ donc gRg^{-1} est aussi d'angle π. Si le vecteur u appartient à l'axe Δ de R, on a alors $(gRg^{-1})(g(u)) = g(u)$, ce qui montre que gRg^{-1} est le retournement d'axe $g(\Delta)$. Or, étant donné une droite D de $\mathbb{R}^3$, il est toujours possible de trouver une rotation g de $\mathbb{R}^3$ telle que $D = g(\Delta)$, en prenant un axe orthogonal à D et Δ et un angle *ad hoc*[3]. Ainsi, G contient tous les retournements.

Or, toute rotation s'écrit comme produit de deux retournements (voir l'exercice 1.33 pour la preuve de ce résultat du cours). On conclut que G contient toutes les rotations *i.e.* $G = SO_3$.

4. On suppose simplement que G est distingué. Considérons G_0 la composante connexe par arcs de Id. D'après la question 2, G_0 est un sous-groupe distingué et, par définition, connexe par arcs de SO_3. D'après la question 3, si G_0 est différent de $\{\mathrm{Id}\}$, $G_0 = SO_3$, auquel cas $G = SO_3$.

Supposons $G_0 = \{\mathrm{Id}\}$ et montrons que $G = \{\mathrm{Id}\}$. Remarquons que dans ces conditions, toutes les composantes connexes par arcs de G sont des singletons. En effet, si g' est dans la composante de g, relié par le chemin γ, alors $t \longmapsto g^{-1}\gamma(t)$ est un chemin de G reliant Id à $g^{-1}g'$. Ainsi, $g^{-1}g' \in G_0 = \{\mathrm{Id}\}$ et il en résulte $g' = g$.

Raisonnons par l'absurde et supposons que G contienne un élément g différent de Id. Soit h une rotation quelconque, différente de Id, θ une mesure de l'angle de h. Pour tout $t \in [0,1]$, on note h_t la rotation de même axe et d'angle $t\theta$. L'application $t \longmapsto h_t$ est continue, car,

[3] On dit que SO_3 agit transitivement sur les droites de $\mathbb{R}^3$.

matriciellement, elle se traduit dans une certaine base orthonormale par $t \longmapsto \begin{pmatrix} \cos(t\theta) & -\sin(t\theta) & 0 \\ \sin(t\theta) & \cos(t\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix}$. L'application $t \in [0,1] \longmapsto h_t g h_t^{-1}$ est un chemin de G, car G est distingué, d'origine g et d'extrémité hgh^{-1}. Ainsi hgh^{-1} appartient à la composante connexe par arcs de g.

Cette dernière étant réduite à un singleton, on obtient $hgh^{-1} = g$. Or, si g est une rotation d'axe Δ, hgh^{-1} est une rotation d'axe $h(\Delta)$, donc $h(\Delta) = \Delta$. C'est impossible : une droite ne peut pas être invariante par toutes les rotations de l'espace. On conclut que $G = \{Id\}$.

Conclusion. Les seuls sous-groupes distingués de SO_3 sont SO_3 et $\{Id\}$. $\lhd$

Un polynôme réel de quatre variables peut être vu comme une application définie pour un couple de vecteurs de $\mathbb{R}^2$. Il s'agit ici de prouver que de tels polynômes invariants lorsqu'on fait agir une rotation quelconque sur un couple de vecteurs (x,y) s'écrivent comme fonction polynomiale des normes et du produit scalaire de x et y.

1.38. Polynômes de quatre variables invariants sous l'action de $O_2(\mathbb{R})$

Soit $G = O_2(\mathbb{R})$ le groupe des matrices orthogonales de $\mathcal{M}_2(\mathbb{R})$. Soit $F \in \mathbb{R}[x_1, x_2, y_1, y_2]$. Si $x = (x_1, x_2)$ et $y = (y_1, y_2)$, on note $F(x,y)$ au lieu de $F(x_1, x_2, y_1, y_2)$. On fait agir G sur $\mathbb{R}[x_1, x_2, y_1, y_2]$ par $gF(x,y) = F(gx, gy)$. On suppose enfin que, pour tout $g \in G$, on a $gF = F$.

1. Soit $K(a,b,c) = F(a,0,b,c)$. Montrer que K est un polynôme en a^2, b^2, c^2 et ab.

2. Montrer qu'il existe $N \in \mathbb{R}[u,v,w]$ et $\alpha \in \mathbb{N}$ tels que, pour tous x et y de $\mathbb{R}^2$, x non nul

$$F(x,y) = \frac{N(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle)}{\langle x,x\rangle^\alpha}.$$

3. Soit $R \in \mathbb{R}[u,v,w]$ tel que, pour tous x et y de $\mathbb{R}^2$, on ait $R(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle) = 0$. Montrer que $R = 0$.

4. En déduire que F est de la forme $H(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle)$, où $H \in \mathbb{R}[u,v,w]$.

(École normale supérieure)

Solution.

1. K est un polynôme en a, b, c. En considérant $g = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, on obtient $K(a,b,c) = K(a,b,-c)$, donc K ne contient que des puissances paires de c. De même, $g = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$ conduit à $K(-a,-b,c) = K(a,b,c)$. On en déduit que si le coefficient du monôme $a^i b^j c^{2k}$ est non nul alors $i+j$ est pair et les entiers i et j sont de même parité. Si $i = 2i'$ et $j = 2j'$ sont pairs, $a^i b^j c^{2k} = (a^2)^{i'}(b^2)^{j'}(c^2)^k$; s'ils sont tous les deux impairs, $i = 2i'+1$ et $j = 2j'+1$, $a^i b^j c^{2k} = (a^2)^{i'}(b^2)^{j'}(c^2)^k ab$. Ainsi K est un polynôme en a^2, b^2, c^2 et ab.

2. On suppose que $x = (x_1, x_2) \neq (0,0)$ et on considère une matrice de rotation telle que $g(x_1,x_2) = (\sqrt{x_1^2+x_2^2}, 0) = (\|x\|, 0)$ afin d'utiliser K. On note $a = \|x\|$ et $gy = (b,c)$ et $F(x,y) = K(a,b,c)$. Il existe d'après la question précédente $L \in \mathbb{R}[u,v,w,t]$ tel que $K(a,b,c) = L(a^2, b^2, c^2, ab)$. Comme g conserve la norme et le produit scalaire, on a

$$a^2 = \langle x, x\rangle, \quad ab = \langle x, y\rangle \quad \text{et} \quad b^2 + c^2 = \|y\|^2,$$

ou bien que

$$b^2 = \frac{\langle x,y\rangle^2}{\langle x,x\rangle} \quad \text{et} \quad c^2 = \langle y,y\rangle^2 - \frac{\langle x,y\rangle^2}{\langle x,x\rangle}.$$

Ainsi, on a donc

$$F(x,y) = L\left(\langle x,x\rangle, \frac{\langle x,y\rangle^2}{\langle x,x\rangle}, \langle y,y\rangle - \frac{\langle x,y\rangle^2}{\langle x,x\rangle}, \langle x,y\rangle\right).$$

On note α le degré total de L en (v,w). En multipliant $F(x,y)$ par $\langle x,x\rangle^\alpha$, on obtient un polynôme en $\langle x,x\rangle$, $\langle x,y\rangle$ et $\langle y,y\rangle$ que l'on note N.

3. Soit λ et μ deux réels strictement positifs, $\theta \in \mathbb{R}$, $x = \left(\sqrt{\lambda}, 0\right)$, $y = (\sqrt{\mu}\cos\theta, \sqrt{\mu}\sin\theta)$. On a alors $\langle x,x\rangle = \lambda$, $\langle y,y\rangle = \mu$ et $\langle x,y\rangle = \sqrt{\lambda\mu}\cos\theta$. On a donc $R(\lambda, \mu, \sqrt{\lambda\mu}\cos\theta) = 0$. Pour tout λ et μ strictement positifs, le polynôme d'une variable $R(\lambda,\mu,w)$ est nul, car il possède une infinité de racines. On en déduit que, pour tout $\lambda > 0$, $\nu \in \mathbb{R}$, le polynôme d'une variable $R(\lambda, v, \nu)$ est nul, car il possède une infinité de racines puis que, pour tous réels μ et ν le polynôme d'une variable $R(u,\mu,\nu)$ est nul. On a donc $R(\lambda,\mu,\nu) = 0$ pour tous réels λ, μ, ν, donc $R = 0$.

4. En échangeant le rôle de x et de y, la question 2 nous donne l'existence d'un polynôme $P \in \mathbb{R}[u,v,w]$ et de $\beta \in \mathbb{N}$ tels que pour tout couple (x,y) avec $y \neq 0$, on ait

$$F(x,y) = \frac{P(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle)}{\langle y,y\rangle^\beta}.$$

Si l'on note $\mathrm{R} = v^\beta \mathrm{N}(u,v,w) - u^\alpha \mathrm{P}(u,v,w) \in \mathbb{R}[u,v,w]$, on a $\mathrm{R}(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle) = 0$ pour x et y non nuls dans $\mathbb{R}^2$. Cela reste encore vrai pour x ou y nul par continuité. La question précédente nous assure que le polynôme formel R est nul. L'égalité $v^\beta \mathrm{N}(u,v,w) = u^\alpha \mathrm{P}(u,v,w)$ implique que si le coefficient du monôme $u^i v^j w^k$ dans $\mathrm{N}(u,v,w)$ est non nul, alors $i \geqslant \alpha$. Ainsi u^α se factorise dans N et il existe $\mathrm{H} \in \mathbb{R}[u,v,w]$ tel que $\mathrm{H}(u,v,w) = \dfrac{\mathrm{N}(u,v,w)}{u^\alpha}$ et finalement pour x non nul,

$$\mathrm{F}(x,y) = \mathrm{H}(\langle x,x\rangle, \langle y,y\rangle, \langle x,y\rangle),$$

formule qui reste valable pour $x = 0$ par continuité. $\lhd$

Le commutant du groupe orthogonal $\mathrm{O}(\mathrm{E})$ *dans* $\mathcal{L}(\mathrm{E})$ *est réduit aux homothéties : en effet, considérons un endomorphisme u commutant avec tout élément du groupe orthogonal de* E. *Si x est un vecteur non nul de* E *et si u commute avec la réflexion s d'hyperplan* $x^\perp$, *alors u laisse stable le sous-espace propre* $\mathrm{Ker}(s + \mathrm{Id_E}) = \mathbb{R}x$. *Autrement dit, tout vecteur non nul est un vecteur propre pour u, qui est donc nécessairement une homothétie. Dans l'exercice suivant, on retrouve ces idées pour résoudre une équation fonctionnelle portant sur le groupe orthogonal.*

1.39. Équation fonctionnelle faisant intervenir le groupe orthogonal

Déterminer les applications $f : \mathbb{R}^n \longrightarrow \mathcal{M}_n(\mathbb{R})$ telles que :

$$\forall \mathrm{X} \in \mathbb{R}^n, \quad \forall \mathrm{O} \in \mathrm{O}_n(\mathbb{R}), \quad f(\mathrm{OX}) = \mathrm{O}f(\mathrm{X})\,{}^t\mathrm{O}.$$

(**École normale supérieure**)

$\rhd$ **Solution.**

Analyse. Soit X un vecteur non nul de $\mathbb{R}^n$. Pour obtenir des informations sur la matrice $f(\mathrm{X})$, il est naturel de considérer les matrices orthogonales O telles que $\mathrm{OX} = \mathrm{X}$. On identifie les matrices et les endomorphismes de $\mathbb{R}^n$ qu'elles représentent dans la base canonique. Considérons O_1 la symétrie orthogonale par rapport à la droite $\mathbb{R}\mathrm{X}$. On a $f(\mathrm{O}_1\mathrm{X}) = f(\mathrm{X}) = \mathrm{O}_1 f(\mathrm{X})\,{}^t\mathrm{O}_1$ et $f(\mathrm{X})\mathrm{O}_1 = \mathrm{O}_1 f(\mathrm{X})$: les endomorphismes $f(\mathrm{X})$ et O_1 commutent et donc les sous-espaces propres $\mathrm{Ker}(\mathrm{O}_1 - \mathrm{I}_n)$ et $\mathrm{Ker}(\mathrm{O}_1 + \mathrm{I}_n)$ sont stables par $f(\mathrm{X})$. Il en résulte que X est un vecteur propre de $f(\mathrm{X})$ et que l'hyperplan $\mathrm{H} = \mathrm{X}^\perp$ est stable par $f(\mathrm{X})$.

Prenons maintenant $\mathrm{Y} \in \mathrm{H}$ non nul et notons O_2 la réflexion par rapport à l'orthogonal de Y. Comme X est orthogonal à Y, on a $\mathrm{O}_2\mathrm{X} = \mathrm{X}$

et donc $f(X) = O_2 f(X)\,{}^tO_2$ et finalement $f(X)O_2 = O_2 f(X)$. Comme précédemment, $f(X)$ laisse stable $\mathrm{Ker}(O_2 + I_n) = \mathbb{R}Y$. Ainsi, les vecteurs non nuls de H sont tous vecteurs propres de l'endomorphisme induit par $f(X)$ sur H. Il est classique d'en déduire que $f(X)$ restreinte à H est une homothétie.

Ainsi, on peut écrire $f(X)$ sous la forme $f(X) = a(X)p_X + b(X)I_n$ où p_X désigne la matrice du projecteur orthogonal sur la droite $\mathbb{R}X$ et où $a(X), b(X)$ sont deux réels dépendant *a priori* de X.

Pour le vecteur nul, on voit comme ci-dessus que la matrice $f(0)$ commute avec $O_n(\mathbb{R})$: c'est donc une homothétie et l'écriture précédente reste correcte.

Synthèse. Regardons maintenant si ces applications conviennent. Soit $X \in \mathbb{R}^n$ et $O \in O_n(\mathbb{R})$. On a

$$f(OX) = a(OX)p_{OX} + b(OX)I_n$$

et

$$Of(X)O^{-1} = a(X)Op_XO^{-1} + b(X)I_n.$$

On a bien $Op_XO^{-1} = p_{OX}$. Si X est nul, l'égalité est vérifiée. Supposons X non nul. Alors la famille (p_{OX}, I_n) est libre et il y a égalité si, et seulement si, $a(OX) = a(X)$ et $b(OX) = b(X)$. Si X est un vecteur fixé, l'ensemble $\{OX, O \in O_n(\mathbb{R})\}$ est la sphère de centre 0 et de rayon $\|X\|$. Il en résulte que f convient si, et seulement si, a et b sont des fonctions qui ne dépendent que de la norme euclidienne de X.

Conclusion. Les solutions sont les applications de la forme

$$X \longmapsto \lambda(\|X\|)p_X + \mu(\|X\|)I_n,$$

où λ, μ sont des applications quelconques de $\mathbb{R}_+$ dans $\mathbb{R}$. ◁

Les exercices suivants sont consacrés à la dimension 3 et notamment au produit vectoriel.

1.40. Endomorphismes conservant le produit vectoriel

Soit E un espace euclidien orienté de dimension 3 et $f \in \mathcal{L}(E)$ non nulle. Montrer que f est une rotation si, et seulement si,

$$\forall (u, v) \in E^2 \quad f(u) \wedge f(v) = f(u \wedge v).$$

(École normale supérieure)

▷ **Solution.**

Soit $\mathcal{B}$ une base orthonormale directe de E. On a alors, par définition du produit vectoriel, pour tout $(u,v,w) \in \mathrm{E}^3$, $\det_{\mathcal{B}}(u,v,w) = \langle u \wedge v, w\rangle$.

• Supposons que f soit une rotation. On a alors, pour $(u,v,w) \in \mathrm{E}^3$,

$$\begin{aligned}\langle f(u) \wedge f(v), f(w)\rangle &= \det_{\mathcal{B}}(f(u), f(v), f(w)) = \det_{\mathcal{B}}(u,v,w)\\ &= \langle u \wedge v, w\rangle = \langle f(u \wedge v), f(w)\rangle,\end{aligned}$$

car f est de déterminant 1 et conserve le produit scalaire. Puisque f est bijective, le vecteur $f(u\wedge v) - f(u)\wedge f(v)$ est orthonormal à tout vecteur de E donc, pour tout u et v de E,

$$f(u \wedge v) = f(u) \wedge f(v).$$

• Supposons réciproquement que cette condition soit réalisée.

Considérons (i,j,k) une base orthonormale directe. Comme $i \wedge j = k$, on a $f(i) \wedge f(j) = f(k)$ et donc $f(k)$ est orthogonal à $f(i)$ et $f(j)$. Comme $k \wedge i = j$, on a $f(k) \wedge f(i) = f(j)$ et on obtient finalement que $(f(i), f(j), f(k))$ forme un système orthogonal. D'autre part, on a $\|f(i)\|\|f(j)\| = \|f(k)\|$ et $\|f(k)\|\|f(i)\| = \|f(j)\|$. D'où $\|f(i)\|^2\|f(k)\| = \|f(k)\|$. Le cas $\|f(k)\| = 0$ est à exclure car alors on aurait $f(i) = f(j) = f(k) = 0$ (car $f(i) = f(j)\wedge f(k)$, puisque $i = j\wedge k$), et f serait nulle. Il s'ensuit que $\|f(i)\| = 1$ et il en va de même pour $f(j)$ et $f(k)$ par permutation circulaire. Ainsi, $(f(i), f(j), f(k))$ est une base orthonormale et elle est même directe puisque $f(k) = f(i) \wedge f(j)$. L'endomorphisme f transformant une base orthonormale directe en une autre base orthonormale directe, il s'agit d'une rotation. ◁

Dans l'exercice suivant on donne une construction abstraite du corps non commutatif des quaternions de Hamilton.

1.41. Le groupe des quaternions

Soit V un espace euclidien de dimension 3, $\mathrm{E} = \mathbb{R} \times \mathrm{V}$ muni de la structure euclidienne produit. On note $(p_0, \overrightarrow{p})$ un élément de E. On pose, pour p et q dans E,

$$pq = (p_0 q_0 - (\overrightarrow{p}, \overrightarrow{q}), p_0\overrightarrow{q} + q_0\overrightarrow{p} + \overrightarrow{p} \wedge \overrightarrow{q}).$$

On admet que cette loi est associative et que $\|pq\| = \|p\|\|q\|$.

1. On note U la sphère unité de E. Montrer que U est un groupe (appelé le groupe des quaternions).

2. On pose, pour $p \in \mathrm{U}$ et $x \in \mathrm{E}$, $a_p(x) = pxp^{-1}$. Montrer que l'application $p \longmapsto a_p$ est un morphisme de U dans le groupe des rotations de $\{0\} \times \mathrm{V}$.

(École polytechnique)

Solution.

L'associativité et la relation sur la norme, qui sont admises dans l'énoncé peuvent se vérifier à l'aide d'un logiciel de calcul formel.

1. Le produit de deux éléments de U est un élément de U car

$$\|p \cdot q\| = \|p\|\|q\|.$$

On vérifie que $1_{\mathrm{E}} = (1, \vec{0}) \in \mathrm{U}$ est élément neutre pour la multiplication. Pour tout élément $p = (p_0, \vec{p})$ de E, on note $\overline{p} = (p_0, -\vec{p})$. Si $p = (p_0, \vec{p}) \in \mathrm{U}$, on a

$$p\overline{p} = (p_0^2 + \|\vec{p}\|^2, -p_0\vec{p} + p_0\vec{p} - \vec{p} \wedge \vec{p}) = 1_{\mathrm{E}},$$

car $p_0^2 + \|\vec{p}\|^2 = \|(p_0, \vec{p})\|^2 = 1$. On montre de même que $\overline{p} \cdot p = 1_{\mathrm{E}}$. Donc $\overline{p}$ est l'inverse de p et U est un groupe multiplicatif.

2. Notons $\mathrm{P} = \{0\} \times \mathrm{V}$. Il faut démontrer que, si p appartient à U et x à P, alors $a_p(x)$ est dans P. On utilise la caractérisation suivante : si $x \in \mathrm{E}$, alors $x \in \mathrm{P}$ si, et seulement si, $\overline{x} = -x$. Pour $p = (p_0, \vec{p})$ et $q = (q_0, \vec{q})$ dans E, on a

$$\begin{aligned}\overline{pq} &= (p_0q_0 - \langle\vec{p}, \vec{q}\rangle, -p_0\vec{q} - q_0\vec{p} - \vec{p} \wedge \vec{q}) \\ &= (q_0p_0 - \langle\vec{q}, \vec{p}\rangle, -q_0\vec{p} - p_0\vec{q} + \vec{q} \wedge \vec{p}) = \overline{q}.\overline{p}.\end{aligned}$$

On en déduit que, pour $p \in \mathrm{U}$ et $x \in \mathrm{P}$,

$$\overline{a_p(x)} = \overline{px\overline{p}} = p\,\overline{x}\,\overline{p} = -p\,x\,\overline{p} = -a_p(x).$$

Ceci montre que $a_p(x)$ appartient à P. D'autre part, l'application a_p est clairement linéaire. Enfin, pour tout $x \in \mathrm{P}$,

$$\|a_p(x)\| = \|p\|\|x\|\|\overline{p}\| = \|x\|,$$

car p et $\overline{p}$ sont dans U. L'application a_p est un endomorphisme orthogonal de P.

Cherchons les éléments de P invariants par a_p. On a, pour $x \in \mathrm{P}$, $a_p(x) = x$ si, et seulement si, $px = xp$. Si $p = (p_0, \vec{p})$ et $x = (0, \vec{q})$, on obtient

$$\begin{aligned}px &= (-\langle\vec{p}, \vec{q}\rangle, p_0\vec{q} + \vec{p} \wedge \vec{q}) \\ xp &= (-\langle\vec{q}, \vec{p}\rangle, p_0\vec{q} + \vec{q} \wedge \vec{p}).\end{aligned}$$

On a donc

$$xp = px \iff \overrightarrow{p} \wedge \overrightarrow{q} = \overrightarrow{0} \iff \overrightarrow{q} \in \mathrm{Vect}(\overrightarrow{p}).$$

Si $\overrightarrow{p} \neq \overrightarrow{0}$, l'application a_p est une rotation de P, d'axe $\mathrm{Vect}(\overrightarrow{p})$. Si $\overrightarrow{p} = \overrightarrow{0}$, et donc $p = \pm 1_{\mathrm{E}}$, a_p est l'identité. L'application $\varphi : p \longmapsto a_p$ va de U dans SO(P).

On a d'autre part, $a_p \circ a_{p'}(x) = pp'x\overrightarrow{p'}\overrightarrow{p} = pp'x\overrightarrow{pp'} = a_{pp'}(x)$. L'application $\varphi : p \longmapsto a_p$ est donc un morphisme de groupes de U dans SO(P). $\lhd$

On peut remarquer que si $p = (0, \overrightarrow{p}) \in \mathrm{P} \cap \mathrm{U}$, *on a*

$$p^2 = p(-\overline{p}) = -1_{\mathrm{E}} \quad et \quad a_{p^2} = \mathrm{Id}_{\mathrm{P}}\,.$$

Ainsi, a_p *est le retournement d'axe* $\mathrm{Vect}(\overrightarrow{p})$. *Le groupe* $\varphi(\mathrm{U})$ *contient donc tous les retournements de* P *et comme les retournements engendrent* SO(P), *l'application* φ *est surjective.*

D'autre part, il résulte de l'étude de a_p *que le noyau de* φ *est* $\{1_{\mathrm{E}}, -1_{\mathrm{E}}\}$. *Le groupe* SO(P) *est donc isomorphe à* $\mathrm{U}/\{-1_{\mathrm{E}}, 1_{\mathrm{E}}\}$. *L'espace vectoriel euclidien* P *est isomorphe à* V *et donc à* $\mathbb{R}^3$. *On obtient donc que* $\mathrm{SO}_3(\mathbb{R})$ *est isomorphe à* $\mathrm{U}/\{-1, 1\}$.

On peut montrer que $(\mathrm{E}, +, \times)$ *est un corps non commutatif appelé corps des quaternions. C'est aussi une* $\mathbb{R}$*-algèbre de dimension* 4.

Voici une équation fonctionnelle portant sur $\mathbb{R}^2$ *muni de sa structure euclidienne canonique.*

1.42. Une équation fonctionnelle

Trouver toutes les applications V de $(\mathbb{R}^2 \setminus \{0\}) \times \mathbb{R}^2$ vers $\mathbb{R}$ telles que, pour tout $\mathrm{A} \in \mathrm{SL}_2(\mathbb{R})$, pour tout X et Y de $\mathbb{R}^2$, avec $\mathrm{X} \neq 0$, pour tout réel a et b, avec $a \neq 0$, on ait (la norme étant la norme euclidienne usuelle de $\mathbb{R}^n$)

$$\mathrm{V}(a\mathrm{AX}, a^2\mathrm{AY} + b\mathrm{AX}) = \frac{\|\mathrm{X}\|}{\|\mathrm{AX}\|}\mathrm{V}(\mathrm{X}, \mathrm{Y}).$$

(École normale supérieure)

▷ **Solution.**

On note (e_1, e_2) la base canonique de $\mathbb{R}^2$ et det le déterminant dans cette base.

• *Analyse.* Supposons que (X, Y) soit une famille libre de $\mathbb{R}^2$. Montrons

que l'on peut trouver $A \in SL_2(\mathbb{R})$, $a \in \mathbb{R}^*$ et $b \in \mathbb{R}$ tels que $aAe_1 = X$ et $a^2Ae_2 + bAe_1 = Y$. Ceci équivaut à $Ae_1 = \frac{1}{a}X$ et $Ae_2 = \frac{1}{a^2}(Y - \frac{b}{a}X)$. On a alors

$$\det A = \det\left(\frac{1}{a}X, \frac{1}{a^2}(Y - \frac{b}{a}X)\right) = \frac{1}{a^3}\det(X, Y).$$

Comme $\det(X, Y) \neq 0$, on peut choisir a pour que $\det A = 1$: a doit être la racine cubique de $\det(X, Y)$. Pour un tel choix de a, b et A (en fait, b est quelconque), on obtient

$$\begin{aligned} V(X, Y) = V(aAe_1, a^2Ae_2 + bAe_1) &= \frac{\|e_1\|}{\|Ae_1\|}V(e_1, e_2) = \frac{|a|}{\|X\|}V(e_1, e_2) \\ &= \frac{|\det(X, Y)|^{\frac{1}{3}}}{\|X\|}V(e_1, e_2). \end{aligned}$$

Montrons que cette formule reste valable si (X, Y) est liée, c'est-à-dire que l'on obtient alors $V(X, Y) = 0$.

Supposons d'abord que $Y = 0$. En choisissant $A \in SL_2(\mathbb{R})$ telle que $AX = 2X$ (c'est possible), $a = 1$, $b = 0$, on arrive à $V(2X, 0) = \frac{1}{2}V(X, Y)$, puis avec $AX = X$, $a = 2$ et $b = 0$, on obtient $V(2X, 0) = V(X, Y)$. On en déduit que $V(X, Y) = 0$.

Si $Y = \lambda X$, on obtient avec $A = I_2$, $a = 1$ et $b = -\lambda$, $V(X, Y - \lambda X) = V(X, Y)$ et donc $V(X, Y) = V(X, 0) = 0$.

En posant $k = V(e_1, e_2)$, on obtient, pour tout X et Y dans $\mathbb{R}^2$,

$$\boxed{V(X, Y) = k\frac{|\det(X, Y)|^{\frac{1}{3}}}{\|X\|}}.$$

• *Synthèse.* On montre, réciproquement que toute application de cette forme convient. Pour $A \in SL_2(\mathbb{R})$, $a \in \mathbb{R}$ et $b \in \mathbb{R}$, on a en effet

$$\begin{aligned} V(aAX, a^2AY + bAX) &= k\frac{|\det(aAX, a^2AY + bAX)|^{\frac{1}{3}}}{\|aAX\|} \\ &= k\frac{|a^3\det(AX, AY)|^{\frac{1}{3}}}{\|aAX\|} = k\frac{|\det(X, Y)|^{\frac{1}{3}}}{\|AX\|} \\ &= \frac{\|X\|}{\|AX\|}V(X, Y). \quad \triangleleft \end{aligned}$$

Nous terminons la série d'exercices sur les espaces euclidiens par le théorème de projection sur un convexe fermé (et quelques applications) : la distance d'un point d'un espace euclidien (ou plus généralement d'un

espace de Hilbert) à un convexe fermé non vide est atteinte en un unique point. En cela, il prolonge le théorème de projection orthogonale sur un sous-espace (de dimension finie). C'est un résultat essentiel pour la théorie des espaces de Hilbert et l'étude des opérateurs en analyse fonctionnelle. Si la vision géométrique de cette propriété s'impose, des arguments d'analyse (compacité ou complétude) sont requis pour prouver l'existence du projeté.

1.43. Projection sur un convexe fermé

Soit E un espace euclidien et $\mathcal{C}$ un convexe fermé non vide de E, $a \in \mathrm{E}$ et $d = d(a, \mathcal{C})$.

1. Établir le *lemme de la médiane* : pour trois points A, B et C de E, et I le milieu de [B, C] on a

$$\|\overrightarrow{\mathrm{AB}}\|^2 + \|\overrightarrow{\mathrm{AC}}\|^2 = 2\|\overrightarrow{\mathrm{AI}}\|^2 + \frac{1}{2}\|\overrightarrow{\mathrm{BC}}\|^2.$$

2. Montrer qu'il existe $h \in \mathcal{C}$ tel que $\|h - a\| = d$. Vérifier l'unicité de h. Le point h est le *projeté orthogonal de a sur $\mathcal{C}$*.

3. Soit $x \in \mathcal{C}$. Démontrer que $\langle h - a, h - x\rangle \leqslant 0$.

4. En déduire lorsque $a \notin \mathcal{C}$, l'existence d'un demi-espace fermé qui contient $\mathcal{C}$ mais pas a.

(**École polytechnique**)

▷ **Solution.**

1. Il s'agit d'une réécriture de l'identité du parallélogramme : on a

$$\begin{aligned} 2\|\overrightarrow{\mathrm{AB}}\|^2 + 2\|\overrightarrow{\mathrm{AC}}\|^2 &= \|\overrightarrow{\mathrm{AC}} + \overrightarrow{\mathrm{AB}}\|^2 + \|\overrightarrow{\mathrm{AC}} - \overrightarrow{\mathrm{AB}}\|^2 \\ &= 4\|\overrightarrow{\mathrm{AI}}\| + \|\overrightarrow{\mathrm{BC}}\|^2 \end{aligned}$$

car $\overrightarrow{\mathrm{AB}} + \overrightarrow{\mathrm{AC}} = 2\overrightarrow{\mathrm{AI}}$, I étant le milieu de [B, C]. En divisant par 2, on obtient le résultat.

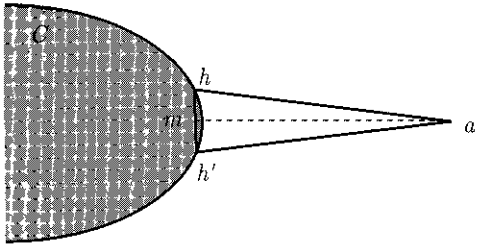
2. *Existence.* Il existe une suite $(h_n)_{n\in\mathbb{N}}$ de $\mathcal{C}$ telle que $\|h_n - a\|$ converge vers d. Cette suite est bornée, puisque $\|h_n\| \leqslant \|h_n - a\| + \|a\|$. Elle admet donc une sous-suite convergente d'après le théorème de Bolzano-Weierstrass (E est de dimension finie). Cette sous-suite converge vers un point h qui est donc dans l'adhérence de $\mathcal{C}$. Le convexe $\mathcal{C}$ étant fermé, h est dans $\mathcal{C}$ et par passage à la limite. $\|h - a\| = d$.

Unicité. La convexité intervient ici. Considérons deux points h et h' de $\mathcal{C}$ tels que $\|h - a\| = \|h' - a\| = d$. Introduisons $m = \dfrac{h + h'}{2}$ qui correspond au milieu de h et h' et qui est un point de $\mathcal{C}$ par convexité.

D'après le lemme de la médiane portant sur les points a, h et h', on obtient

$$\|h-a\|^2 + \|h'-a\|^2 = 2\|m-a\|^2 + \frac{1}{2}\|h'-h\|^2.$$

Dans ces conditions, $2\|m-a\|^2 = 2d^2 - \frac{1}{2}\|h'-h\|^2$ et comme $\|m-a\| \geqslant d$, on a $2d^2 \leqslant 2\|m-a\|^2$ et nécessairement $\|h'-h\| = 0$: les points h et h' sont confondus.



3. Pour utiliser la convexité, on va écrire que pour tout $y \in [h,x]$, $\|y-a\| \geqslant \|h-a\|$. Plus précisément, si $t \in [0,1]$, $y = (1-t)h + tx$ décrit $[a,x] \subset C$ et on a donc l'inégalité

$$\begin{aligned}\|h-a\|^2 \leqslant \|(1-t)h + tx - a\|^2 &= \|-t(h-x) + h - a\|^2 \\ &\leqslant t^2\|h-x\|^2 - 2t\langle h-x, h-a\rangle + \|h-a\|^2,\end{aligned}$$

ce qui donne $0 \leqslant t^2\|h-x\| - 2t\langle h-x, h-a\rangle$. Si on l'avait $\langle h-x, h-a\rangle > 0$, la quantité $t^2\|h-x\| - 2t\langle h-x, h-a\rangle \underset{t\to 0^+}{\sim} -2t\langle h-x, h-a\rangle$ serait strictement négative au voisinage de 0^+, ce qui est exclu. Par conséquent, on a prouvé que $\langle h-x, h-a\rangle \leqslant 0$.

4. Comme a n'est pas dans C, $a \neq h$ et $\langle h-x, h-a\rangle \leqslant 0$ est l'équation d'un demi-espace fermé dans lequel on retrouve C d'après la question précédente. Sa frontière est l'hyperplan[4] passant par h, de direction l'orthogonal de $h-a$. Il ne contient pas a, puisque $\langle h-a, h-a\rangle = \|h-a\|^2 > 0$. ◦

Dans le tome 3 d'Analyse, nous verrons que pour prouver l'existence de h, nous pouvons utiliser des arguments de complétude (plutôt que de s'appuyer sur la compacité comme nous l'avons fait ici à travers le théorème de Bolzano-Weierstrass). Ainsi, les résultats de cet exercice restent valables si C est un convexe fermé d'un espace de Hilbert (de dimension finie ou pas).

Nous proposons maintenant deux exercices où le théorème de projection sur un convexe fermé se révèle incontournable.

[4] Un tel hyperplan est appelé hyperplan d'appui au convexe C en h.

1.44. Lemme de Farkas

Soit $J_1, \ldots, J_p$ des points de $\mathbb{R}^n$ muni de sa structure canonique d'espace euclidien. On note C l'ensemble des vecteurs $\sum_{i=1}^{p} \lambda_i J_i$ où les λ_i sont des réels positifs ou nuls quelconques.

1. Montrer que C est un cône convexe fermé. On note

$$C' = \{X \in \mathbb{R}^n, \forall Y \in C, \langle X, Y\rangle \leqslant 0\}$$

$$C'' = \{X \in \mathbb{R}^n, \forall Y \in C', \langle X, Y\rangle \leqslant 0\}.$$

2. Justifier l'inclusion $C \subset C''$. On veut démontrer que $C = C''$. Soit $X \in C''$ et P le projeté orthogonal de X sur le convexe fermé C (voir l'exercice précédent). Démontrer que P et $X - P$ sont orthogonaux. Démontrer que $X - P \in C'$. Conclure.

3. Soit $B \in \mathbb{R}^n$. On suppose que pour tout $X \in \mathbb{R}^n$ vérifiant pour chaque $1 \leqslant i \leqslant p$, $\langle J_i, X\rangle \geqslant 0$, on a $\langle B, X\rangle \geqslant 0$. Montrer qu'il existe $\lambda_1, \ldots, \lambda_p$ dans $\mathbb{R}_+$ tels que

$$B = \sum_{i=1}^{p} \lambda_i J_i.$$

(**École polytechnique**)

▷ **Solution.**

1. Un barycentre à coefficients positifs de deux éléments de C est clairement une combinaison linéaire à coefficients positifs des J_i, donc un élément de C : l'ensemble C est convexe. Il s'agit bien d'un cône, puisque C est stable par multiplication par un réel positif.

Il reste à vérifier que C est fermé et pour cela, nous allons procéder par récurrence sur $p \geqslant 0$. Pour $p = 0$, $C = \{0\}$ est bien fermé. Supposons maintenant $p \geqslant 1$ et le résultat vrai au rang $p-1$. Penchons-nous d'abord sur le cas particulier où pour chaque k, $-J_k$ est dans C. Comme C est stable par homothétie de rapport positif et par addition, C est clairement le sous-espace engendré par les J_i. En particulier, C est fermé, puisqu'il est de dimension finie.

Supposons désormais que l'un des opposés des J_k, disons $-J_p$ pour fixer les idées, ne soit pas dans C. Notons $C_0 = \sum_{i=1}^{p-1} \mathbb{R}_+ J_i$ l'ensemble des combinaisons linéaires à coefficients positifs de $J_1, \ldots, J_{p-1}$. Soit $(X_k)_{k\in\mathbb{N}}$ une suite de C convergente vers un vecteur X. Il s'agit de montrer que $X \in C$. Écrivons pour tout k, $X_k = Y_k + \lambda_k J_p$ avec $Y_k \in C_0$ et $\lambda_k \in \mathbb{R}_+$.

Supposons que λ_k tende vers $+\infty$ lorsque k tend vers $+\infty$. Pour k assez grand, λ_k est non nul et

$$\frac{Y_k}{\lambda_k} + J_p = \frac{X_k}{\lambda_k} \xrightarrow[k\to\infty]{} 0,$$

puisque $(X_k)_{k\geqslant 0}$ est bornée. Il s'ensuit que $-J_p = \lim_{k\to\infty} \frac{Y_k}{\lambda_k}$ et $-J_p$, en tant que limite de points de C_0, est aussi un point de C_0, puisque cet ensemble est fermé par hypothèse de récurrence. Or ce cas a été exclu. On en déduit que λ_k ne peut diverger vers $+\infty$, et par conséquent elle admet une sous-suite majorée, donc bornée, et d'après le théorème de Bolzano-Weierstrass, elle admet une sous-suite $(\lambda_{\varphi(k)})_{k\in\mathbb{N}}$ ($\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ strictement croissante) convergente vers $\lambda \geqslant 0$. Dans ces conditions, C_0 étant fermé par hypothèse de récurrence, $Y_{\varphi(k)}$ converge vers $Y = X - \lambda J_p$ qui est donc un point de C_0. Au final, $X = Y + \lambda J_p$ est bien un élément de C.

2. Soit $X \in C$. Pour tout $Y \in C'$, on a par définition de C', $\langle X, Y\rangle \leqslant 0$. Donc X appartient à C'' par définition de C''. Donc, on a $C \subset C''$.

Soit $X \in C''$. Considérons le projeté orthogonal P de X sur le convexe fermé C.

• Montrons que P et $X - P$ sont orthogonaux. C'est trivial si $P = 0$. Supposons donc P non nul. Nous allons utiliser le fait que C est un cône en écrivant que pour tout $\lambda \geqslant 0$, $\lambda P \in C$ et donc

$$\|X - P\|^2 \leqslant \|X - \lambda P\|^2,$$

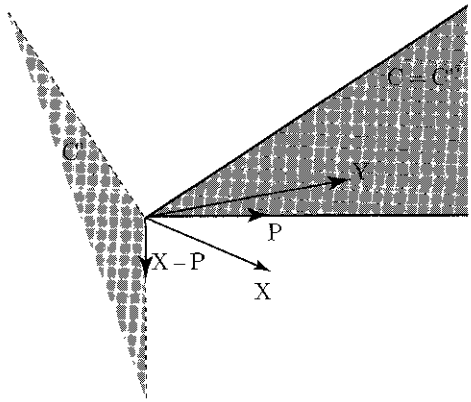
ce qui donne en développant

$$0 \leqslant \lambda^2\|P\|^2 - 2\lambda\langle P, X\rangle + 2\langle P, X\rangle - \|P\|^2.$$

Il s'agit là d'un trinôme du second degré admettant 1 comme racine. Or cette racine est aussi un minimum, autrement dit, $\lambda = 1$ est une racine double, également racine de la dérivée du trinôme : 1 est racine de $2\lambda\|P\|^2 - 2\langle X, P\rangle$. On en déduit que $\|P\|^2 = \langle X, P\rangle$ et $\langle X - P, P\rangle = 0$.

• Comme P est le projeté orthogonal de X sur C, d'après la question 3 de l'exercice précédent, pour tout $Y \in C$, $\langle Y - P, X - P\rangle \leqslant 0$. Or, comme P et $X - P$ sont orthogonaux, cela donne $\langle Y, X - P\rangle \leqslant 0$, autrement dit $X - P \in C'$ (voir figure ci-après).

• Comme $X \in C''$ par hypothèse, on obtient $\langle X, X - P\rangle \leqslant 0$ qui entraîne $\langle X - P, X - P\rangle \leqslant 0$, puisque P et $X - P$ sont orthogonaux. On conclut que $X = P \in C$, et finalement $C = C''$.



3. Il s'agit de montrer que $B \in C$. Notons $C_B = \mathbb{R}_+ B$ et comme précédemment

$$C'_B = \{X \in \mathbb{R}^n, \forall Y \in C_B, \langle X, Y\rangle \leqslant 0\}$$
$$\text{et } C''_B = \{X \in \mathbb{R}^n, \forall Y \in C'_B, \langle X, Y\rangle \leqslant 0\}.$$

Remarquons que $C'_B = \{X \in \mathbb{R}^n, \langle B, X\rangle \leqslant 0\}$. Traduisons l'hypothèse sur ces ensembles : nous avons $C' \subset C'_B$. En effet, prenons $X \in C'$. Alors $-X$ a un produit scalaire positif ou nul avec chaque J_k et par hypothèse, il en va de même avec B. Ainsi, $\langle B, X\rangle \leqslant 0$ et on a bien $X \in C'_B$.

Comme nous avons $C' \subset C'_B$, il est clair que $C''_B \subset C''$ et d'après la question précédente, $C_B \subset C$. On conclut que $B \in C$. ◁

L'intérêt de ce résultat apparaît en programmation linéaire, lorsqu'il s'agit d'obtenir des solutions «optimales » pour un système linéaire avec contrainte. On peut d'ailleurs traduire en termes de systèmes ce lemme : soit A *une matrice réelle à n lignes et p colonnes de colonnes* $J_1, \ldots, J_p$ *et le système* $AX = B$ *d'inconnue* $X \in \mathbb{R}^p$. *Pour que ce système possède une solution à composantes positives, il suffit que pour tout* $Y \in \mathbb{R}^n$ *vérifiant* ${}^tYJ_i \geqslant 0$ *(*$1 \leqslant i \leqslant p$*), on ait* ${}^tYB \geqslant 0$.

L'exercice suivant vient également de la programmation linéaire.

1.45. Éléments de $\mathbb{R}^n$ à composantes positives

Soit $E = \mathbb{R}^n$ muni de sa structure euclidienne canonique,

$$A^* = \{(x_1, \ldots, x_n) \in E \setminus \{0\},\ \ \forall i \in [\![1, n]\!]\ \ x_i \geqslant 0\}$$

et S un sous-espace de E.

1. Soit $C = \left\{(x_1, \ldots, x_n) \in (\mathbb{R}_+)^n, \sum_{i=1}^{n} x_i = 1\right\}$ et p la projection orthogonale sur S. Démontrer que $K = p(C)$ est un convexe compact non vide.

2. En déduire que $S \cap A^* \neq \emptyset$ ou $S^\perp \cap A^* \neq \emptyset$.

(**École polytechnique**)

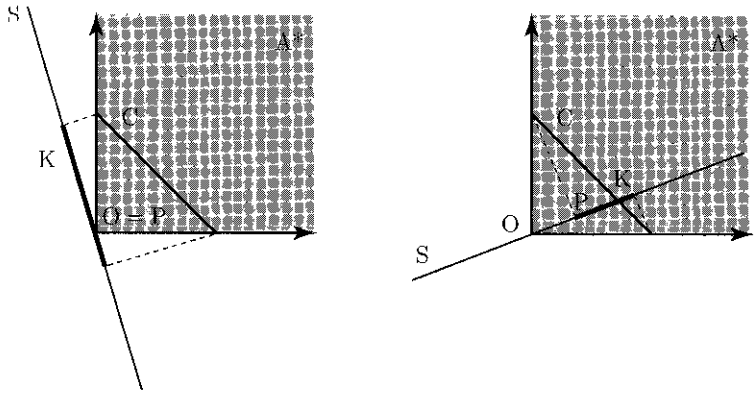
Solution.

1. Si l'on considère une suite de points de C convergente vers $X \in \mathbb{R}^n$, comme les composantes des vecteurs de cette suite sont positives ou nulles, de somme égale à 1, il en va de même des composantes de X par passage à la limite. Donc X est dans C et C est fermé. De plus, C est borné donc C est compact. Comme p est continue, $K = p(C)$ est un compact.

L'ensemble C étant l'intersection du convexe $(\mathbb{R}_+)^n$ et de l'hyperplan affine $x_1 + \cdots + x_n = 1$ (qui est convexe), il est lui-même convexe. L'image d'un convexe par une application linéaire étant un convexe, K est un convexe. Il est non vide, puisque C contient les vecteurs de la base canonique et K leurs images.

On conclut que K est un convexe compact non vide contenu dans S.

2. Considérons le projeté orthogonal P du vecteur nul sur le convexe K comme il a été introduit dans l'exercice 1.43. Par définition de K, il existe $Q \in C$ tel que $P = p(Q)$. Si $P = 0$, le point $Q \in S^\perp$ répond à la question puisqu'il est non nul à composantes positives.



Dans le cas où P est non nul, c'est P qui convient : il est dans S et à composantes positives. En effet, toujours d'après l'exercice 1.43, on a pour tout $X \in K$, $\langle 0 - P, X - P\rangle \leqslant 0$, ou encore $\|P\|^2 \leqslant \langle P, X\rangle$. C'est en particulier vrai pour $X = p(e_i)$ où $e_i \in C$ désigne le i-ième vecteur de la

base canonique. Cela donne $\|P\|^2 \leqslant \langle P, p(e_i)\rangle = \langle P, e_i\rangle$, puisque $P \in S$ et p est la projection orthogonale sur S. Autrement dit, les composantes de P sont positives. $\lhd$

Les énoncés suivants s'inscrivent dans le thème des espaces hermitiens (thème désormais à la limite du programme des classes préparatoires mais qui intéressera les candidats aux concours de recrutement de professeurs). L'étude des formes hermitiennes trouve son origine dans l'analyse fonctionnelle, plus précisément dans la théorie des espaces de Hilbert, illustrée notamment par les séries de Fourier. Étant donné un espace vectoriel complexe E, *une forme sesquilinéaire est une application* $\varphi : E^2 \longrightarrow \mathbb{C}$ *semi-linéaire à gauche et linéaire à droite (si* $\lambda \in \mathbb{C}$, $x, y \in E$, *on a* $\varphi(\lambda x, y) = \overline{\lambda}\varphi(x, y)$ *et* $\varphi(x, \lambda y) = \lambda\varphi(x, y)$*). Elle est dite hermitienne si* $\varphi(x, y) = \overline{\varphi(y, x)}$. *La forme quadratique hermitienne associée est alors* $q : x \longmapsto \varphi(x, x) \in \mathbb{R}$. *Si* $q(x) > 0$ *pour* x *non nul, on dit que* φ *est définie positive : c'est un produit scalaire hermitien. La norme hermitienne associée est* $x \longmapsto \sqrt{q(x)}$.

Le produit hermitien canonique de $\mathbb{C}^n$ *est donné par*

$$\varphi(X, Y) = \overline{x_1}y_1 + \overline{x_2}y_2 + \cdots + \overline{x_n}y_n,$$

avec $X = (x_1, \ldots, x_n)$, $Y = (y_1, \ldots, y_n) \in \mathbb{C}^n$. *Dans l'espace des fonctions continues* 2π*-périodiques à valeurs complexes, l'application*

$$(f, g) \longmapsto (f, g) = \frac{1}{2\pi}\int_0^{2\pi} \overline{f(t)}g(t)\,dt$$

définit un produit scalaire hermitien. Un espace de dimension finie muni d'un produit scalaire hermitien est un espace hermitien.

On le voit, le parallèle avec les espaces euclidiens est tout tracé : on introduit de même les notions d'orthogonalité, l'inégalité de Cauchy-Schwarz est vérifiée dans les mêmes conditions... Si bon nombre de résultats géométriques sur les espaces euclidiens trouvent une traduction sur les espaces hermitiens, nous mettrons ici plutôt en avant les différences.

1.46. Inégalités

1. Soit E hermitien, u, v, w trois vecteurs de norme 1 de E. Montrer que

$$\sqrt{1 - |\langle u, w\rangle|^2} \leqslant \sqrt{1 - |\langle u, v\rangle|^2} + \sqrt{1 - |\langle v, w\rangle|^2}.$$

2. Soit A et B deux matrices unitaires de $\mathcal{M}_n(\mathbb{C})$. Montrer que

$$\sqrt{n^2 - |\operatorname{Tr}(AB)|^2} \leqslant \sqrt{n^2 - |\operatorname{Tr}(A)|^2} + \sqrt{n^2 - |\operatorname{Tr}(B)|^2}.$$

(École normale supérieure)

Solution.

1. Les trois produits scalaires sont de module inférieur à 1 d'après l'inégalité de Cauchy-Schwarz. Si u et w sont colinéaires, on a $w = \lambda u$ avec $|\lambda| = 1$ et le premier membre de l'inégalité est nul ; il n'y a rien à démontrer. On suppose donc que la famille (u, w) est libre ; on note $P = \operatorname{Vect}(u, w)$. Soit $(\lambda, \mu) \in \mathbb{C}^2$ et $t \in P^\perp$ tel que $v = \lambda u + \mu w + t$; on pose $a = \langle u, w\rangle$ qui est de module inférieur ou égal à 1. On a alors $\langle u, v\rangle = \lambda + \mu a$ et $\langle v, w\rangle = \overline{\lambda} a + \overline{\mu}$. Par ailleurs, on obtient en écrivant que v est unitaire, $1 = |\lambda|^2 + |\mu|^2 + 2\operatorname{Re}(\overline{\lambda}\mu a) + \|t\|^2$. On en déduit que

$$\begin{aligned} 1 - |\langle u, v\rangle|^2 &= 1 - |\lambda + \mu a|^2 = 1 - |\lambda|^2 - |\mu a|^2 - 2\operatorname{Re}(\overline{\lambda}\mu a) \\ &= |\mu|^2 - |\mu a|^2 + \|t\|^2 = |\mu|^2(1 - |a|^2) + \|t\|^2, \end{aligned}$$

et de même $1 - |\langle v, w\rangle|^2 = |\lambda|^2(1 - |a|^2) + \|t\|^2$. Il s'agit donc de démontrer que

$$\sqrt{1 - |a|^2} \leqslant \underbrace{\sqrt{|\mu|^2(1 - |a|^2) + \|t\|^2} + \sqrt{|\lambda|^2(1 - |a|^2) + \|t\|^2}}_{A}.$$

En notant A le second membre de l'inégalité précédente, on a

$$\begin{aligned} A^2 &\geqslant |\mu|^2(1 - |a|^2) + |\lambda|^2(1 - |a|^2) + 2\|t\|^2 + 2|\lambda||\mu|(1 - |a|^2) \\ &\geqslant (|\lambda| + |\mu|)^2(1 - |a|^2) + 2\|t\|^2 \geqslant ((|\lambda| + |\mu|)^2 + 2\|t\|^2)(1 - |a|^2). \end{aligned}$$

Comme $1 = |\lambda|^2 + |\mu|^2 + 2\operatorname{Re}(\overline{\lambda}\mu a) + \|t\|^2 \leqslant (|\lambda| + |\mu|)^2 + 2\|t\|^2$, on a bien $A^2 \geqslant 1 - |a|^2$.

2. L'application $(A, B) \longmapsto \operatorname{Tr}(A^*B)$ est une forme hermitienne sur $\mathcal{M}_n(\mathbb{C})$. Elle est définie positive. En effet, on a pour $A = (a_{i,j}) \in \mathcal{M}_n(\mathbb{C})$, $\operatorname{Tr}(A^*A) = \sum_{1 \leqslant i,j \leqslant n} |a_{i,j}|^2 \geqslant 0$ et $\operatorname{Tr}(A^*A) = 0$ si, et seulement si, $A = 0$. Il s'agit là du produit hermitien canonique de $\mathcal{M}_n(\mathbb{C})$.

Pour une matrice unitaire, on a $\operatorname{Tr}(A^*A) = n$. Ainsi $\frac{1}{\sqrt{n}}A$ est de norme 1. Le résultat demandé résulte alors directement de la première question appliquée à $u = \frac{1}{\sqrt{n}}A^*$, $v = \frac{1}{\sqrt{n}}I_n$ et $w = \frac{1}{\sqrt{n}}B$. ◁

Le produit hermitien canonique de $\mathcal{M}_n(\mathbb{C})$ vérifie des propriétés analogues à celles du produit scalaire canonique de $\mathcal{M}_n(\mathbb{R})$ (voir exercice

1.17). En particulier, si l'on note $\mathrm{N}(\mathrm{A}) = \sqrt{\mathrm{Tr}(\mathrm{A}^*\mathrm{A})}$ *la norme de* A, *on obtient en appliquant l'inégalité de Cauchy-Schwarz une majoration du module de la trace en fonction de la norme :*

$$|\operatorname{Tr}(\mathrm{A})| = |\operatorname{Tr}(\mathrm{IA})| \leqslant \mathrm{N}(\mathrm{I})\mathrm{N}(\mathrm{A}) = \sqrt{n}\mathrm{N}(\mathrm{A}).$$

D'autre part, N *est une norme d'algèbre, autrement dit on a toujours* $\mathrm{N}(\mathrm{AB}) \leqslant \mathrm{N}(\mathrm{A})\mathrm{N}(\mathrm{B})$. *En effet, posons* $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n}$, $\mathrm{B} = (b_{ij})_{1\leqslant i,j\leqslant n}$ *et* $\mathrm{C} = \mathrm{AB}$ *avec* $\mathrm{C} = (c_{ij})_{1\leqslant i,j\leqslant n}$. *Comme* $c_{ij} = \sum_{k=1}^{n} a_{ik}b_{kj}$, *on a*

$$\begin{aligned}
\mathrm{N}(\mathrm{AB})^2 &= \sum_{1\leqslant i,j\leqslant n} |c_{ij}|^2 = \sum_{1\leqslant i,j\leqslant n} \left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|^2 \\
&\leqslant \sum_{1\leqslant i,j\leqslant n} \left(\sum_{k=1}^{n} |a_{ik}|^2 \sum_{l=1}^{n} |b_{lj}|^2\right) \quad \textit{(inégalité de Cauchy-Schwarz)} \\
&\leqslant \left(\sum_{1\leqslant i,k\leqslant n} |a_{ik}|^2\right)\left(\sum_{1\leqslant j,l\leqslant n} |b_{lj}|^2\right) = \mathrm{N}(\mathrm{A})^2\mathrm{N}(\mathrm{B})^2.
\end{aligned}$$

Voici maintenant une inégalité propre aux espaces hermitiens.

1.47. Étude de normes sur $\mathcal{L}(\mathrm{E})$ où E est hermitien

Soit E un espace hermitien de dimension $n \geqslant 1$. Si $u \in \mathcal{L}(\mathrm{E})$, on note $\mathrm{N}(u) = \sup_{\|x\|=1} |\langle u(x), x\rangle|$.

1. Montrer que N est une norme sur $\mathcal{L}(\mathrm{E})$.

2. On pose $\|u\| = \sup_{\|x\|=1} \|u(x)\|$. Montrer que

$$\frac{1}{2}\|u\| \leqslant \mathrm{N}(u) \leqslant \|u\|.$$

(École normale supérieure)

▷ **Solution.**

1. N est bien à valeurs réelles positives ou nulles. L'homogénéité est vérifiée : si $\lambda \in \mathbb{C}$, $u \in \mathcal{L}(\mathrm{E})$, on a

$$\begin{aligned} N(\lambda u) &= \sup_{\|x\|=1} |\langle \lambda u(x), x\rangle| = \sup_{\|x\|=1} |\lambda| |\langle u(x), x\rangle| = |\lambda| \sup_{\|x\|=1} |\langle u(x), x\rangle| \\ &= |\lambda| N(u). \end{aligned}$$

Prouvons l'inégalité triangulaire : si u et v sont dans $\mathcal{L}(E)$, pour tout $x \in E$ de norme 1, on a

$$\begin{aligned} |\langle (u+v)(x), x\rangle| &= |\langle u(x), x\rangle + \langle v(x), x\rangle| \leqslant |\langle u(x), x\rangle| + |\langle v(x), x\rangle| \\ &\leqslant N(u) + N(v), \end{aligned}$$

ce qui entraîne $N(u+v) \leqslant N(u) + N(v)$.

Il reste à montrer l'axiome de séparation à savoir que pour $u \in \mathcal{L}(E)$, si $N(u) = 0$, alors $u = 0$. Bien que l'on souligne la plupart du temps les similitudes entre la théorie des espaces euclidiens et celle des espaces hermitiens, l'analogue de ce résultat dans un espace euclidien est inexact : par exemple dans un plan euclidien une rotation d'angle $\pi/2$ correspondrait à une borne supérieure nulle.

Dans un espace euclidien deux formes bilinéaires différentes φ et φ' peuvent donner la même forme quadratique $q(x) = \varphi(x,x) = \varphi'(x,x)$. Par exemple, dans un plan euclidien, si u désigne la rotation d'angle $\pi/2$, $\varphi = 0$ et $\varphi' : (x,y) \longmapsto \langle u(x), y\rangle$, φ et φ' déterminent la même forme quadratique $q = 0$. Bien sûr, si l'on suppose que φ et φ' sont symétriques et qu'elles donnent la même forme quadratique, $\varphi = \varphi'$ d'après l'identité de polarisation :

$$\varphi(x,y) = \frac{1}{2}\left(q(x+y) - q(x) - q(y)\right).$$

En revanche, dans un espace hermitien, si l'on dispose de deux formes sesquilinéaires (c'est-à-dire semi-linéaires à gauche, linéaires à droite) φ et φ', que l'on ne suppose pas hermitiennes, telles que pour tout x, $\varphi(x,x) = \varphi'(x,x)$, alors $\varphi = \varphi'$. C'est une conséquence de l'identité de polarisation suivante.

Lemme. *Soit $\varphi : (x,y) \in E^2 \longmapsto \varphi(x,y) \in \mathbb{C}$ sesquilinéaire, $q : x \in E \longmapsto \varphi(x,x)$. Alors, pour tout $(x,y) \in E^2$, on a*

$$\varphi(x,y) = \frac{1}{4}\left(q(x+y) - q(x-y) - i(q(x+iy) - q(x-iy))\right).$$

Démonstration.

Il suffit de développer les termes suivants :

$$q(x+y) = q(x) + q(y) + \varphi(x,y) + \varphi(y,x)$$

$$q(x-y) = q(x) + q(y) - \varphi(x,y) - \varphi(y,x)$$

$$q(x+iy) = q(x) - q(y) + i\varphi(x,y) - i\varphi(y,x)$$
$$q(x-iy) = q(x) - q(y) - i\varphi(x,y) + i\varphi(y,x).$$

La relation est alors immédiate. $\diamondsuit$

Soit $u \in \mathcal{L}(\mathrm{E})$ tel que $\mathrm{N}(u) = 0$. On a donc $\langle u(x), x\rangle = 0$ pour tout $x \in \mathrm{E}$. Posons $\varphi(x,y) = \langle u(x), y\rangle$ pour tout $(x,y) \in \mathrm{E}^2$ et $q(x) = \varphi(x,x)$. Alors φ vérifie les hypothèses du lemme. Comme $q = 0$, d'après l'identité de polarisation $\varphi = 0$. En particulier, si $x \in \mathrm{E}$ et $y = u(x)$, on obtient $\langle u(x), y\rangle = \|u(x)\|^2 = 0$. Il s'ensuit que $u(x) = 0$ et que u est nul. On conclut que N est une norme sur $\mathcal{L}(\mathrm{E})$.

2. Soit $u \in \mathcal{L}(\mathrm{E})$. On peut supposer u non nul. Si $x \in \mathrm{E}$, $\|x\| = 1$, on a, en vertu de l'inégalité de Cauchy-Schwarz,

$$|\langle u(x), x\rangle| \leqslant \|u(x)\|\|x\| = \|u(x)\| \leqslant \|u\|.$$

Par conséquent, $\mathrm{N}(u) \leqslant \|u\|$.

Passons à l'autre inégalité : comme $u \neq 0$, on peut poser $v = \dfrac{u}{\|u\|}$ et alors $\|v\| = 1$. Par homogénéité, il suffit de prouver que $\mathrm{N}(v) \geqslant \dfrac{1}{2}$. Considérons la forme sesquilinéaire définie par $\varphi(x,y) = \langle v(x), y\rangle$ pour tout $(x,y) \in \mathrm{E}^2$ et posons $q(x) = \varphi(x,x)$. Pour $x \in \mathrm{E}$ non nul, on a

$$|q(x)| = |\langle v(x), x\rangle| = \|x\|^2 \left|\left\langle v\left(\frac{x}{\|x\|}\right), \frac{x}{\|x\|}\right\rangle\right| \leqslant \|x\|^2 \mathrm{N}(v),$$

cette inégalité restant valable si $x = 0$. Par conséquent, si $(x,y) \in \mathrm{E}^2$, d'après le lemme, on a la majoration

$$\begin{aligned}|\varphi(x,y)| &\leqslant \frac{1}{4}\left(|q(x+y)| + |q(x-y)| + |q(x+iy)| + |q(x-iy)|\right)\\ &\leqslant \frac{1}{4}\mathrm{N}(v)\left(\|x+y\|^2 + \|x-y\|^2 + \|x+iy\|^2 + \|x-iy\|^2\right)\\ &\leqslant \frac{1}{4}\mathrm{N}(v)\left(4\|x\|^2 + 4\|y\|^2\right) \text{ après développement.}\end{aligned}$$

Si l'on prend x de norme 1 et $y = v(x)$, il vient

$$|\varphi(x,y)| = \|v(x)\|^2 \leqslant \frac{1}{4}\mathrm{N}(v)\left(4 + 4\|v(x)\|^2\right) \leqslant 2\mathrm{N}(v),$$

puisque $\|v\| = 1$. En passant à la borne supérieure sur les vecteurs x de norme 1, on obtient $\|v\| = 1 \leqslant 2\mathrm{N}(v)$ et finalement l'inégalité voulue. On conclut

$$\boxed{\frac{1}{2}\|u\| \leqslant \mathrm{N}(u) \leqslant \|u\|}. \quad \triangleleft$$

Dans l'exercice suivant, on utilise le fait qu'une matrice unitaire est diagonalisable dans une base orthonormée et que ses valeurs propres sont de module 1.

1.48. Condition suffisante pour que deux matrices unitaires commutent

On munit $\mathcal{M}_n(\mathbb{C})$ de la norme $\|A\| = \sqrt{\mathrm{Tr}(A^*A)}$. On note $\mathrm{U}_n(\mathbb{C})$ le groupe unitaire d'ordre n.

1. Montrer que

$$\forall U \in \mathrm{U}_n(\mathbb{C}),\ \forall A \in \mathcal{M}_n(\mathbb{C}),\ \|UA\| = \|AU\| = \|A\|.$$

On considère dans la suite A, B deux matrices de $\mathrm{U}_n(\mathbb{C})$ et on pose $C = ABA^{-1}B^{-1}$. On suppose que $AC = CA$, que $\|I - B\| < \sqrt{2}$ et on veut montrer que $AB = BA$.

2. Montrer que A et BAB^{-1} commutent.

3. Montrer qu'il existe $U \in \mathrm{U}_n(\mathbb{C})$, des complexes $\lambda_1, \ldots, \lambda_n$ de module 1 et une permutation σ de $[\![1, n]\!]$ tels que $U^*AU = D$ et $U^*BAB^{-1}U = D'$ où $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ et $D' = \mathrm{Diag}(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(n)})$.

4. Conclure.

(École normale supérieure)

Solution.

Notons pour commencer que l'application $(A, B) \longmapsto \mathrm{Tr}(A^*B)$ est un produit scalaire hermitien sur $\mathcal{M}_n(\mathbb{C})$ et que $\|\ \|$ est tout simplement la norme associée à ce produit hermitien (voir page 85).

1. Pour $U \in \mathrm{U}_n(\mathbb{C})$ et $A \in \mathcal{M}_n(\mathbb{C})$, on a

$$\mathrm{Tr}((UA)^*(UA)) = \mathrm{Tr}(A^*U^*UA) = \mathrm{Tr}(A^*A)$$

donc $\|UA\| = \|A\|$ et de même,

$$\mathrm{Tr}((AU)^*(AU)) = \mathrm{Tr}(U^*A^*AU) = \mathrm{Tr}(A^*AUU^*) = \mathrm{Tr}(A^*A)$$

donc $\|AU\| = \|A\|$.

2. Il s'agit de montrer que $ABAB^{-1} = BAB^{-1}A$ ou encore en passant aux inverses que $BA^{-1}B^{-1}A^{-1} = A^{-1}BA^{-1}B^{-1}$, ce qui s'écrit $A^{-1}CA^{-1} = A^{-1}A^{-1}C$. Or cette dernière relation est vraie, car $CA = AC$ entraîne $A^{-1}C = CA^{-1}$: les matrices A et BAB^{-1} commutent.

3. Les matrices A et BAB^{-1} sont unitaires donc diagonalisables en base orthonormée. Comme elles commutent, elles sont codiagonalisables en base orthonormée. Considérons en effet les endomorphismes a et b canoniquement associés à A et BAB^{-1}. Les sous-espaces propres de a sont

deux à deux orthogonaux et stables par b. La restriction de b à l'un de ces sous-espaces propres F est un endomorphisme unitaire, diagonalisable en base orthonormale. On peut donc trouver une base orthonormale de F constituée de vecteurs propres pour b (et pour a). La réunion de ces bases orthonormales donne une base orthonormale de $\mathbb{C}^n$ constituée de vecteurs propres pour a et b.

Il existe donc $U \in U_n(\mathbb{C})$, des complexes $\lambda_1, \ldots, \lambda_n$ de module 1, et des complexes $\lambda'_1, \ldots, \lambda'_n$ de module 1 tels que $U^*AU = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n) = D$ et $U^*BAB^{-1}U = \mathrm{Diag}(\lambda'_1, \ldots, \lambda'_n) = D'$. Mais comme A et BAB^{-1} sont semblables, il existe une permutation σ de $[\![1, n]\!]$ telle que $\lambda'_i = \lambda_{\sigma(i)}$ pour tout $i \in [\![1, n]\!]$.

4. Il ne reste plus qu'à faire intervenir la dernière hypothèse qui n'a pas encore servi, à savoir que $\|I - B\| < \sqrt{2}$. Supposons qu'il existe i tel que $\lambda_{\sigma(i)} \neq \lambda_i$. Si l'on note e_i le i-ième vecteur de la base canonique de $\mathbb{C}^n$ et $f_i = Ue_i$ on a : $Af_i = \lambda_i f_i$ et $BAB^{-1}f_i = \lambda_{\sigma(i)}f_i$ ou encore, $AB^{-1}f_i = \lambda_{\sigma(i)}B^{-1}f_i$. Donc $B^{-1}f_i$ est orthogonal à f_i, car les sous-espaces propres de A sont deux à deux orthogonaux. Ainsi, $Bf_i \perp f_i$ et en vertu du théorème de Pythagore, on a $\|(B-I)f_i\|^2 = \|Bf_i\|^2 + \|f_i\|^2 = 2$ (f_i est unitaire). Cela contredit le fait que $\|I - B\| < \sqrt{2}$.

Conclusion. $A = BAB^{-1}$ *i.e.* A et B commutent. $\lhd$

Le lecteur trouvera beaucoup plus d'exercices sur les espaces hermitiens à la fin du chapitre suivant.

Chapitre 2

Réduction des endomorphismes auto-adjoints

Le théorème spectral est le résultat central de la théorie des endomorphismes auto-adjoints. Il affirme que si u est un endomorphisme symétrique d'un espace euclidien E, alors u est diagonalisable dans une base orthonormée de E. En termes matriciels, cela signifie que pour une matrice symétrique réelle A, il existe une matrice orthogonale P telle que $\mathrm{P}^{-1}\mathrm{AP}$ soit diagonale. Comme il y a une correspondance bijective entre les endomorphismes symétriques de E et les formes bilinéaires symétriques[1]*, on peut le traduire aussi dans le langage des formes quadratiques : pour toute forme φ bilinéaire symétrique de l'espace euclidien E, il existe une base orthonormée pour le produit scalaire de E et orthogonale pour φ. Matriciellement, l'écriture $\mathrm{P}^{-1}\mathrm{AP} = \mathrm{D}$ avec P orthogonale peut aussi se lire ${}^{t}\mathrm{PAP} = \mathrm{D}$: les matrices A et D sont non seulement semblables mais aussi congruentes.*

Ce deuxième aspect du théorème sera surtout développé dans le chapitre 3. Même si une forme quadratique peut apparaître à l'occasion d'un exercice (notamment pour ceux qui font intervenir des matrices symétriques positives), nous avons regroupé ici les énoncés qui portent plutôt sur des propriétés spectrales : diagonalisation et applications diverses, décompositions matricielles (QR, décomposition polaire, décomposition de Choleski, ...) principes variationnels et inégalités concernant les valeurs propres des matrices symétriques, réduction des matrices antisymétriques. La fin du chapitre regroupe des énoncés posés dans un cadre hermitien. La richesse des points de vue se retrouvera dans les exercices : nous verrons qu'il y a souvent plusieurs solutions possibles à un même énoncé.

Pour conclure cette introduction, notons que les variantes du théorème spectral en dimension infinie dans les espaces de Hilbert sont des outils puissants de l'analyse fonctionnelle, adaptés aux équations sur des opérateurs intégraux ou au calcul des variations.

Les premiers exercices correspondent à des applications directes de la diagonalisation des endomorphismes symétriques.

---

1 En associant à u la forme bilinéaire symétrique $(x, y) \longmapsto \langle u(x), y\rangle$.

2.1. Codiagonalisation

Soit E un espace vectoriel euclidien, f, g, h des éléments de $\mathcal{L}(\mathrm{E})$ tels que $fg = h$. $gh = f$, $hf = g$, $f = f^*$, $\det f \neq 0$. Prouver que f, g et h sont diagonalisables dans une même base.

(École polytechnique)

▷ **Solution.**

L'endomorphisme f est bijectif et, puisqu'il est symétrique, diagonalisable dans une base orthonormée avec des valeurs propres réelles. On a $f = gh = gfg$ et $f^2 = (gh)f = g(hf) = g^2$. On en déduit que

$$f = gfg = g(gfg)g = g^2fg^2 = f^5$$

et donc $f^4 = \mathrm{Id}$. Si λ est une valeur propre de f, on a $\lambda^4 = 1$ et donc $\lambda^2 = 1$. L'endomorphisme f^2 est diagonalisable et a pour seule valeur propre 1 : c'est l'identité. On a donc $f^2 = g^2 = \mathrm{Id_E}$. Ainsi g, qui est une symétrie, est également diagonalisable. Enfin, l'égalité $f = gfg$ peut s'écrire $fg = gfg^2 = gf$. Les endomorphismes f et g sont diagonalisables et commutent. Il est classique alors de vérifier qu'ils sont diagonalisables dans la même base (voir l'exercice 2.19 du tome 2 d'algèbre). La matrice dans cette base de $h = fg$ est aussi diagonale. ◁

Il est bien connu que le calcul des puissances d'une matrice est relativement simple lorsque celle-ci est diagonalisable.

2.2. Puissances d'une matrice symétrique

Soit $\mathrm{A} = \begin{pmatrix} p & q & r \\ q & r & p \\ r & p & q \end{pmatrix} \in \mathcal{M}_3(\mathbb{R})$. Déterminer, lorsqu'elle existe, la limite de la suite $(\mathrm{A}^n)_{n\geqslant 0}$.

(École polytechnique)

▷ **Solution.**

La matrice A est symétrique réelle donc diagonalisable dans une base orthonormale. Le vecteur $(1, 1, 1)$ est vecteur propre pour la valeur propre $\sigma_1 = p+q+r$. On montre facilement que le polynôme caractéristique de A est

$$(\mathrm{X} - \sigma_1)(\mathrm{X}^2 - p^2 - q^2 - r^2 + pq + qr + rp) = (\mathrm{X} - \sigma_1)(\mathrm{X}^2 - \sigma_1^2 + 3\sigma_2),$$

où $\sigma_2 = pq + qr + rp$. Les valeurs propres de A sont donc σ_1 et

$\sqrt{\sigma_1^2 - 3\sigma_2}$ (il n'est pas utile de vérifier que $\sigma_1^2 - 3\sigma_2$ est positif, car nous le savons *a priori*). Il existe $P \in O_3(\mathbb{R})$ tel que $A = PD\,{}^tP$, où $D = \mathrm{Diag}(\sigma_1, \sqrt{\sigma_1^2 - 3\sigma_2}, -\sqrt{\sigma_1^2 - 3\sigma_2})$. On peut choisir $\frac{1}{\sqrt{3}}\begin{pmatrix}1\\1\\1\end{pmatrix}$ comme première colonne de P. La suite $(A^n)_{n\geqslant 0}$ possède une limite si, et seulement si, les trois valeurs propres appartiennent à $]-1,1]$, c'est-à-dire si $\sigma_1 \in]-1,1]$ et $\sigma_1^2 - 3\sigma_2 < 1$.

• Si $\sigma_1 \in]-1,1[$ et $\sigma_1^2 - 3\sigma_2 < 1$, la suite $(A^n)_{n\geqslant 0}$ converge clairement vers 0.

• Si $\sigma_1 = 1$ et $\sigma_1^2 - 3\sigma_2 < 1$, alors la suite $(D^n)_{n\geqslant 0}$ converge vers $J = \begin{pmatrix}1&0&0\\0&0&0\\0&0&0\end{pmatrix}$ et $(A^n)_{n\geqslant 0}$ converge vers $PJ\,{}^tP$. La matrice $J\,{}^tP$ est égale à $\begin{pmatrix}a&b&c\\0&0&0\\0&0&0\end{pmatrix}$, où (a,b,c) est la première ligne de ${}^tP$, c'est-à-dire $\left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$. Connaissant la première colonne de P, on obtient

$$PJ\,{}^tP = \frac{1}{3}\begin{pmatrix}1&1&1\\1&1&1\\1&1&1\end{pmatrix}. \quad \lhd$$

On notera que, dans le second cas, la limite est la matrice de la projection orthogonale sur l'espace propre de A *relatif à la valeur propre* 1.

L'exercice suivant s'intéresse à la convergence d'une suite arithmético-géométrique de matrices. La technique est usuelle : on retranche le point fixe pour se ramener à une suite géométrique et c'est alors le rayon spectral de la matrice qui gouverne le comportement de la suite. Le lecteur se reportera aux exercices 2.59 et 2.60 du tome algèbre 2 pour des énoncés généraux.

2.3. Méthode itérative pour une équation linéaire

Soit $A \in \mathcal{M}_p(\mathbb{R})$ une matrice symétrique définie positive, $b \in \mathbb{R}^p$, $u = A^{-1}b$, $x_0 \in \mathbb{R}^p$ et $\alpha > 0$. On considère la suite définie par

$$\forall n \in \mathbb{N}, \quad x_{n+1} = x_n + \alpha(b - Ax_n).$$

1. Trouver une condition nécessaire et suffisante sur $\alpha > 0$ pour que la suite $(x_n)_{n\in\mathbb{N}}$ converge pour tout choix de x_0.

2. On pose $e_n = a - x_n$. Déterminer le plus petit réel $K(\alpha)$ tel que, tout choix de x_0 et pour tout $n \in \mathbb{N}$, on ait $\|e_{n+1}\| \leqslant K(\alpha)\|e_n\|$ (il s'agit de la norme euclidienne canonique).

(**École polytechnique**)

▷ **Solution.**

1. Si la suite $(x_n)_{n\in\mathbb{N}}$ converge, sa limite ℓ vérifie $\alpha(b - A\ell) = 0$ et donc $\ell = A^{-1}b = a$. On pose donc $e_n = a - x_n$. On a, pour tout $n \in \mathbb{N}$,

$$e_{n+1} = e_n - \alpha(b - AX_n) = e_n - \alpha Ae_n = (I_p - \alpha A)e_n.$$

On en déduit, pour tout n, $e_n = (I_p - \alpha A)^n e_0$. La suite $(x_n)_{n\in\mathbb{N}}$ converge pour tout choix de x_0 si, et seulement si, la suite $(I_p - \alpha A)^n$ tend vers 0.

La matrice A étant symétrique réelle, il existe $P \in O_p(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_p)$ tels que $A = {}^tPDP$. On a alors $I_p - \alpha A = {}^tP\Delta P$, où $\Delta = \mathrm{Diag}(1-\alpha\lambda_1, \ldots, 1-\alpha\lambda_p)$ et, pour tout $n \in \mathbb{N}$, $(I_p - \alpha A)^n = {}^tP\Delta^n P$. La suite $(I_p - \alpha A)^n$ converge vers 0 si, et seulement si, Δ^n converge vers 0, c'est-à-dire si, et seulement si, pour $1 \leqslant i \leqslant p$, on a $|1 - \alpha\lambda_i| < 1$. Comme α et les λ_i sont positifs (car A est définie positive), cela équivaut à ce que pour tout i, $\alpha\lambda_i < 2$. *i.e.* $\alpha \max(\lambda_1, \ldots, \lambda_p) < 2$.

Conclusion. La suite $(x_n)_{n\in\mathbb{N}}$ converge (vers a) pour tout choix de x_0 si, et seulement si,

$$\alpha < \frac{2}{\max(\lambda_1, \ldots, \lambda_p)} = \frac{2}{\rho(A)}.$$

2. On cherche le plus petit réel $K(\alpha)$ tel que $\|e_{n+1}\| \leqslant K(\alpha)\|e_n\|$, pour tout x_0. Comme $e_{n+1} = (I_p - \alpha A)e_n$ pour tout n, il est clair que $K(\alpha) = |||I_p - \alpha A|||$ convient. C'est la plus petite valeur possible, car on souhaite qu'elle soit valable pour toute condition initiale e_0. D'après le cours, comme $I_p - \alpha A$ est symétrique, on a

$$K(\alpha) = \rho(I_p - \alpha A) = \max(|1 - \alpha\lambda_1|, \ldots, |1 - \alpha\lambda_p|). \quad \triangleleft$$

Le lecteur se reportera à la note qui suit l'exercice 1.5 pour la preuve du fait que $|||A||| = \rho(A)$ pour une matrice symétrique.

Nous mettons ici quelques exercices supplémentaires sur les projecteurs orthogonaux. Ce sont des endomorphismes symétriques, et donc redevables du théorème spectral.

2.4. Décomposition en somme de droites et plans stables

Soit p et q deux projecteurs orthogonaux d'un espace vectoriel euclidien E. Montrer que E est somme directe orthogonale de sous-espaces vectoriels de dimension 1 ou 2 stables par p et q.

(École normale supérieure)

Solution.

Il semble clair qu'il faut raisonner par récurrence sur la dimension de E. La propriété est évidente pour $n = 1$ et $n = 2$. Si l'on suppose qu'elle est vérifiée pour un espace de dimension inférieure ou égale à $n - 1$ et que l'on considère un espace euclidien E de dimension n, il suffit alors de démontrer l'existence d'un sous-espace F de dimension 1 ou 2 stable par p et q. En effet, les endomorphismes p et q étant symétriques (puisque ce sont des projecteurs orthogonaux), $F^\perp$ est également stable par p et q : pour $x \in F$ et $y \in F^\perp$, $(p(y), x) = (y, p(x)) = 0$, car $p(x) \in F$; donc $p(y) \in F^\perp$ et $F^\perp$ est stable par p et il en est de même pour q. La dimension de $F^\perp$ est inférieure ou égale à $n - 1$ et il est stable par p et q. On peut lui appliquer l'hypothèse de récurrence. On obtient une décomposition de $F^\perp$ en sous-espaces stables par p et q de dimension 1 ou 2, à laquelle on ajoute F pour obtenir la décomposition cherchée de E.

Il reste à démontrer l'existence de F. Un sous-espace stable par p et q qui contient x contient aussi $p(x)$ et $q(x)$: si sa dimension est inférieure ou égale à 2, il y a donc une relation de liaison entre x, $p(x)$ et $q(x)$. Cela suggère de considérer l'endomorphisme $p + q$; il est symétrique, donc ses valeurs propres sont réelles. Soit $\lambda \in \mathrm{Sp}(p + q)$ et x un vecteur propre associé. On a $p(x) + q(x) = \lambda x$. Considérons $F = \mathrm{Vect}(x, p(x))$. C'est un sous-espace de dimension 1 ou 2. Comme $p(p(x)) = p(x)$, F est stable par p. Les deux relations $q(x) = \lambda x - p(x)$ et $q(p(x)) = \lambda q(x) - q^2(x) = (\lambda - 1)q(x)$ montrent que F est stable par q, ce qui achève la démonstration. ◁

Bien entendu au lieu de $p + q$ on peut prendre n'importe quelle combinaison linéaire de p et q à coefficients non nuls.

2.5. Spectre de la différence de deux projecteurs orthogonaux

Soit E un espace euclidien de dimension $2n$. F et F′ deux sous-espaces supplémentaires de E qui sont de dimension n. On note π (resp. π') le projecteur orthogonal sur F (resp. F′). $(\sigma_1, \ldots, \sigma_n, 0, \ldots, 0)$ les valeurs propres de $\pi - \pi\pi'\pi$. Montrer que les valeurs propres de $\pi - \pi'$ sont les $\pm\sqrt{\sigma_i}$.

(École polytechnique)

▷ **Solution.**

Remarquons que $\pi - \pi'$ et $\pi - \pi\pi'\pi$ sont auto-adjoints et donc diagonalisables en base orthonormée. En particulier, le spectre de ces endomorphismes est réel. Comme $\mathrm{Ker}(\pi - \pi\pi'\pi)$ contient $\mathrm{F}^\perp = \mathrm{Ker}\,\pi$, il est de dimension supérieure ou égale à $\dim \mathrm{F}^\perp = n$. Cela explique que 0 est valeur propre de $\pi - \pi\pi'\pi$ avec une multiplicité supérieure ou égale à n.

Soit $\mathcal{B} = (\mathcal{B}_1, \mathcal{B}_2)$ une base de E obtenue en réunissant une base $\mathcal{B}_1$ de F et une base $\mathcal{B}_2$ de F'. On note P (resp. P') la matrice de π (resp. π') dans la base $\mathcal{B}$. Elles s'écrivent par blocs de la manière suivante

$$\mathrm{P} = \begin{pmatrix} \mathrm{I}_n & \mathrm{A} \\ 0 & 0 \end{pmatrix} \quad \text{et} \quad \mathrm{P}' = \begin{pmatrix} 0 & 0 \\ \mathrm{B} & \mathrm{I}_n \end{pmatrix}.$$

Dans ces conditions, la matrice dans $\mathcal{B}$ de $\pi - \pi\pi'\pi$ est

$$\mathrm{P} - \mathrm{PP'P} = \begin{pmatrix} \mathrm{I}_n - \mathrm{AB} & \mathrm{A} - \mathrm{ABA} \\ 0 & 0 \end{pmatrix}$$

et son polynôme caractéristique est $\chi_{\pi-\pi\pi'\pi} = \mathrm{X}^n \det((\mathrm{X}-1)\mathrm{I}_n + \mathrm{AB})$. Par hypothèse, on a donc $\prod_{i=1}^{n}(\mathrm{X} - \sigma_i) = \det((\mathrm{X}-1)\mathrm{I}_n + \mathrm{AB})$.

Quant à $\pi - \pi'$, sa matrice dans $\mathcal{B}$ est $\begin{pmatrix} \mathrm{I}_n & \mathrm{A} \\ -\mathrm{B} & -\mathrm{I}_n \end{pmatrix}$ et son polynôme caractéristique est, d'après un résultat classique (que le lecteur trouvera dans l'exercice 1.27 du tome 2 d'algèbre),

$$\chi_{\pi-\pi'} = \det((\mathrm{X}-1)(\mathrm{X}+1)\mathrm{I}_n^2 + \mathrm{AB}) = \det((\mathrm{X}^2-1)\mathrm{I}_n + \mathrm{AB}).$$

Ainsi, on a l'identité $\chi_{\pi-\pi'} = \prod_{i=1}^{n}(\mathrm{X}^2 - \sigma_i)$. Les racines de $\chi_{\pi-\pi'}$ sont réelles comme nous l'avons dit en introduction, donc les σ_i sont tous positifs et les racines de $\chi_{\pi-\pi'}$ sont les $\pm\sqrt{\sigma_i}$. ◁

L'énoncé suivant est un cas particulier de l'exercice 2.13, un peu plus simple à traiter.

2.6. Produit de deux projecteurs orthogonaux

Soit E un espace euclidien et p, q deux projecteurs orthogonaux de E. Montrer que pq est diagonalisable et que son spectre est inclus dans $[0, 1]$.

(École polytechnique)

Solution.

On peut noter que le résultat est évident lorsque p et q commutent, car dans ce cas pq est aussi un projecteur.

Soit λ une valeur propre non nulle de pq et x un vecteur propre associé. On a $p(q(x)) = \lambda x$ donc $x \in \operatorname{Im} p$. Par suite, comme $p = p^*$,

$$\lambda \|x\|^2 = (pq(x), x) = (q(x), p(x)) = (q(x), x) = \|q(x)\|^2$$

ce qui prouve que λ est positif et aussi que $\lambda \leqslant 1$, puisque $\|q(x)\| \leqslant \|x\|$. On a donc $\operatorname{Sp} pq \subset [0,1]$.

Pour la diagonalisabilité on va travailler matriciellement. En prenant une base orthonormale adaptée à la décomposition $\mathrm{E} = \operatorname{Im} q \oplus \operatorname{Ker} q$ on peut supposer que q a pour matrice $\mathrm{B} = \begin{pmatrix} \mathrm{I}_r & 0 \\ 0 & 0 \end{pmatrix}$ avec $r = \operatorname{rg} q$ et on note $\mathrm{A} \quad \begin{pmatrix} \mathrm{A}_1 & {}^t\mathrm{A}_2 \\ \mathrm{A}_2 & \mathrm{A}_3 \end{pmatrix}$ la matrice de p dans la même base (on laisse de côté les cas triviaux $r = 0$ et $r = n$). Alors pq a pour matrice $\mathrm{AB} = \begin{pmatrix} \mathrm{A}_1 & 0 \\ \mathrm{A}_2 & 0 \end{pmatrix}$. La matrice A_1 est symétrique donc diagonalisable et le spectre de AB est la réunion de $\{0\}$ et du spectre de A_1. Il est clair que pour toutes les valeurs propres non nulles de A_1 la dimension du sous-espace propre de AB est égale à la multiplicité de cette valeur propre. Il nous faut donc juste prouver que $\operatorname{rg} \mathrm{AB} = \operatorname{rg} \mathrm{A}_1$. Cela revient à montrer que si $\mathrm{A}_1\mathrm{X} = 0$ alors $\mathrm{A}_2\mathrm{X} = 0$ (pour X vecteur colonne de taille r). Ici c'est facile car la propriété $\mathrm{A}^2 = \mathrm{A}$ donne $\mathrm{A}_1^2 + {}^t\mathrm{A}_2\mathrm{A}_2 = \mathrm{A}_1$. Donc si $\mathrm{A}_1\mathrm{X} = 0$ on a ${}^t\mathrm{A}_2\mathrm{A}_2\mathrm{X} = 0$ puis $\|\mathrm{A}_2\mathrm{X}\|^2 = 0$ soit $\mathrm{A}_2\mathrm{X} = 0$ en multipliant par ${}^t\mathrm{X}$ à gauche. ◁

L'exercice 2.13 montrera que si f, g sont des endomorphismes autoadjoints positifs, alors $f \circ g$ est toujours diagonalisable avec des valeurs propres positives.

Le sujet de l'exercice suivant a été développé dans la seconde épreuve écrite du concours Polytechnique MP en 2002.

2.7. Frame d'un espace euclidien

Soit E un espace euclidien, $(y_j)_{j\in\mathrm{J}}$ une famille de vecteurs de E telle qu'il existe A et B strictement positifs vérifiant

$$\forall x \in \mathrm{E}. \quad \mathrm{A}\|x\|^2 \leqslant \sum_{j\in\mathrm{J}} (x, y_j)^2 \leqslant \mathrm{B}\|x\|^2.$$

1. Montrer que la famille $(y_j)_{j\in\mathrm{J}}$ engendre E.

2. Avec $E = \mathbb{R}^2$, montrer que $y_1 = (1,0)$, $y_2 = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$, $y_3 = y_2$ conviennent.

3. Si $A = B = 1$ et si $\|y_j\| = 1$ pour tout j, montrer que $(y_j)_{j\in J}$ est une base orthonormale de E.

4. Si $A = B$ montrer que $x = \frac{1}{A}\sum_{j\in J}(x, y_j)y_j$ pour tout $x \in E$.

(**École polytechnique**)

▷ **Solution.**

1. Posons $F = \mathrm{Vect}(y_j)_{j\in J}$. Pour montrer que $F = E$ on va prouver que son orthogonal est nul. C'est ici naturel, car notre hypothèse porte sur les produits scalaires $\langle x, y_j\rangle$. Effectivement, si $x \in F^\perp$ on a pour tout $j \in J$ $\langle x, y_j\rangle = 0$ et donc $A\|x\|^2 = 0$. Par suite $\|x\| = 0$ et $x = 0$. Ainsi, $F^\perp = \{0\}$ et $F = E$: la famille $(y_j)_{j\in J}$ est génératrice.

2. Si $x = (a, b) \in \mathbb{R}^2$, on a

$$S = \sum_{j=1}^{3}\langle x, y_j\rangle^2 = a^2 + 2\left(-\frac{a\sqrt{3}}{2} - \frac{b}{2}\right)^2 = \frac{1}{2}(5a^2 + b^2 + 2ab\sqrt{3}).$$

On en déduit que

$$S = \frac{1}{2}\left(\left(b\frac{\sqrt{3}}{2} + 2a\right)^2 + a^2 + \frac{1}{4}b^2\right) \geqslant \frac{1}{8}\|x\|^2.$$

De l'inégalité $ab \leqslant \dfrac{a^2 + b^2}{2}$, on tire

$$S \leqslant \frac{1}{2}(5a^2 + b^2 + \sqrt{3}(a^2 + b^2)) \leqslant \frac{5 + \sqrt{3}}{2}\|x\|^2.$$

On a donc les inégalités voulues, avec $A = \dfrac{1}{8}$ et $B = \dfrac{5 + \sqrt{3}}{2}$.

3. On a, pour tout $x \in E$, $\|x\|^2 = \sum_{j\in J}\langle x, y_j\rangle^2$. En particulier, pour $x = y_i$, on obtient

$$1 = \sum_{j\in J}\langle y_i, y_j\rangle^2 = 1 + \sum_{j\neq i}\langle y_i, y_j\rangle^2.$$

On en déduit que $\langle y_i, y_j\rangle = 0$ pour $i \neq j$. La famille $(y_j)_{j\in J}$ est donc orthonormale. Cette famille est donc libre et comme elle est également génératrice, il s'agit d'une base orthonormale de E.

4. Soit f l'endomorphisme de E défini par $f(x) = \sum_{j\in J}\langle x, y_j\rangle y_j$ pour $x \in E$. On a, pour $(x, x') \in E^2$,

$$\langle f(x), x'\rangle = \sum_{j\in J}\langle x, y_j\rangle\langle y_j, x'\rangle = \langle x, \sum_{j\in J}\langle x', y_j\rangle y_j\rangle = \langle x, f(x')\rangle.$$

L'endomorphisme f est symétrique donc diagonalisable. Si x est un vecteur propre relatif à la valeur propre λ, on a

$$\lambda\|x\|^2 = \langle \lambda x, x\rangle = \langle f(x), x\rangle = \sum_{j\in J}\langle x, y_j\rangle^2 = A\|x\|^2,$$

par hypothèse. On a donc $\lambda = A$. L'endomorphisme f est diagonalisable et a pour seule valeur propre A. C'est donc l'homothétie de rapport A. On a, pour tout $x \in E$, $f(x) = Ax$ et donc $x = \frac{1}{A}\sum_{j\in J}\langle x, y_j\rangle y_j$. ◁

Des familles de vecteurs $(y_j)_{j\in J}$ d'un espace de Hilbert vérifiant la condition de l'énoncé sont appelées des frames (d'où le titre donné à notre exercice). Elles jouent un rôle important dans la théorie des ondelettes.

C'est un exercice classique que de montrer qu'une matrice de trace nulle à coefficients dans un corps de caractéristique nulle est semblable à une matrice à diagonale nulle (le lecteur trouvera cela dans l'exercice 7.11 du tome 1 d'algèbre). Dans l'exercice suivant, on montre que pour une matrice réelle on peut même imposer une matrice de passage orthogonale.

2.8. Matrices à termes diagonaux égaux

Soit $A \in \mathcal{M}_n(\mathbb{R})$. Montrer qu'il existe $P \in O_n(\mathbb{R})$ telle que ${}^tPAP$ ait tous ses termes diagonaux égaux.

(École polytechnique)

Solution.

Notons qu'il suffit de prouver le résultat dans le cas d'une matrice symétrique. En effet, si on écrit $A = A_1 + A_2$, avec A_1 symétrique et A_2 antisymétrique on a, pour toute matrice $P \in O_n(\mathbb{R})$,

$${}^tPAP = {}^tPA_1P + {}^tPA_2P.$$

Comme ${}^tPA_2P$ est antisymétrique, les termes de sa diagonale sont nuls : les termes diagonaux de ${}^tPAP$ sont ceux de ${}^tPA_1P$. On supposera donc désormais A symétrique. Quitte à retrancher à A la matrice scalaire

$\frac{\operatorname{Tr} A}{n} I_n$ on peut aussi supposer que $\operatorname{Tr} A = 0$. On cherche donc à montrer que A est orthogonalement semblable à une matrice de diagonale nulle.

Pour cela, on va raisonner par récurrence sur n, le cas $n = 1$ étant évident. Supposons le résultat vrai au rang $n-1$ et soit $A \in \mathcal{S}_n(\mathbb{R})$ de trace nulle. On note f l'endomorphisme de $\mathbb{R}^n$ euclidien canoniquement associé à A. On va commencer par chercher une base orthonormale $(e_1, e_2, \ldots, e_n)$ de $\mathbb{R}^n$ telle que le coefficient $(1,1)$ de la matrice de f dans cette base soit nul. Or, ce coefficient vaut $\langle f(e_1), e_1 \rangle$. Il nous suffit donc de trouver un vecteur unitaire e_1 qui est orthogonal à son image par f puis de le compléter n'importe comment en une base orthonormale. Comme f est symétrique, il est diagonalisable dans une base orthonormale $(\varepsilon_1, \ldots, \varepsilon_n)$. Notons $\lambda_1, \lambda_2, \ldots, \lambda_n$ les valeurs propres associées. On va chercher notre vecteur e_1 dans cette base de vecteurs propres. Si $e_1 = \sum_{i=1}^{n} a_i \varepsilon_i$ on a

$$\langle f(e_1), e_1 \rangle = \sum_{i=1}^{n} \lambda_i a_i^2.$$

Comme $\operatorname{Tr} f = \operatorname{Tr} A = \sum_{i=1}^{n} \lambda_i = 0$ il suffit de prendre les a_i tous égaux à $\frac{1}{\sqrt{n}}$ (afin que le vecteur e_1 soit unitaire). La matrice de f dans une base orthonormale dont le premier vecteur est e_1 est de la forme

$$B = \left(\begin{array}{c|ccc} 0 & \times & \cdots & \times \\ \hline \times & & & \\ \vdots & & A' & \\ \times & & & \end{array}\right)$$

où A' est une matrice symétrique d'ordre $n-1$ telle que $\operatorname{Tr} A' = \operatorname{Tr} B = \operatorname{Tr} A = 0$. Dit matriciellement, A est orthogonalement semblable à B et il suffit donc de prouver le résultat pour B. On applique pour cela l'hypothèse de récurrence à la matrice A'. Il existe $Q \in O_{n-1}(\mathbb{R})$ et A'' de diagonale nulle telles que ${}^tQA'Q = A''$. Posons alors $P = \begin{pmatrix} 1 & 0 \\ 0 & Q \end{pmatrix}$. Il est clair que $P \in O_n(\mathbb{R})$ et un calcul par blocs prouve que

$${}^tPBP = \left(\begin{array}{c|ccc} 0 & \times & \cdots & \times \\ \hline \times & & & \\ \vdots & & A'' & \\ \times & & & \end{array}\right).$$

C'est une matrice de diagonale nulle. Cela termine la récurrence. ◁

Voici la formulation en termes de formes quadratiques du résultat qu'on vient de montrer : si u est un endomorphisme auto-adjoint et de trace nulle d'un espace euclidien E, et si $q : x \longmapsto \langle u(x), x\rangle$ est la forme quadratique associée à u, alors on peut trouver une base orthonormale de E *formée de vecteurs q-isotropes. Le lecteur est invité à étudier l'exercice 3.3 sur l'existence de bases formées de vecteurs isotropes.*

Dans les exercices suivants, on s'intéresse aux matrices symétriques réelles positives (resp. définies positives). Ce sont les matrices A *de $\mathcal{M}_n(\mathbb{R})$ symétriques telles que la forme quadratique associée sur $\mathbb{R}^n$, à savoir $q_A : X \longmapsto {}^tXAX$, est une forme quadratique positive (resp. définie positive). À l'aide des valeurs propres, on dispose de la caractérisation suivante : la matrice symétrique réelle* A *est positive (resp. définie positive) si, et seulement si, son spectre est contenu dans $\mathbb{R}_+$ (resp. $\mathbb{R}_+^*$). Dans ces conditions, on retiendra que les coefficients diagonaux de* A *sont positifs ou nuls (resp. strictement positifs), puisqu'il s'agit des réels $q_A(e_i) = {}^te_iAe_i$ où $(e_1, \ldots, e_n)$ désigne la base canonique de $\mathbb{R}^n$.*

2.9. Caractérisation des matrices positives avec la trace

Soit $A \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique. Montrer que A est positive, si et seulement si $\mathrm{Tr}(AB) \geqslant 0$ pour toute matrice symétrique positive B de $\mathcal{M}_n(\mathbb{R})$.

(École polytechnique)

▷ **Solution.**

Il existe $P \in O_n(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, où les λ_i sont réels, tel que $A = PDP^{-1} = PD\,{}^tP$.

• Supposons A positive. Alors, les λ_i sont dans $\mathbb{R}_+$ et si B est une matrice symétrique positive, on a

$$\mathrm{Tr}(AB) = \mathrm{Tr}(PD\,{}^tPB) = \mathrm{Tr}(D({}^tPBP)) = \mathrm{Tr}(DB'),$$

avec $B' = {}^tPBP$ congruente à B. Il s'ensuit que la matrice B' est positive puisqu'elle représente la même forme quadratique que B. Si on écrit $B' = (b_{ij})_{1\leqslant i,j\leqslant n}$, il vient $\mathrm{Tr}(DB') = \sum_{i=1}^{n} \lambda_i b_{ii} \geqslant 0$, puisque chaque b_{ii} est positif ou nul. On conclut que $\mathrm{Tr}(AB) \geqslant 0$.

• Réciproquement, supposons que A n'est pas positive. Une des valeurs propres de A est strictement négative, disons λ_1 pour fixer les idées. On pose $B' = \mathrm{Diag}(1, 0, \ldots, 0)$ et $B = PB'\,{}^tP$. La matrice B est positive et symétrique et $\mathrm{Tr}(AB) = \mathrm{Tr}(DB') = \lambda_1 < 0$. L'équivalence est prouvée. ◁

L'exercice 2.13 éclairera ce résultat puisqu'il montrera que le produit AB *de deux matrices symétriques positives est diagonalisable avec des valeurs propres dans* $\mathbb{R}_+$. *Nous verrons aussi dans l'exercice 2.43 une majoration fine de* $\mathrm{Tr}(AB)$ *toujours pour* A, B *symétriques positives.*

2.10. Étude d'un sous-ensemble convexe de $\mathbb{R}^n \times \mathcal{M}_n(\mathbb{R})$

Soit $\mathcal{S}$ (resp. $\mathcal{S}_+$) l'ensemble des matrices symétriques (resp. symétriques positives) de $\mathcal{M}_n(\mathbb{R})$. On considère

$$\mathcal{C} = \{(X, S) \in \mathbb{R}^n \times \mathcal{S},\ S - X\,{}^tX \in \mathcal{S}^+\}.$$

1. Montrer que $\mathcal{C}$ est convexe.

2. Soit $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$. $k \in [\![1, n]\!]$, $\lambda_1, \ldots, \lambda_k$ des réels non nuls et $S = \mathrm{Diag}(\lambda_1^2, \ldots, \lambda_k^2, 0, \ldots, 0)$. Montrer que

$$(X, S) \in \mathcal{C} \iff \forall i \in [\![k+1, n]\!]\ \ x_i = 0 \quad \text{et} \quad \sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2} \leqslant 1.$$

3. Montrer que tout élément de $\mathcal{C}$ est barycentre à coefficients positifs d'éléments de $\{(X, S) \in \mathbb{R}^n \times \mathcal{S},\ S = X\,{}^tX\}$. On étudiera d'abord le cas $n = 1$.

(École polytechnique)

▷ **Solution.**

1. Notons que si $X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, alors $X\,{}^tX = (x_i x_j)_{1 \leqslant i,j \leqslant n}$ est symétrique. Pour X et Y dans $\mathbb{R}^n$, on a ${}^tYX\,{}^tXY = ({}^tXY)^2 \geqslant 0$. La matrice $X\,{}^tX$ est donc symétrique positive. Pour (X, S), (X', S') dans $\mathcal{C}$ et $\lambda \in [0, 1]$, on pose $(Z, T) = \lambda(X, S) + (1 - \lambda)(X', S')$. On obtient

$$\begin{aligned} T - Z\,{}^tZ &= \lambda S + (1-\lambda)S' - (\lambda X + (1-\lambda)X')\,{}^t(\lambda X + (1-\lambda)X') \\ &= \lambda S + (1-\lambda)S' - \lambda^2 X\,{}^tX - (1-\lambda)^2 X'\,{}^tX' - \lambda(1-\lambda)(X\,{}^tX' + X'\,{}^tX) \\ &= \lambda(S - X\,{}^tX) + (1-\lambda)(S' - X'\,{}^tX') \\ &\qquad + \lambda(1-\lambda)(X\,{}^tX + X'\,{}^tX' - X\,{}^tX' - X'\,{}^tX) \\ &= \lambda(S - X\,{}^tX) + (1-\lambda)(S' - X'\,{}^tX') + \lambda(1-\lambda)(X - X')\,{}^t(X - X'). \end{aligned}$$

Cette matrice est symétrique positive, car somme de matrices symétriques positives. Ainsi $\lambda(X, S) + (1 - \lambda)(X', S')$ appartient à $\mathcal{C}$, donc $\mathcal{C}$ est convexe.

2. On a, pour tout $Y = {}^t(y_1, \dots, y_n)$,

$$ {}^tY(S - X\,{}^tX)Y = {}^tYSY - ({}^tXY)^2 = \sum_{i=1}^{k} \lambda_i^2 y_i^2 - \Big(\sum_{i=1}^{n} x_i y_i\Big)^2 .$$

• Si $(X, S) \in \mathcal{C}$, on a $\sum_{i=1}^{k} \lambda_i^2 y_i^2 - \Big(\sum_{i=1}^{n} x_i y_i\Big)^2 \geqslant 0$ pour tout Y. En prenant $y_1 = \cdots = y_k = 0$, on obtient $\Big(\sum_{i=k+1}^{n} x_i y_i\Big)^2 \leqslant 0$ et donc $\sum_{i=k+1}^{n} x_i y_i = 0$, pour tout $(y_{k+1}, \dots, y_n) \in \mathbb{R}^{n-k}$. Ceci implique clairement $x_{k+1} = \cdots = x_n = 0$. L'hypothèse devient donc $\sum_{i=1}^{k} \lambda_i^2 y_i^2 - \Big(\sum_{i=1}^{k} x_i y_i\Big)^2 \geqslant 0$ pour tout $(y_1, \dots, y_k) \in \mathbb{R}^k$. En prenant $y_i = \dfrac{x_i}{\lambda_i^2}$ pour $1 \leqslant i \leqslant k$, on obtient

$$\sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2} - \left(\sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2}\right)^2 \geqslant 0.$$

Si $\sum_{i=1}^{k} \dfrac{x_i^2}{\lambda_i^2} > 0$, on simplifie et on obtient $\sum_{i=1}^{k} \dfrac{x_i^2}{\lambda_i^2} \leqslant 1$, inégalité qui reste vraie si la somme est nulle.

• Réciproquement, si les conditions $x_i = 0$ pour $i > k$ et $\sum_{i=1}^{k} \dfrac{x_i^2}{\lambda_i^2} \leqslant 1$ sont réalisées, on a, par l'inégalité de Cauchy-Schwarz,

$$\begin{aligned}\left(\sum_{i=1}^{n} x_i y_i\right)^2 &= \left(\sum_{i=1}^{k} x_i y_i\right)^2 = \left(\sum_{i=1}^{k} \frac{x_i}{\lambda_i} \lambda_i y_i\right)^2 \\ &\leqslant \left(\sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2}\right)\left(\sum_{i=1}^{k} \lambda_i^2 y_i^2\right) \leqslant \sum_{i=1}^{k} \lambda_i^2 y_i^2 .\end{aligned}$$

On en déduit que $(X, S) \in \mathcal{C}$.

3. Notons $\mathcal{F} = \{(X, S) \in \mathbb{R}^n \times \mathcal{S},\quad S = X\,{}^tX\}$. Commençons par traiter le cas $n = 1$. Soit $(x, s) \in \mathbb{R}^2$ tel que $s - x^2 \geqslant 0$. Si $x \neq 0$, on a $0 \leqslant \dfrac{x^2}{s} \leqslant 1$ et on peut écrire

$$(x, s) = \frac{x^2}{s}\left(\frac{s}{x}, \frac{s^2}{x^2}\right) + \left(1 - \frac{x^2}{s}\right)(0, 0).$$

Si $x = 0$, on a $s \geqslant 0$ et

$$(0, s) = \frac{1}{2}\Big((\sqrt{s}, s) + (-\sqrt{s}, s)\Big).$$

Pour traiter le cas général, on se ramène au cas d'une matrice S diagonale. Il existe $P \in O_n(\mathbb{R})$ et D diagonale telles que $S = PD\,{}^tP$. On peut écrire

$$S - X\,{}^tX = P(D - {}^tPX\,{}^tXP)\,{}^tP = P(D - Z\,{}^tZ)\,{}^tP,$$

où $Z = {}^tPX$. La matrice $D - Z\,{}^tZ$ est symétrique positive, car congruente à $S - X\,{}^tX$. Il suffit de démontrer le résultat pour (Z, D). En effet, si (Z, D) est barycentre à coefficients positifs d'éléments de $\mathcal{F}$, il existe un entier r, des réels positifs $\alpha_1, \ldots, \alpha_r$ de somme 1 et des vecteurs $X_1, \ldots, X_r$ tels que $(Z, D) = \sum_{i=1}^{r} \alpha_i (X_i, X_i\,{}^tX_i)$. On a alors

$$\begin{aligned}(X, S) = (PZ.\, PD\,{}^tP) &= \left(\sum_{i=1}^{r} \alpha_i PX_i, \sum_{i=1}^{r} PX_i\,{}^tX_i\,{}^tP\right) \\ &= \sum_{i=1}^{r} \alpha_i (PX_i, (PX_i)\,{}^t(PX_i)),\end{aligned}$$

ce qui est le résultat cherché pour (S, X).

On suppose dorénavant que S est diagonale. Si $S = 0$, alors $X\,{}^tX = 0$ et donc $X = 0$. Il n'y a rien à démontrer. Sinon, on peut supposer qu'il existe $k \in [\![1, n]\!]$, $\lambda_1, \ldots, \lambda_k$ non nuls tels que $S = \mathrm{Diag}(\lambda_1^2, \ldots, \lambda_k^2, 0 \ldots, 0)$. D'après la question 2, on a $x_i = 0$ si $i \in [\![k+1, n]\!]$ et $\sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2} \leqslant 1$. Cette dernière relation incite à prendre comme coefficients de la combinaison linéaire les $\frac{x_i^2}{\lambda_i^2}$.

Si $x_i \neq 0$, on peut écrire $(x_i, \lambda_i^2) = \frac{x_i^2}{\lambda_i^2}\left(\frac{\lambda_i^2}{x_i}, \frac{\lambda_i^4}{x_i^2}\right)$. On note X_i le vecteur dont la seule la coordonnée non nulle est la i-ième qui vaut $\frac{\lambda_i^2}{x_i^2}$. La matrice $X_i\,{}^tX_i$ est diagonale, le seul terme non nul de la diagonale étant le i-ième qui vaut $\frac{\lambda_i^4}{x_i^2}$. On en déduit que si $x_i \neq 0$ pour tout $i \in [\![1, k]\!]$,

$$(X, S) = \sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2}(X_i, X_i\,{}^tX_i) = \sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2}(X_i, X_i\,{}^tX_i) + \left(1 - \sum_{i=1}^{k} \frac{x_i^2}{\lambda_i^2}\right)(0, 0).$$

Si les x_i ne sont pas tous non nuls (pour $i \leqslant k$), on peut supposer qu'il existe $j \in [\![2, k]\!]$ tel que $x_i \neq 0$ si $i \leqslant j$ et $x_i = 0$ si $i > j$. En raisonnant comme précédemment, on obtient

$$(X, \mathrm{Diag}(\lambda_1^2, \ldots, \lambda_j^2, 0 \ldots 0)) = \sum_{i=1}^{j} \frac{x_i^2}{\lambda_i^2}(X_i, X_i\,{}^tX_i) + \left(1 - \sum_{i=1}^{j} \frac{x_i^2}{\lambda_i^2}\right)(0, 0)$$

et donc

$$(X,S) = \frac{x_1^2}{\lambda_1^2}(X_1,S_1) + \sum_{i=2}^{j} \frac{x_i^2}{\lambda_i^2}(X_i, X_i\,{}^tX_i) + \left(1 - \sum_{i=1}^{j} \frac{x_i^2}{\lambda_i^2}\right)(0,0) \quad (1),$$

où $S_1 = \mathrm{Diag}(\mu_1^2, \ldots, \mu_n^2)$, avec $\mu_i = \dfrac{\lambda_1\lambda_i}{x_1}$ si $i = 1$ ou $j+1 \leqslant i \leqslant k$ et $\mu_i = 0$ sinon.

Il suffit de démontrer que (X_1, S_1) peut s'écrire comme barycentre à coefficients positifs d'éléments de $\mathcal{F}$. Considérons les 2^{k-j} vecteurs-colonnes de la forme

$${}^t(\mu_1, 0, \ldots, 0, \varepsilon_{j+1}\mu_{j+1}, \ldots, \varepsilon_k\mu_k, 0, \ldots, 0),$$

où les ε_i sont égaux à ± 1. On les note $Y_1, \ldots, Y_{2^{n-k}}$. On obtient $\dfrac{1}{2^{n-k}} \displaystyle\sum_{i=1}^{2^{n-k}} Y_i = X_1$, car hormis le premier, les coefficients s'annulent deux à deux et $\dfrac{1}{2^{n-k}} \displaystyle\sum_{i=1}^{2^{n-k}} Y_i\,{}^tY_i = S_1$. En effet les coefficients diagonaux de $Y_i\,{}^tY_i$ sont égaux à ceux de S_1 et les autres sont deux à deux opposés. Ainsi, $(X_1, S_1) = \dfrac{1}{2^{n-k}} \displaystyle\sum_{i=1}^{2^{n-k}} (Y_i, Y_i\,{}^tY_i)$ et (X,S) peut s'écrire comme barycentre à coefficients positifs d'éléments de $\mathcal{F}$ comme on le voit en remplaçant (X_1, S_1) par son expression dans l'égalité (1). ◁

Dans les exercices suivants, il est souvent demandé de vérifier que si u est un endomorphisme d'un espace euclidien (resp. A une matrice de $\mathcal{M}_n(\mathbb{R})$), alors $u^* \circ u$ (resp. ${}^tAA$) est un endomorphisme symétrique positif (resp. une matrice symétrique positive). De plus, si u est un isomorphisme (resp. A inversible), $u^* \circ u$ (resp. ${}^tAA$) est défini positif (resp. définie positive). Nous détaillerons la preuve de ce fait important les premières fois et nous y ferons référence par la suite.

2.11. Caractérisation des symétries orthogonales

Soit E un espace euclidien de dimension n, s une symétrie de E.

1. Que dire de $s^* \circ s$?

2. Montrer que $P(X) = \det(X\,\mathrm{Id}_E + s^* \circ s)$ est un polynôme réciproque, c'est-à-dire vérifie $P(X) = X^n P\left(\dfrac{1}{X}\right)$ $(n = \deg P)$.

3. Montrer que $P(1) \geqslant 2^n$. À quelle condition y a-t-il égalité ?

(École polytechnique)

▷ **Solution.**

1. L'endomorphisme $s^* \circ s$ est symétrique. Pour $x \in \mathrm{E} \setminus \{0\}$, on a

$$\langle s^* \circ s(x), x\rangle = \|s(x)\|^2 > 0,$$

donc $s^* \circ s$ est défini positif. Ainsi $s^* \circ s$ est diagonalisable dans une base orthonormale et ses valeurs propres sont strictement positives. On peut noter de plus que $\det(s^* \circ s) = \det s^* \det s = (\det s)^2 = 1$, puisque $s^2 = \mathrm{Id}_\mathrm{E}$.

2. Si P est réciproque et si z est une racine non nulle de P alors $\dfrac{1}{z}$ est aussi racine de P. Or les racines de P sont les opposés des valeurs propres de $s^* \circ s$. On va donc commencer par regarder si l'ensemble de ces valeurs propres (qui sont non nulles) est stable par passage à l'inverse.

Soit λ une valeur propre de $s^* \circ s$ et x un vecteur propre. On pose $y = s(x)$. On a alors $s^*(y) = \lambda x = \lambda s(y)$. On en déduit que

$$(s^* \circ s)(y) = \frac{1}{\lambda}(s^* \circ s^*)(y) = \frac{1}{\lambda} y.$$

Autrement dit, $\dfrac{1}{\lambda}$ est aussi valeur propre de $s^* \circ s$. Si $\lambda \neq 1$, on a $\dfrac{1}{\lambda} \neq \lambda$ et $s(\mathrm{E}_\lambda) \subset \mathrm{E}_{\frac{1}{\lambda}}$, où l'on note E_λ l'espace propre relatif à la valeur propre λ. Mais le même résultat appliqué à $\dfrac{1}{\lambda}$ montre qu'on fait $s(\mathrm{E}_\lambda) = \mathrm{E}_{\frac{1}{\lambda}}$: s réalise un isomorphisme entre ces deux sous-espaces propres et les valeurs propres λ et $\dfrac{1}{\lambda}$ ont même ordre de multiplicité. Outre 1, dont on note m l'ordre de multiplicité, les valeurs propres distinctes de $s^* \circ s$ peuvent se noter $\lambda_1, \dfrac{1}{\lambda_1}, \ldots, \lambda_r, \dfrac{1}{\lambda_r}$, les valeurs propres λ_i et $\dfrac{1}{\lambda_i}$, pour $1 \leqslant i \leqslant r$ ayant même multiplicité n_i. On a alors

$$\mathrm{Q} = \det(s^* \circ s - \mathrm{X}\,\mathrm{Id}_\mathrm{E}) = (-1)^n(\mathrm{X}-1)^m \prod_{i=1}^{r} \left((\mathrm{X}-\lambda_i)\left(\mathrm{X}-\frac{1}{\lambda_i}\right)\right)^{n_i}$$

et

$$\begin{aligned}\mathrm{P} = \mathrm{Q}(-\mathrm{X}) &= (\mathrm{X}+1)^m \prod_{i=1}^{r} \left((\mathrm{X}+\lambda_i)\left(\mathrm{X}+\frac{1}{\lambda_i}\right)\right)^{n_i} \\ &= (\mathrm{X}+1)^m \prod_{i=1}^{r} \left(\mathrm{X}^2 + (\lambda_i + \lambda_i^{-1})\mathrm{X} + 1\right)^{n_i}.\end{aligned}$$

Il est alors clair sur cette expression que $\mathrm{X}^n \mathrm{P}\left(\dfrac{1}{\mathrm{X}}\right) = \mathrm{P}(\mathrm{X})$.

3. On a $\mathrm{P}(1) = 2^m \prod_{i=1}^{r} (2 + \lambda_i + \lambda_i^{-1})^{n_i}$. On vérifie aisément que pour tout $t > 0$, on a $t + t^{-1} \geqslant 2$, avec égalité si, et seulement si, $t = 1$. On

on déduit que $P(1) \geqslant 2^m 4^{n_1+\cdots+n_r}$. Comme $2\sum_{i=1}^{r} n_i + m = n$, cela donne exactement l'inégalité $P(1) \geqslant 2^n$. Il y a égalité si, et seulement si, 1 est la seule valeur propre de $s^* \circ s$. Comme $s^* \circ s$ est diagonalisable, on a alors $s^* \circ s = \mathrm{Id}_E$, c'est-à-dire $s^* = s$.

Conclusion. On a $P(1) = 2^n$ si, et seulement si, s est une symétrie orthogonale. ◁

Le résultat de l'exercice suivant est classique et très important.

2.12. Racine carrée d'un endomorphisme auto-adjoint positif

Soit E un espace euclidien et u un endomorphisme auto-adjoint positif. Montrer qu'il existe un unique endomorphisme auto-adjoint positif h tel que $u = h^2$. Montrer que h est un polynôme en u.

(**École polytechnique**)

▷ **Solution.**

Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base orthonormale de vecteurs propres de u, $\lambda_i \in \mathbb{R}_+$ étant la valeur propre associée à e_i. On définit h dans la base $\mathcal{B}$, on posant $h(e_i) = \sqrt{\lambda_i} e_i$. L'endomorphisme h est auto-adjoint, car diagonalisable dans une base orthonormale, positif et vérifie clairement $h^2 = u$.

Si k est un endomorphisme auto-adjoint positif tel que $k^2 = u$, k commute avec u, donc laisse invariants les sous-espaces propres de u. La restriction de k à E_λ, espace propre de u relatif à la valeur propre λ, est un endomorphisme auto-adjoint positif donc le carré est $\lambda \mathrm{Id}_{E_\lambda}$. Sa seule valeur propre (positive) possible est $\sqrt{\lambda}$. Comme il est diagonalisable, c'est $\sqrt{\lambda}\,\mathrm{Id}_{E_\lambda}$. Ainsi, pour tout i, $k(e_i) = \sqrt{\lambda_i} e_i = h(e_i)$ et on a donc $k = h$.

Il est facile d'exhiber un polynôme P tel que $P(\lambda) = \sqrt{\lambda}$ pour tout $\lambda \in \mathrm{Sp}\, u$ (par interpolation de Lagrange). La construction de h montre que pour un tel P on a $h = P(u)$. ◁

Le résultat admet bien entendu une traduction matricielle immédiate : si $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice symétrique positive, il existe une unique matrice B symétrique positive telle que $A = B^2$. De plus B est un polynôme en A. On généralise sans peine aux racines k-ièmes pour tout $k \geqslant 2$.

Voici une première application de l'existence de la racine carrée.

2.13. Produit de deux matrices symétriques positives

Soit $A, B \in \mathcal{M}_n(\mathbb{R})$ avec A et B symétriques positives.

1. On suppose de plus que A est définie positive. Montrer que AB est diagonalisable et que son spectre est contenu dans $\mathbb{R}_+$.

2. Montrer que le résultat de la première question reste vrai si on ne suppose plus A définie positive, mais seulement positive.

(École polytechnique)

▷ **Solution.**

1. On peut écrire $A = S^2$ avec S symétrique définie positive. Dans ces conditions, le produit AB s'écrit $AB = S^2B = S(SBS)S^{-1}$ et AB est semblable à la matrice SBS qui est symétrique et positive. En effet, $SBS = {}^tSBS$ est congruente à B (elles représentent la même forme quadratique). En particulier, AB est diagonalisable et ses valeurs propres sont positives.

Notons que la réciproque est vraie. Si M *est diagonalisable à valeurs propres positives, on peut l'écrire* $M = P^{-1}DP$ *avec* D *diagonale et* P *inversible. Il suffit alors d'écrire* $M = P^{-1}\,{}^tP^{-1}\,{}^tPDP$ *et de constater que la matrice* $A = P^{-1}\,{}^tP^{-1}$ *est symétrique définie positive et que la matrice* $B = {}^tPDP$ *est symétrique positive.*

On peut donner un autre éclairage du fait que AB *est diagonalisable. Comme* A^{-1} *est définie positive, la forme bilinéaire symétrique* $(X, Y) \longmapsto \langle X, A^{-1}Y\rangle$ *est un produit scalaire. On observe alors que* AB *définit un endomorphisme auto-adjoint pour ce produit scalaire, puisque*

$$\langle ABX, A^{-1}Y\rangle = \langle BX, Y\rangle = \langle X, BY\rangle = \langle X, A^{-1}(ABY)\rangle.$$

2. On peut encore écrire $A = S^2$, mais S est seulement symétrique positive. On utilise encore la matrice SBS qui est symétrique positive, car congruente à B. On note r le rang de SBS et on considère des vecteurs propres $X_1, \ldots, X_r$ associés aux valeurs propres strictement positives $\lambda_1, \ldots, \lambda_r$ de SBS. Pour tout $k \in [\![1, r]\!]$, on a $SBSX_k = \lambda_k X_k$ et donc $AB(SX_k) = \lambda_k SX_k$.

Comme $\mathrm{Vect}(X_1, \ldots, X_r) \cap \mathrm{Ker}(SBS) = \{0\}$, on a *a fortiori* $\mathrm{Vect}(X_1, \ldots, X_r) \cap \mathrm{Ker}\, S = \{0\}$. On en déduit que la famille $(SX_1, \ldots, SX_r)$ est libre, car la restriction de S à un supplémentaire de $\mathrm{Ker}\, S$ est injective. On a donc une famille libre de r vecteurs propres de AB associés à des valeurs propres strictement positives.

Pour conclure que AB est diagonalisable et que ses valeurs propres sont positives, il suffit de montrer que l'espace propre relatif à la valeur propre 0 est de dimension $n - r$, c'est-à-dire que $\mathrm{rg}(AB) = r = \mathrm{rg}(SBS)$.

Par construction (voir exercice 2.12), on a $\operatorname{Ker} A = \operatorname{Ker} S$. On en déduit que $\operatorname{Ker}(AB) = \operatorname{Ker}(SB)$ et donc $\operatorname{rg}(AB) = \operatorname{rg}(SB)$.

Soit T une matrice symétrique positive telle que $B = T^2$. On a, si $X \in \operatorname{Ker}(SBS)$,

$$0 = {}^tXSBSX = {}^tXST^2SX = \|TSX\|^2.$$

On en déduit $TSX = 0$ et donc $BSX = 0$. On a donc $\operatorname{Ker}(SBS) \subset \operatorname{Ker}(BS)$ et comme l'inclusion inverse est évidente, $\operatorname{Ker}(SBS) = \operatorname{Ker}(BS)$. On en déduit que

$$\operatorname{rg}(SBS) = \operatorname{rg}(BS) = \operatorname{rg}{}^t(BS) = \operatorname{rg}(SB) = \operatorname{rg}(AB). \lhd$$

On peut aussi se placer dans une base orthonormale qui diagonalise A, isoler le noyau, et raisonner par blocs comme dans l'exercice 2.6.

Présenté comme un critère original de diagonalisabilité pour une matrice réelle, l'exercice suivant n'est qu'une autre écriture de la première question de celui qui précède.

2.14. Condition de diagonalisabilité d'une matrice réelle

1. Soit $P \in GL_n(\mathbb{R})$. Montrer que ${}^tPP$ est une matrice symétrique définie positive.

2. Montrer que toute matrice S symétrique définie positive s'écrit $S = {}^tPP$ avec P inversible.

3. Soit $A \in \mathcal{M}_n(\mathbb{R})$. Montrer que A est diagonalisable si, et seulement si, il existe une matrice S symétrique définie positive telle que ${}^tA = S^{-1}AS$.

(École polytechnique)

Solution.

1. La matrice ${}^tPP$ est clairement symétrique et vérifie, pour tout $X \in \mathbb{R}^n$, ${}^tX\,{}^tPPX = \|PX\|^2 \geqslant 0$. De plus, P étant inversible, si $X \neq 0$, PX n'est pas nul et $\|PX\|^2 > 0$. Donc ${}^tPP$ est symétrique définie positive.

2. La forme quadratique q_S associée à S est définie positive et on peut donc trouver une base orthonormée de $\mathbb{R}^n$ pour q_S. Si $P \in GL_n(\mathbb{R})$ est la matrice de passage de cette base à la base canonique, on a en vertu de la formule de changement de base. $S = {}^tPI_nP = {}^tPP$.

On peut également utiliser la racine carrée de S *(voir exercice 2.12) : la matrice* S *s'écrit* $S = P^2$ *avec* P *symétrique positive. En particulier, on a* $S = {}^tPP$. *On se reportera aussi à l'exercice 2.29 où l'on montrera*

que P *peut être choisie triangulaire supérieure à coefficients diagonaux positifs.*

3. Si A est diagonalisable, il existe D diagonale et $P \in GL_n(\mathbb{R})$ telles que $A = PDP^{-1}$. On a alors

$$^tA = {}^tP^{-1}D\,{}^tP = {}^tP^{-1}P^{-1}AP\,{}^tP = (P\,{}^tP)^{-1}A(P\,{}^tP).$$

D'après la question 1, la matrice $S = P\,{}^tP$ est symétrique définie positive et $^tA = S^{-1}AS$.

Supposons réciproquement qu'il existe S symétrique définie positive telle que $^tA = S^{-1}AS$. Cette égalité s'écrit $AS = S\,{}^tA = {}^t(AS)$: la matrice $B = AS$ est symétrique et $A = BS^{-1}$. C'est une situation analogue à celle de l'exercice précédent (la matrice définie positive est cette fois à droite). Les questions préliminaires de l'énoncé invitent à écrire la matrice symétrique définie positive S^{-1} sous la forme $S^{-1} = {}^tPP$ avec $P \in GL_n(\mathbb{R})$. On obtient alors

$$A = B\,{}^tPP = P^{-1}PB\,{}^tPP.$$

La matrice $PB\,{}^tP$ est symétrique réelle donc diagonalisable et A qui lui est semblable est également diagonalisable. ◁

Si l'on dispose du théorème de réduction de Jordan, il est possible de démontrer que toute matrice de $\mathcal{M}_n(K)$ *(où* K *est un corps commutatif) est semblable à sa transposée.*

L'énoncé suivant apporte une généralisation un peu plus difficile.

2.15. Condition d'existence de m vecteurs propres indépendants

Montrer que les deux propriétés suivantes d'une matrice A de $\mathcal{M}_n(\mathbb{R})$ sont équivalentes :

(i) A possède m vecteurs propres linéairement indépendants ;

(ii) il existe une matrice symétrique $S \in \mathcal{M}_n(\mathbb{R})$ de rang m ayant des valeurs propres positives telle que $AS = S\,{}^tA$.

(École polytechnique)

▷ **Solution.**

• Si A possède m vecteurs propres indépendants, on voit en choisissant une base dont les m premiers vecteurs sont ces vecteurs propres, que A est semblable à une matrice A' de la forme $A' = \begin{pmatrix} D & B \\ 0 & C \end{pmatrix}$, où D est diagonale d'ordre m, $B \in \mathcal{M}_{m,n-m}(\mathbb{R})$ et $C \in \mathcal{M}_{n-m}(\mathbb{R})$. On pose

$J_m = \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}$. On vérifie aisément que $A'J_m = J_m\,{}^tA' = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}$. Si P est une matrice de $GL_n(\mathbb{R})$ telle que $A' = PAP^{-1}$, on obtient $PAP^{-1}J_m = J_m\,{}^tP^{-1}\,{}^tA\,{}^tP$ et donc $AP^{-1}J_m\,{}^tP^{-1} = P^{-1}J_m\,{}^tP^{-1}\,{}^tA$. La matrice $S = P^{-1}J_m\,{}^tP^{-1}$ est symétrique et de rang m comme J_m. Comme J_m et S expriment la même forme quadratique dans des bases différentes, S est positive puisque J_m l'est manifestement. Les valeurs propres de S sont donc positives et S répond à la question.

• Supposons réciproquement qu'il existe une matrice S symétrique de rang m telle que $AS = S\,{}^tA$. Comme S est positive de rang m elle est congruente[2] à la matrice J_m : il existe $Q \in GL_n(\mathbb{R})$ tel que $J_m = {}^tQSQ$. Si l'on note $P = {}^tQ$, on a $S = P^{-1}J_m\,{}^tP^{-1}$. La relation $AS = S\,{}^tA$ devient $A'J_m = J_m\,{}^tA'$ avec $A' = PAP^{-1}$. Si l'on écrit $A' = \begin{pmatrix} D & B \\ E & C \end{pmatrix}$ avec D carrée de taille m, on obtient ${}^tD = D$ et $E = 0$. Autrement dit, $A' = \begin{pmatrix} D & B \\ 0 & C \end{pmatrix}$, avec D symétrique réelle. La matrice D est diagonalisable et la restriction de l'endomorphisme canoniquement associé à A' au sous-espace engendré par les m premiers vecteurs possède donc une base de vecteurs propres. Cette base est constituée de m vecteurs propres indépendants pour A'. Comme A est semblable à A', A possède également m vecteurs propres indépendants. ◁

Il s'agit dans les deux exercices qui suivent de résoudre une équation dans $\mathcal{L}(E)$ ou $\mathcal{M}_n(\mathbb{R})$ faisant intervenir des endomorphismes définis positifs ou des matrices définies positives.

2.16. Une équation matricielle

Soit B une matrice symétrique définie positive, A une matrice dont toutes les valeurs propres sont de module strictement inférieur à 1. Prouver qu'il existe une unique matrice symétrique définie positive K telle que $K - AK\,{}^tA = B$.

(École polytechnique)

Solution.

Notons $\mathcal{S}$ le sous-espace des matrices symétriques de $\mathcal{M}_n(\mathbb{R})$. Définissons l'application

$$\varphi : \begin{array}{ccc} \mathcal{S} & \longrightarrow & \mathcal{S} \\ K & \longmapsto & K - AK\,{}^tA. \end{array}$$

2. Le lecteur pourra se reporter aux rappels sur la classification des formes quadratiques page 212.

Cette application est bien définie car si K est symétrique,

$$^t(K - AK\,^tA) = {}^tK - {}^t({}^tA)\,{}^tK\,{}^tA = K - AK\,^tA$$

et $K - AK\,^tA$ est symétrique. Clairement, φ est linéaire.

Montrons que φ est injective. Soit $K \in \operatorname{Ker}\varphi$. On a $K = AK\,^tA$ et par une récurrence immédiate, $K = A^mK\,^tA^m$ pour tout $m \in \mathbb{N}$. Or $\lim\limits_{m\to+\infty} A^m = 0$, puisque le rayon spectral $\rho(A)$ de A est strictement inférieur à 1 (se reporter à l'exercice 2.59 du tome 2 d'algèbre). Donc $K = 0$ et φ est injective.

L'espace $\mathcal{S}$ étant de dimension finie, φ est surjective. Il existe donc $K \in \mathcal{S}$ tel que $\varphi(K) = B$ *i.e.* $K - AK\,^tA = B$. Il reste à prouver que K est définie positive. Soit $x \in \mathbb{R}^n$ non nul. On a, en remplaçant K par $AK\,^tA + B$ de manière itérée,

$$\begin{aligned}
\langle Kx, x\rangle &= \langle AK\,^tAx, x\rangle + \langle Bx, x\rangle = \langle K\,^tAx, {}^tAx\rangle + \langle Bx, x\rangle \\
&= \langle AK\,^tA\,^tAx, {}^tAx\rangle + \langle B\,^tAx, {}^tAx\rangle + \langle Bx, x\rangle \\
&= \langle K\,^tA^2x, {}^tA^2x\rangle + \langle B\,^tAx, {}^tAx\rangle + \langle Bx, x\rangle \\
&\;\;\vdots \\
&= \langle K\,^tA^mx, {}^tA^mx\rangle + \cdots + \langle B\,^tAx, {}^tAx\rangle + \langle Bx, x\rangle \\
&= \langle K\,^tA^mx, {}^tA^mx\rangle + \sum_{k=0}^{m-1} \langle B\,^tA^kx, {}^tA^kx\rangle.
\end{aligned}$$

Or $\lim\limits_{m\to+\infty} {}^tA^m = 0$, donc $\lim\limits_{m\to+\infty} \langle K\,^tA^mx, {}^tA^mx\rangle = 0$ (par continuité des applications linéaires et du produit scalaire en dimension finie) et finalement en faisant tendre m vers l'infini

$$(Kx|x) = \sum_{k=0}^{+\infty} \underbrace{\langle B\,^tA^kx, {}^tA^kx\rangle}_{\substack{\geqslant 0 \text{ car} \\ \text{B est définie positive}}} \geqslant \langle Bx, x\rangle > 0.$$

Donc K est définie positive. ◁

2.17. Équation dans $\mathcal{L}(E)$

Soit E un espace euclidien, a et b deux endomorphismes de E symétriques définis positifs.

1. Montrer qu'il existe un unique $c \in \mathcal{L}(E)$ tel que $b = ac + ca$. Montrer que c est symétrique défini positif.

2. Dans le cas où $\dim E = 2$, montrer que si a et c sont des endomorphismes de E symétriques définis positifs, $ac + ca$ n'est pas nécessairement défini positif.

(École polytechnique)

▷ **Solution.**

1. Considérons l'application :

$$\begin{array}{rccc} & \mathcal{L}(\mathrm{E}) & \longrightarrow & \mathcal{L}(\mathrm{E}) \\ \varphi : & c & \longmapsto & ac + ca. \end{array}$$

Il est aisé de vérifier que φ est linéaire. Montrons que φ est injective. Comme a est symétrique, considérons $(e_1, \ldots, e_n)$ une base orthonormée formée de vecteurs propres pour a. Pour $i \in [\![1, n]\!]$, on note $\lambda_i \in \mathbb{R}_+^*$ la valeur propre associée à e_i : $a(e_i) = \lambda_i e_i$. Prenons $c \in \operatorname{Ker}\varphi$. Soit $i \in \{1, 2, \ldots, n\}$, on a

$$0 = (ac + ca)(e_i) = a(c(e_i)) + c(a(e_i)) = a(c(e_i)) + c(\lambda_i e_i)$$

et donc $a(c(e_i)) = -\lambda_i c(e_i)$. Or $-\lambda_i < 0$ et λ_i ne peut donc être dans le spectre de a qui est contenu dans $\mathbb{R}_+^*$. Donc $c(e_i) = 0$. Par conséquent, $c = 0$ et φ est injective.

Comme $\mathcal{L}(\mathrm{E})$ est de dimension finie, φ est un isomorphisme. En particulier, il existe un unique $c \in \mathcal{L}(\mathrm{E})$ tel que $ac + ca = b$.

On a

$$b = b^* = (ac + ca)^* = c^* a^* + a^* c^* = c^* a + ac^*.$$

Ainsi, on a vérifié $\varphi(c) = \varphi(c^*)$ et par conséquent $c = c^*$, puisque φ est injectif. L'endomorphisme c est symétrique.

Soit $\lambda \in \operatorname{Sp} c$ et $x \neq 0$ dans E tel que $c(x) = \lambda x$. On a

$$\begin{aligned} 0 < \langle b(x), x\rangle &= \langle ac(x), x\rangle + \langle ca(x), x\rangle = \langle a(\lambda x), x\rangle + \langle a(x), c(x)\rangle \\ &= \lambda\langle a(x), x\rangle + \lambda\langle a(x), x\rangle = 2\lambda\langle a(x), x\rangle \end{aligned}$$

Par conséquent, comme $\langle a(x), x\rangle > 0$ puisque a est défini positif, λ est strictement positif. Le spectre de c est contenu dans $\mathbb{R}_+^*$ et c est symétrique défini positif.

2. Une matrice réelle $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ est symétrique définie positive si, et seulement si, $\alpha > 0$ et son déterminant $\alpha\gamma - \beta^2 > 0$. En effet, si $\alpha > 0$ et $\alpha\gamma - \beta^2 > 0$, le produit des valeurs propres de $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ est strictement positif et ces valeurs propres sont de même signe. Leur somme vaut $\alpha + \gamma$. Or α est du même signe que γ, car $\alpha\gamma > \beta^2$. Ainsi $\alpha + \gamma > 0$ et finalement les valeurs propres sont strictement positives. La matrice $\begin{pmatrix} \alpha & \beta \\ \beta & \gamma \end{pmatrix}$ est bien définie positive. La réciproque est immédiate (voir aussi l'exercice suivant).

Considérons les matrices

$$\mathrm{A} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}, \quad \mathrm{C} = \begin{pmatrix} 1 & \beta \\ \beta & \gamma \end{pmatrix}, \quad \text{et} \quad \mathrm{B} = \mathrm{AC} + \mathrm{CA}$$

avec $(\beta, \gamma) \in \mathbb{R}^2$. La matrice C est définie positive dès lors que $\gamma - \beta^2 > 0$. On a

$$B = \begin{pmatrix} 2 & 3\beta \\ 3\beta & 4\gamma \end{pmatrix}.$$

B n'est pas définie positive si $8\gamma - 9\beta^2 \leqslant 0$. Prenons $\beta = 1$ et $\gamma \in]1, 9/8[$, par exemple $\gamma = 17/16$. Alors ces inéquations sont vérifiées (B est en fait de signature $(1, 1)$).

Dans E, on prend une base orthonormée et les endomorphismes a et c associés à A et C respectivement dans cette base et les contre-exemples voulus sont construits. ◁

La caractérisation analytique des matrices définies positives de taille $(2, 2)$ donnée dans la seconde question de l'exercice précédent se généralise pour les matrices de taille quelconque. Elle est basée sur le signe des mineurs principaux.

2.18. Caractérisation de Sylvester des matrices définies positives

Soit $M = (m_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique réelle. Montrer que M est définie positive si, et seulement si, tous les mineurs principaux $D_k = \det(m_{ij})_{1 \leqslant i,j \leqslant k}$ (pour $1 \leqslant k \leqslant n$) sont strictement positifs.

(École polytechnique)

▷ **Solution.**

On note $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$. Supposons M définie positive et notons q_M la forme quadratique canoniquement associée à M : $q_M(X) = {}^tXMX$ pour $X \in \mathbb{R}^n$. Alors, $(m_{ij})_{1 \leqslant i,j \leqslant k}$ n'est autre que la matrice de la restriction de q_M à $\mathrm{Vect}(e_1, \ldots, e_k)$. Cette restriction étant évidemment définie positive, sa matrice a un déterminant D_k strictement positif.

Pour la réciproque, procédons par récurrence sur n. Le résultat est clair pour $n = 1$. On suppose que la propriété est vraie au rang $n - 1$ $(n \geqslant 2)$ et on considère une matrice M symétrique réelle d'ordre n dont les n déterminants extraits correspondants aux k premières lignes et aux k premières colonnes $(1 \leqslant k \leqslant n)$ sont positifs. On note N la matrice obtenue en supprimant la dernière ligne et la dernière colonne de M. Elle vérifie les mêmes hypothèses que M. Par hypothèse de récurrence, elle est définie positive. Il existe $P \in O_{n-1}(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_{n-1})$, à coefficients diagonaux strictement positifs telles que ${}^tPNP = D$. On pose $Q = \begin{pmatrix} P & 0 \\ 0 & 1 \end{pmatrix} \in GL_n(\mathbb{R})$. La matrice Q est orthogonale et il existe

$(a_1, \dots, a_n) \in \mathbb{R}^n$ tel que

$$ {}^t\mathrm{QMQ} = \begin{pmatrix} \lambda_1 & & & a_1 \\ & \ddots & & \vdots \\ & & \lambda_{n-1} & a_{n-1} \\ a_1 & \dots & a_{n-1} & a_n \end{pmatrix} = \mathrm{M}'. $$

Pour tout $\mathrm{X} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$, on a

$$ \begin{aligned} {}^t\mathrm{XM'X} &= \sum_{i=1}^{n-1} \lambda_i x_i^2 + a_n x_n^2 + 2 \sum_{i=1}^{n-1} a_i x_i x_n \\ &= \sum_{i=1}^{n-1} \lambda_i \left(x_i + \frac{a_i}{\lambda_i} x_n \right)^2 + \left(a_n - \sum_{i=1}^{n-1} \frac{a_i^2}{\lambda_i} \right) x_n^2. \end{aligned} $$

Il faut démontrer que le dernier coefficient est strictement positif. En ajoutant à la dernière ligne de M′ la i-ième multipliée par $-\frac{a_i}{\lambda_i}$, pour $1 \leqslant i \leqslant n-1$, on obtient

$$ \det \mathrm{M}' = \begin{vmatrix} \lambda_1 & & & a_1 \\ & \ddots & & \vdots \\ & & \lambda_{n-1} & a_{n-1} \\ 0 & \dots & 0 & a_n - \sum_{i=1}^{n-1} \frac{a_i^2}{\lambda_i} \end{vmatrix} = \lambda_1 \dots \lambda_{n-1} \left(a_n - \sum_{i=1}^{n-1} \frac{a_i^2}{\lambda_i} \right). $$

Par hypothèse le déterminant de M est strictement positif, donc $\det \mathrm{M}' = (\det \mathrm{Q})^2 \det \mathrm{M} = \det \mathrm{M} > 0$. Les λ_i étant positifs, on en déduit que $a_n - \sum_{i=1}^{n-1} \frac{a_i^2}{\lambda_i} > 0$. De l'expression de ${}^t\mathrm{XM'X}$, on déduit que M′ est définie positive. Il en est de même de M qui lui est congruente. ◁

Une autre solution plus simple, cependant moins dans l'esprit du programme des classes préparatoires, utilisant l'orthogonalité au sens d'une forme quadratique sera présentée dans l'exercice 3.15.

La caractérisation établie dans l'exercice précédent montre que l'ensemble des matrices symétriques définies positives forme un ouvert de l'espace des matrices symétriques réelles de taille n.

Notons que l'on ne peut caractériser les matrices symétriques positives par la positivité des mineurs principaux comme le montre la matrice $\begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix}$. *En prenant tous les mineurs diagonaux c'est toutefois possible comme le montre l'exercice suivant.*

2.19. Caractérisation des matrices positives

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique réelle. On note A_I la sous-matrice $(a_{ij})_{(i,j) \in I^2}$ pour toute partie non vide $I \subset [\![1, n]\!]$. Montrer que A est positive si et seulement si $\det A_I \geqslant 0$ pour toute partie non vide I.

(École polytechnique)

▷ **Solution.**

On notera $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$ et $q : X \longmapsto {}^tXAX$ la forme quadratique canoniquement associée à A.

• Supposons A est positive et soit $I \subset [\![1, n]\!]$ non vide. La restriction de la forme quadratique q au sous-espace $\mathrm{Vect}(e_i)_{i \in I}$ est positive et sa matrice dans la base $(e_i)_{i \in I}$ est A_I. On a donc $\det A_I \geqslant 0$.

• Supposons réciproquement que $\det A_I \geqslant 0$ pour toute partie non vide I. On va montrer que $A + \varepsilon I_n$ est définie positive pour tout $\varepsilon > 0$. Cela suffit pour conclure, puisque si $X \in \mathbb{R}^n$ est fixé et non nul, alors $q(X) + \varepsilon \|X\|^2 > 0$ pour tout $\varepsilon > 0$, et donc $q(X) \geqslant 0$ en faisant tendre ε vers 0^+. Pour prouver que $A + \varepsilon I_n$ est définie positive on va utiliser la caractérisation de Sylvester démontrée dans l'exercice précédent et montrer que les mineurs principaux de cette matrice sont strictement positifs. Comme les matrices extraites $(a_{ij})_{1 \leqslant i,j \leqslant k}$ pour $1 \leqslant k \leqslant n$ vérifient la même propriété que A il suffit donc de prouver que $\det(A + \varepsilon I_n) > 0$ pour tout $\varepsilon > 0$. Il s'agit évidemment d'une fonction polynôme en ε de degré n (à savoir $\chi_A(-\varepsilon)$) qu'on peut écrire

$$\det(A + \varepsilon I_n) = a_0 + a_1 \varepsilon + a_2 \varepsilon^2 + \cdots + a_{n-1} \varepsilon^{n-1} + a_n \varepsilon^n$$

avec $a_0 = \det A \geqslant 0$ et $a_n = 1$. Le résultat provient de ce que tous les coefficients a_k sont positifs ou nuls. En effet, on a pour tout $k \in [\![1, n]\!]$,

$$a_{n-k} = \sum_{|I| = k} \det A_I.$$

Une manière agréable de prouver cela est de considérer plutôt le déterminant de la matrice $A + \mathrm{Diag}(\varepsilon_1, \ldots, \varepsilon_n)$ qui conduit à un polynôme en $\varepsilon_1, \ldots, \varepsilon_n$. Soit alors $i_1 < i_2 < \cdots i_k$ des indices de $[\![1, n]\!]$. En développant par multilinéarité, il est facile de voir que le coefficient du terme $\varepsilon_{i_1} \varepsilon_{i_2} \ldots \varepsilon_{i_k}$ est exactement le déterminant de la matrice A_I avec $I = [\![1, n]\!] \setminus \{i_1, \ldots, i_k\}$. La formule annoncée plus haut s'en déduit en prenant $\varepsilon_1 = \cdots = \varepsilon_n = \varepsilon$. ◁

Le thème suivant est l'étude des matrices symétriques réelles à coefficients positifs. Le premier énoncé étudie en dimension 2 un problème difficile qui sera abordé plus loin.

2.20. Matrice symétrique de taille 2 à coefficients positifs

Soit $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ et $D = \mathrm{Diag}(\lambda_1, \lambda_2)$. À quelle condition nécessaire et suffisante existe-t-il une matrice symétrique à coefficients réels strictement positifs orthogonalement semblable à D ?

(École polytechnique)

▷ **Solution.**

Supposons qu'il existe $A \in \mathcal{M}_2(\mathbb{R})$ symétrique à coefficients strictement positifs orthogonalement semblable à D. Il existe donc $P \in O_2(\mathbb{R})$ telle que $A = PD\,{}^tP$. Quitte à remplacer P par $P\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ on peut supposer que $P \in SO_2(\mathbb{R})$: dit plus abstraitement, si A se diagonalise dans une base orthonormée, elle se diagonalise aussi dans une base orthonormée directe (il suffit de remplacer l'un des vecteurs propres par son opposé). Posons alors $P = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix}$. On a alors, après deux lignes de calcul,

$$A = PD\,{}^tP = \begin{pmatrix} \lambda_1 \cos^2 t + \lambda_2 \sin^2 t & (\lambda_1 - \lambda_2) \sin t \cos t \\ (\lambda_1 - \lambda_2) \sin t \cos t & \lambda_1 \sin^2 t + \lambda_2 \cos^2 t \end{pmatrix}$$

ce qui donne encore en utilisant les formules de duplication

$$A = \frac{1}{2}\begin{pmatrix} \lambda_1 + \lambda_2 + (\lambda_1 - \lambda_2)\cos 2t & (\lambda_1 - \lambda_2)\sin 2t \\ (\lambda_1 - \lambda_2)\sin 2t & \lambda_1 + \lambda_2 - (\lambda_1 - \lambda_2)\cos 2t \end{pmatrix}.$$

On cherche à avoir tous les coefficients de A strictement positifs. Une condition nécessaire est que $\mathrm{Tr}\, A = \lambda_1 + \lambda_2$ soit strictement positif. Il est aussi nécessaire que $\lambda_1 \neq \lambda_2$ en raison des coefficients non diagonaux. Ces deux conditions sont en fait suffisantes : en effet, il suffit alors de choisir t de sorte que $\cos 2t$ soit assez petit pour que les coefficients diagonaux soit strictement positifs et que $\sin 2t$ ait le signe de $\lambda_1 - \lambda_2$.

Conclusion. Il existe une matrice symétrique à coefficients réels strictement positifs semblable à D si, et seulement si, $\lambda_1 + \lambda_2 > 0$ et $\lambda_1 \neq \lambda_2$.

Si on souhaite seulement avoir des coefficients positifs ou nuls la condition est $\lambda_1 + \lambda_2 \geqslant 0$. C'est ce résultat qui sera généralisé dans l'exercice 2.22.

L'énoncé suivant concerne le théorème de Perron-Frobenius dans le cadre de matrices symétriques. La preuve est plus simple que dans le cas général que le lecteur trouvera dans le tome algèbre 2 (exercice 2.10).

2.21. Théorème de Perron-Frobenius pour les matrices symétriques

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique réelle à coefficients strictement positifs. On note $\rho = \max_{\lambda \in Sp_{\mathbb{C}} A} |\lambda|$ le rayon spectral de A. Montrer que ρ est valeur propre simple de A et que le sous-espace propre associé à ρ est engendré par une matrice colonne à coefficients strictement positifs.

(École normale supérieure)

▷ **Solution.**

Notons $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ les valeurs propres de A. Nous savons que si X est un vecteur unitaire alors ${}^tXAX \leqslant \lambda_1$. En effet, il suffit de considérer une base orthonormale $(e_1, \ldots, e_n)$ de vecteurs propres de A, e_i étant associé à λ_i, et de décomposer X dans cette base : si $X = x_1e_1 + \cdots + x_ne_n$ on a

$$ {}^tXAX = \sum_{i=1}^{n} \lambda_i x_i^2 \leqslant \lambda_1 \sum_{i=1}^{n} x_i^2 = \lambda_1. $$

On peut aussi noter qu'il y a égalité si et seulement si X est un vecteur propre associé à la valeur propre λ_1 (toutes les composantes x_i pour les $\lambda_i < \lambda_1$ doivent être nulles).

Pour $X \in \mathbb{R}^n$ on notera $|X|$ le vecteur dont les coordonnées sont les valeurs absolues de celles de X. Pour tout X unitaire, l'inégalité triangulaire permet d'écrire

$$ |{}^tXAX| \leqslant {}^t|X|A|X| \leqslant \lambda_1 $$

car A est à coefficients positifs. Si on prend pour X un vecteur propre unitaire associé à une valeur propre λ quelconque de A on en déduit que $|\lambda| \leqslant \lambda_1$. Autrement dit, $\lambda_1 = \rho$ et le rayon spectral de A est bien une valeur propre (en particulier λ_1 est positif, et même strictement, puisque A n'est pas nulle).

L'inégalité ci-dessus montre de plus que si X est un vecteur propre unitaire associé à λ_1 alors $|\lambda_1| = {}^t|X|A|X| \leqslant \lambda_1$. On a donc ${}^t|X|A|X| = \lambda_1$ et $|X|$ est aussi un vecteur propre associé à λ_1. Montrons que toutes les composantes de $|X|$ sont strictement positives. En effet, si $x_i = 0$ pour un entier i, la i-ième coordonnée de la relation $A|X| = \lambda_1|X|$ conduit à $\sum_{j \neq i} a_{ij}|x_j| = 0$ et comme les coefficients de A sont strictement positifs cela implique $x_j = 0$ pour tout $j \neq i$, donc $X = 0$ ce qui est absurde.

On a donc montré que A admet un vecteur propre pour λ_1 dont les coordonnées sont strictement positives. Pour conclure, il ne reste plus qu'à prouver que l'espace propre associé à λ_1 est de dimension 1. Comme

A est diagonalisable cela revient à dire que λ_1 est une valeur propre simple. Si $X = (x_1, \ldots, x_n)$ et $X' = (x'_1, \ldots, x'_n)$ sont deux vecteurs propres de A pour la valeur propre λ_1, les composantes de X et X' sont, d'après ce qui précède, toutes non nulles. Posons $t = \dfrac{x'_1}{x_1}$. Le vecteur $X' - tX$ est dans l'espace propre pour la valeur propre λ_1 et sa première composante est nulle, donc il est nul. On a donc $X' = tX$ et le résultat demandé. ◁

On s'intéresse désormais à une sorte de réciproque. Étant donnés n réels, à quelles conditions est-il possible de trouver une matrice symétrique à coefficients positifs admettant ces réels comme valeurs propres ? On notera que si dans l'exercice précédent la matrice A est seulement à coefficients positifs ou nuls on a toujours $\rho(A) \in \operatorname{Sp} A$ et l'existence d'un vecteur propre à coordonnées positives ou nulles : en revanche l'espace propre associé n'est plus nécessairement de dimension 1. *Cela fournit déjà une condition nécessaire sur le n-uplet de réels.*

2.22. Théorème de Suleimanova (1949)

1. Soit $A \in \mathcal{M}_p(\mathbb{R})$ et $B \in \mathcal{M}_q(\mathbb{R})$ deux matrices symétriques. Soit α (resp. β) une valeur propre de A (resp. B) et X (resp. Y) un vecteur propre associé et $s \in \mathbb{R}$. Déterminer les valeurs propres de la matrice

$$C = \begin{pmatrix} A & sX\,{}^tY \\ sY\,{}^tX & B \end{pmatrix}.$$

2. Montrer que si M est symétrique à coefficients positifs de plus grande valeur propre $\lambda \geqslant 0$, alors M admet un vecteur propre associé à λ à coefficients positifs.

3. En déduire que si $\lambda_1 \geqslant 0 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ vérifient $\lambda_1 + \cdots + \lambda_n \geqslant 0$, il existe M symétrique de taille n à coefficients positifs de valeurs propres $\lambda_1, \ldots, \lambda_n$ (comptées avec multiplicité).

(École polytechnique)

Solution.

1. La matrice C est symétrique réelle, donc elle est diagonalisable dans une base orthonormée, avec des valeurs propres réelles. On cherche les vecteurs propres de C sous la forme $\begin{pmatrix} Z \\ T \end{pmatrix}$, avec $Z \in \mathcal{M}_{p,1}(\mathbb{R})$ et $T \in \mathcal{M}_{q,1}(\mathbb{R})$. Si Z est un vecteur propre de A associé à la valeur propre λ orthogonal à X, on a

$$C\begin{pmatrix} Z \\ 0 \end{pmatrix} = \begin{pmatrix} AZ \\ sY\,{}^tXZs \end{pmatrix} = \begin{pmatrix} \lambda Z \\ 0 \end{pmatrix} = \lambda \begin{pmatrix} Z \\ 0 \end{pmatrix}.$$

Cela montre que $\lambda \in \operatorname{Sp} C$. Ainsi toute valeur propre de A différente de α est valeur propre de C et la dimension de l'espace propre associé est supérieure ou égale à la dimension de l'espace propre correspondant de A. De plus, α est valeur propre de C lorsque son ordre de multiplicité pour A est supérieur ou égal à 2. En considérant les vecteurs $\begin{pmatrix} 0 \\ T \end{pmatrix}$, où T est un vecteur propre de B orthogonal à Y, on obtient le même résultat pour les valeurs propres de B. On peut ainsi obtenir $p + q - 2$ vecteurs propres linéairement indépendants deux à deux orthogonaux. Notons que si $s = 0$, les vecteurs $\begin{pmatrix} X \\ 0 \end{pmatrix}$ et $\begin{pmatrix} 0 \\ Y \end{pmatrix}$ sont vecteurs propres de C. Dans ce cas, que nous écartons désormais, on a $\operatorname{Sp} C = \operatorname{Sp} A \cup \operatorname{Sp} B$. Cherchons enfin des vecteurs propres de la forme $U = \begin{pmatrix} aX \\ Y \end{pmatrix}$, où a est réel. Un tel vecteur est orthogonal aux vecteurs propres précédemment déterminés. On obtient

$$CU = \begin{pmatrix} aAX + sX\,{}^tYY \\ saY\,{}^tXX + BY \end{pmatrix} = \begin{pmatrix} (a\alpha + s\|Y\|^2)X \\ (sa\|X\|^2 + \beta)Y \end{pmatrix}.$$

Le vecteur U est vecteur propre de C si et seulement si on a $a\alpha + s\|Y\|^2 = a(sa\|X\|^2 + \beta)$. Cette équation du second degré en a qui s'écrit $s\|X\|^2a^2 + (\beta - \alpha)a - s\|Y\|^2 = 0$ a deux solutions

$$a = \frac{1}{2s\|X\|^2}\left(\alpha - \beta \pm \sqrt{(\beta - \alpha)^2 + 4s^2\|X\|^2\|Y\|^2}\right),$$

la valeur propre correspondante étant $sa\|X\|^2 + \beta$. On obtient donc les deux dernières valeurs propres

$$\lambda = \frac{1}{2}\left(\alpha + \beta \pm \sqrt{(\beta - \alpha)^2 + 4s^2\|X\|^2\|Y\|^2}\right).$$

Si $s = 0$, on retrouve α et β, comme prévu.

Conclusion. Si $\alpha_1, \ldots, \alpha_p$ sont les valeurs propres de A et $\beta_1, \ldots, \beta_q$ celles de B, comptées avec multiplicité, avec $\alpha = \alpha_1$ et $\beta = \beta_1$, alors les valeurs propres de C sont $\frac{1}{2}(\alpha + \beta - \sqrt{(\beta - \alpha)^2 + 4s^2\|X\|^2\|Y\|^2})$, $\frac{1}{2}(\alpha + \beta + \sqrt{(\beta - \alpha)^2 + 4s^2\|X\|^2\|Y\|^2})$, $\alpha_2, \ldots, \alpha_p$, $\beta_2, \ldots, \beta_q$.

2. C'est ce que nous avons rappelé en préambule : voir la solution de l'exercice précédent.

3. On raisonne par récurrence sur n. Le cas $n = 1$ est évident ; on prend $M = (\lambda_1)$. Supposons la propriété vérifiée au rang $n - 1$ ($n \geqslant 2$)

et considérons des réels $\lambda_1, \dots, \lambda_n$ tels que $\lambda_1 \geqslant 0 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ et $\lambda_1 + \cdots + \lambda_n \geqslant 0$. On a alors $\lambda_1 + \lambda_n \geqslant -\lambda_2 - \cdots - \lambda_{n-1} \geqslant 0$. En appliquant l'hypothèse de récurrence, on obtient une matrice N de $\mathcal{M}_{n-1}(\mathbb{R})$ symétrique à coefficients positifs, de valeurs propres $\lambda_1 + \lambda_n$, $\lambda_2, \dots, \lambda_{n-1}$. D'après la question 2, N possède pour la valeur propre $\lambda_1 + \lambda_n$ un vecteur propre X à coefficients positifs que nous pouvons supposer unitaire. Considérons la matrice $\mathrm{M} = \begin{pmatrix} \mathrm{N} & s\mathrm{X} \\ s\,{}^t\mathrm{X} & 0 \end{pmatrix} \in \mathcal{M}_n(\mathbb{R})$, où $s \in \mathbb{R}_+$. C'est une matrice symétrique à coefficients positifs. On applique le résultat de la question 1, en prenant $p = n-1$, $q = 1$, $\mathrm{A} = \mathrm{N}$, $\mathrm{B} = 0$, $\mathrm{Y} = 1$, $\alpha = \lambda_1 + \lambda_n$, $\beta = 0$. Les valeurs propres de M sont $\frac{1}{2}(\alpha + \sqrt{\alpha^2 + 4s^2})$, $\frac{1}{2}(\alpha - \sqrt{\alpha^2 + 4s^2})$, $\lambda_2, \dots, \lambda_{n-1}$. On choisit s pour que les deux autres valeurs propres soient λ_1 et λ_n. Leur somme est $\alpha = \lambda_1 + \lambda_n$ et leur produit $-s^2$. Il suffit de choisir s tel que $s^2 = -\lambda_1\lambda_n$. C'est possible, car $\lambda_1\lambda_n \leqslant 0$. ◁

Le problème général de savoir si pour des réels $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ donnés il existe A *symétrique à coefficients positifs de valeurs propres $\lambda_1, \dots, \lambda_n$ est très ardu et à notre connaissance non résolu. Il est évidemment nécessaire que $\lambda_1 + \cdots + \lambda_n = \operatorname{Tr} \mathrm{A} \geqslant 0$ mais aussi, d'après le théorème de Perron-Frobenius, que $\lambda_1 = \rho(\mathrm{A})$ autrement dit que $\lambda_1 \geqslant |\lambda_n|$ (cette seconde condition est automatiquement vérifiée dans la question 3 de l'exercice ci-dessus). Ces seules conditions ne sont toutefois pas suffisantes : on doit aussi avoir $\operatorname{Tr} \mathrm{A}^p = \sum_{i=1}^{n} \lambda_i^p \geqslant 0$ pour tout $p \in \mathbb{N}$. Ainsi le 5-uplet $(2, 1, 1, -2, -2)$ qui vérifie les deux premières conditions ne saurait convenir. Mais cette hypothèse supplémentaire se révèle aussi insuffisante...*

Lorsque qu'on considère une base de vecteurs unitaires formant une famille obtusangle (voir l'exercice 1.3), il s'avère que la matrice de passage de cette base à son orthonormalisé au sens de Gram-Schmidt est à coefficients positifs. C'est ce que l'on propose de démontrer dans l'exercice suivant.

2.23. Matrice de passage à coefficients positifs

1. Soit a un endomorphisme auto-adjoint d'un espace euclidien E. On suppose que les valeurs propres de $\mathrm{Id}_\mathrm{E} - a$ sont strictement positives et que $\operatorname{Tr} a^p \geqslant 0$ pour tout $p \in \mathbb{N}$. Montrer que les valeurs propres de a sont dans $]-1, 1[$.

2. Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ symétrique dont les coefficients sont positifs et telle que $\mathrm{I}_n - \mathrm{A}$ est définie positive. Montrer que les coefficients de $(\mathrm{I}_n - \mathrm{A})^{-1}$ sont positifs.

3. Soit $(f_1, \dots, f_n)$ une base de vecteurs unitaires de E telle que, pour tout $i \neq j$, $\langle f_i, f_j \rangle < 0$. Soit $(e_1, \dots, e_n)$ son orthonormalisé de Gram-Schmidt et M la matrice de passage de $(f_1, \dots, f_n)$ à $(e_1, \dots, e_n)$. Montrer que les coefficients de M sont positifs ou nuls.

(École polytechnique)

▷ **Solution.**

1. Le spectre de $\mathrm{Id}_E - a$ est $\{1 - \lambda,\ \lambda \in \mathrm{Sp}\, a\}$. Les valeurs propres sont réelles, car a est auto-adjoint. Puisqu'elles sont strictement positives, on a $\lambda < 1$, pour tout $\lambda \in \mathrm{Sp}\, a$. Notons $\lambda_1 < \lambda_2 < \cdots < \lambda_k$ les valeurs propres distinctes de a et n_i l'ordre de multiplicité de λ_i. Raisonnons par l'absurde et supposons que $\lambda_1 \leqslant -1$. On a, pour tout $p \in \mathbb{N}$,

$$\mathrm{Tr}\, a^p = \sum_{i=1}^{k} n_i \lambda_i^p = \lambda_1^p \left(n_1 + \sum_{i=2}^{p} n_i \left(\frac{\lambda_i}{\lambda_1} \right)^p \right).$$

Si $\lambda_i \geqslant 0$, on a $\left| \dfrac{\lambda_i}{\lambda_1} \right| \leqslant \lambda_i < 1$; cela reste vrai si $\lambda_i < 0$, car $0 > \lambda_i > \lambda_1$. Ainsi, pour tout $i \geqslant 2$, $\lim\limits_{p \to +\infty} \left(\dfrac{\lambda_i}{\lambda_1} \right)^p = 0$ et $\mathrm{Tr}\, a^p \underset{p \to +\infty}{\sim} n_1 \lambda_1^p$. Pour p assez grand et impair, on a $\mathrm{Tr}\, a^p < 0$, ce qui est contraire à l'hypothèse. On conclut que $\lambda_1 > -1$ et que toutes les valeurs propres de a sont dans l'intervalle $]-1, 1[$.

2. Soit a l'endomorphisme de $\mathbb{R}^n$ dont la matrice dans la base canonique de $\mathbb{R}^n$ est A. Il est auto-adjoint car A est symétrique. Les valeurs propres de $\mathrm{Id}_{\mathbb{R}^n} - a$ sont strictement positives, car cet endomorphisme est défini positif. Enfin, pour tout $p \in \mathbb{N}$, $\mathrm{Tr}\, a^p \geqslant 0$, car les coefficients de A, donc de A^p, sont positifs. D'après la question 1, les valeurs propres de a, et donc de A, sont dans $]-1, 1[$.

La matrice $I_n - A$ est donc inversible. Déterminons son inverse. La matrice A est diagonalisable : il existe $P \in O_n(\mathbb{R})$ et D diagonale telles que $A = {}^tPDP$. Pour tout $k \in \mathbb{N}$, on a $A^k = {}^tPD^kP$. Puisque les valeurs propres de A sont dans $]-1, 1[$, on a $\lim\limits_{k \to +\infty} D^k = 0$ et donc $\lim\limits_{k \to +\infty} A^k = 0$.

On peut écrire, pour tout $p \in \mathbb{N}$, $(I_n - A)\left(\sum\limits_{k=0}^{p-1} A^k \right) = I - A^p$ et donc

$$\sum_{k=0}^{p-1} A^k = (I_n - A)^{-1}(I_n - A^p).$$

En faisant p tendre vers $+\infty$, on en déduit que $(I_n - A)^{-1} = \sum\limits_{k=0}^{+\infty} A^k$. Comme pour tout $k \in \mathbb{N}$, les coefficients de A^k sont positifs, ceux de $(A - I_n)^{-1}$ le sont aussi.

3. Soit $(f_1, \ldots, f_n)$ une base de vecteurs unitaires de E telle que, pour tout $i \neq j$, $\langle f_i, f_j \rangle < 0$. Notons $G = (\langle f_i, f_j \rangle)_{1 \leqslant i,j \leqslant n}$ la matrice de Gram de cette base. La matrice $A = I_n - G$ est symétrique et à coefficients positifs. De plus, $G = I_n - A$ est définie positive (voir exercice 1.26). On peut donc appliquer la question précédente et dire que G^{-1} est à coefficients positifs. Il s'agit maintenant de relier G^{-1} à la matrice de passage M entre la base $(f_1, \ldots, f_n)$ et son orthonormalisée de Gram-Schmidt $(e_1, \ldots, e_n)$. Comme $(e_1, \ldots, e_n)$ est une base orthonormée et que M^{-1} contient les coordonnées des vecteurs $(f_1, \ldots, f_n)$ dans cette base, on a $G = {}^tM^{-1}M^{-1}$. En passant à l'inverse on a donc $G^{-1} = M\,{}^tM$. Or, $M = (m_{ij})_{1 \leqslant i,j \leqslant n}$ est une matrice triangulaire supérieure à coefficients diagonaux strictement positifs. La dernière colonne du produit $M\,{}^tM$ contient donc la dernière colonne de M multipliée par m_{nn}. On en déduit que la dernière colonne de M contient des coefficients positifs. Pour conclure, il suffit de faire une récurrence sur le nombre n de vecteurs. ◁

Nous abordons maintenant le thème très riche des décompositions matricielles. Celle présentée dans l'exercice suivant, facilement équivalente à la décomposition polaire qui sera vue plus loin, est utilisée en analyse numérique pour la résolution des systèmes linéaires et l'inversion des matrices. Elle revient à démontrer que toute matrice A de $\mathcal{M}_n(\mathbb{R})$ est orthogonalement équivalente à une matrice diagonale à coefficients diagonaux positifs. Ces derniers sont appelés valeurs singulières de A : ce sont les racines carrées des valeurs propres de la matrice symétrique positive ${}^tAA$.

2.24. Décomposition en valeurs singulières

Soit $A \in \mathcal{M}_n(\mathbb{R})$.

1. Montrer que ${}^tAA$ est symétrique positive.

2. Montrer qu'il existe deux matrices orthogonales U et V et $\lambda_1, \ldots, \lambda_n$ des réels positifs tels que

$$A = U\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)V.$$

3. En déduire que A est somme d'au plus n matrices de rang 1.

(**École polytechnique**)

Solution.

1. Nous avons déjà rencontré plusieurs fois ce fait classique : la matrice ${}^tAA$ est symétrique et pour montrer qu'elle est positive, on constate que la forme quadratique q associée est positive. Si $X \in \mathbb{R}^n$, on a

$$q(X) = {}^tX({}^tAA)X = \|AX\|^2 \geqslant 0,$$

$\mathbb{R}^n$ étant muni de son produit scalaire canonique. Ainsi ${}^tAA$ est positive.

2. *Analyse.* S'il existe deux matrices orthogonales U et V telles que $A = U\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)V$, alors

$$\begin{aligned} {}^tAA &= {}^tV\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)\,{}^tUU\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)V \\ &= {}^tV\,\mathrm{Diag}(\lambda_1^2, \ldots, \lambda_n^2)V. \end{aligned}$$

donc $\lambda_1^2, \ldots, \lambda_n^2$ sont nécessairement les valeurs propres de ${}^tAA$ ce qui détermine les réels positifs $\lambda_1, \ldots, \lambda_n$ de manière unique.

Synthèse. D'après la question 1 la matrice ${}^tAA$ est symétrique positive donc diagonalisable dans une base orthonormale de $\mathbb{R}^n$. Si on note $\lambda_1^2, \ldots, \lambda_n^2$ ses valeurs propres avec $\lambda_1, \ldots, \lambda_n$ positifs, il existe $V \in O_n(\mathbb{R})$ telle que

$${}^tAA = {}^tV\,\mathrm{Diag}(\lambda_1^2, \ldots, \lambda_n^2)V.$$

Posons alors $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. On a donc ${}^tAA = {}^t(DV)DV$; cela implique l'existence d'une matrice $U \in O_n(\mathbb{R})$ telle que $A = UDV$ comme cela a été montré dans l'exercice 1.28 de manière géométrique (les colonnes des deux matrices forment des familles isométriques). Donnons tout de même une rédaction matricielle de ce résultat important.

On cherche $U \in O_n(\mathbb{R})$ telle que $A\,{}^tV = UD$. Posons $A\,{}^tV = B$ de coefficients $(b_{ij})_{1 \leqslant i,j \leqslant n}$. L'égalité ${}^tBB = \mathrm{Diag}(\lambda_1^2, \ldots, \lambda_n^2)$ se traduit, pour tout $(i,j) \in [\![1,n]\!]^2$, par

$$\sum_{k=1}^{n} b_{ki}b_{kj} = \lambda_i^2\delta_{ij}.$$

En appelant $C_1, \ldots, C_n$ les vecteurs colonnes de B, on a donc, pour tout $(i,j) \in [\![1,n]\!]^2$,

$$\langle C_i, C_j \rangle = \lambda_i^2\delta_{ij}.$$

Autrement dit, la famille $(C_1, \ldots, C_n)$ est orthogonale et pour tout i, $\|C_i\|^2 = \lambda_i^2$. Soit $J = \{i \in [\![1,n]\!],\ \lambda_i \neq 0\}$. Posons $u_i = \dfrac{1}{\lambda_i}C_i$ pour $i \in J$. La famille $(u_i)_{i \in J}$ est orthonormale. On la complète en une base orthonormale $(u_i)_{i \in [\![1,n]\!]}$. La matrice U dont les vecteurs colonnes sont $u_1, \ldots, u_n$ est orthogonale et on a, pour tout $i \in [\![1,n]\!]$, $C_i = \lambda_i u_i$. En effet, si $i \in J$, c'est vrai par construction et si $i \notin J$, cela résulte de ce que $\lambda_i = 0$ et $C_i = 0$.

Matriciellement, cela se traduit par l'égalité $B = U\,\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ souhaitée.

3. La matrice A s'écrit

$$\begin{aligned} A &= \sum_{i=1}^{n} U\,\mathrm{Diag}(0,\dots,0,\lambda_i,0,\dots,0,\dots)V \\ &= \sum_{i\in J} U\,\mathrm{Diag}(0,\dots,0,\lambda_i,0,\dots,0,\dots)V, \end{aligned}$$

avec $\mathrm{rg}(U\,\mathrm{Diag}(0,\dots,0,\lambda_i,0,\dots,0,\dots)V) = 1$, si $i \in J$. La matrice A est somme d'au plus n matrices de rang 1. ◁

Notons que le résultat de la dernière question s'obtient de manière tout à fait élémentaire et reste vrai avec une matrice A *dans* $\mathcal{M}_n(K)$ *pour un corps* K *quelconque : il suffit de considérer les matrices obtenues en gardant une colonne non nulle de* A *et dont les autres colonnes sont nulles. Il y en a au plus* n*, elles sont toutes de rang* 1 *et leur somme fait bien* A.

Par ailleurs, la décomposition en valeurs singulières de la deuxième question permet de donner une expression du pseudo-inverse de A *(voir exercice 1.12). En effet, si on note* $D = \mathrm{Diag}(\lambda_1,\dots,\lambda_n)$ *et* $\widetilde{D} = \mathrm{Diag}(\widetilde{\lambda_1},\dots,\widetilde{\lambda_n})$ *où* $\widetilde{\lambda_i} = \dfrac{1}{\lambda_i}$ *si* λ_i *est non nul et* 0 *sinon, alors* $B = V\widetilde{D}U$ *est le pseudo-inverse de* A.

Nous verrons plus loin que la décomposition en valeurs singulières équivaut très facilement à la décomposition polaire. L'une ou l'autre de ces décompositions peut donc être utilisée pour résoudre l'exercice suivant.

2.25. Dilatation isométrique d'une contraction

Soit E un espace euclidien de dimension $2n$, F un sous-espace de E de dimension n, $u \in \mathcal{L}(F)$ et p la projection orthogonale de E sur F. Montrer l'équivalence entre

(*i*) $\forall x \in F,\ \|u(x)\| \leqslant \|x\|$;

(*ii*) $\exists v \in O(E),\ u = p \circ v_{|F}$.

(**École normale supérieure**)

Solution.

• Une des implications est évidente. En effet, s'il existe $v \in O(E)$ tel que $u = p \circ v_{|F}$, on a, pour tout $x \in F$,

$$\|u(x)\| - \|p \circ v(x)\| \leqslant \|v(x)\| - \|x\|,$$

car une projection orthogonale est de norme triple inférieure à 1.

• Supposons réciproquement que (*i*) est réalisé. Soit $A \in \mathcal{M}_n(\mathbb{R})$ la matrice de u dans une base orthonormale $\mathcal{B}_0$ quelconque de F. On complète $\mathcal{B}_0$ en une base orthonormale $\mathcal{B}$ de E. Un endomorphisme v de E vérifie (*ii*) si sa matrice M dans la base $\mathcal{B}$ est orthogonale et s'écrit par

blocs $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, où B, C et D appartiennent à $\mathcal{M}_n(\mathbb{R})$. Autrement dit, A étant une matrice donnée (de norme triple inférieure à 1), on cherche à la compléter par des matrices B, C, D de manière à ce que M soit orthogonale. En calculant ${}^tMM$, on voit que ceci est réalisé si, et seulement si,

$$\begin{cases} {}^tAA + {}^tCC = I_n \\ {}^tBA + {}^tDC = I_n \\ {}^tBB + {}^tDD = 0_n. \end{cases}$$

La matrice ${}^tAA$ est symétrique positive, car pour tout $X \in \mathcal{M}_{n,1}(\mathbb{R})$, on a ${}^tX\,{}^tAAX = \|AX\|^2 \geqslant 0$. L'hypothèse se traduit par l'inégalité ${}^tX\,{}^tAAX = \|AX\|^2 \leqslant \|X\|^2$. En prenant pour X un vecteur propre de ${}^tAA$ de norme 1, on en déduit que les valeurs propres de ${}^tAA$ sont inférieures à 1 (et positives). On les note $\lambda_1^2, \ldots, \lambda_n^2$, où les λ_i sont dans $[0,1]$. D'après l'exercice précédent, il existe des matrices orthogonales U et V telles que $A = U\Delta_A V$, où $\Delta_A = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$.

On va chercher B, C et D sous la forme $U\Delta_B V$, $U\Delta_C V$, et $U\Delta_D V$ respectivement, avec Δ_B, Δ_C et Δ_D diagonales. On a alors

$$\begin{cases} {}^tAA + {}^tCC = {}^tV(\Delta_A^2 + \Delta_C^2)V, \\ {}^tBA + {}^tDC = {}^tV(\Delta_B\Delta_A + \Delta_D\Delta_C)V, \\ {}^tBB + {}^tDD = {}^tV(\Delta_B^2 + \Delta_D^2)V. \end{cases}$$

La matrice M a les propriétés voulues si, et seulement si,

$$\begin{cases} \Delta_A^2 + \Delta_C^2 = I_n \\ \Delta_B\Delta_A + \Delta_D\Delta_C = 0 \\ \Delta_B^2 + \Delta_D^2 = I_n \end{cases}$$

La première condition est réalisée si $\Delta_C = \mathrm{Diag}(\sqrt{1-\lambda_1^2}, \ldots, \sqrt{1-\lambda_n^2})$, ce qui est possible, car les λ_i appartiennent à $[0,1]$. On constate ensuite qu'en prenant $\Delta_B = \Delta_C$ et $\Delta_D = -\Delta_A$, on obtient une matrice M et donc un endomorphisme v de E ayant les propriétés voulues. ◁

Nous en venons maintenant à la décomposition polaire.

2.26. Décomposition polaire (1)

Soit E un espace euclidien de dimension n.

1. Soit u un endomorphisme quelconque de E. Montrer qu'il existe un automorphisme orthogonal v et un endomorphisme auto-adjoint positif h tel que $u = vh$. Montrer que h est unique. Montrer que si u n'est pas inversible, v n'est pas unique.

2. Montrer que $u^*u = uu^* \iff vh = hv$.

3. Montrer que, pour tout $w \in \mathrm{O(E)}$, on a $|\operatorname{Tr}(wu)| \leqslant \operatorname{Tr} h$.

4. Montrer que l'application N de $\mathcal{L}(\mathrm{E})$ dans $\mathbb{R}$ définie par $\mathrm{N}(u) = \operatorname{Tr} h$ est une norme sur $\mathcal{L}(\mathrm{E})$.

(École polytechnique)

Solution.

1. Si $u = vh$, avec v orthogonal et h auto-adjoint positif, on a $u^* = h^*v^* = hv^{-1}$ et $u^*u = h^2$. L'endomorphisme u^*u est auto-adjoint et positif, car pour tout $x \in \mathrm{E}$, $(u^*u(x), x) = \|u(x)\|^2 \geqslant 0$. Classiquement (voir exercice 2.12), il existe un unique endomorphisme auto-adjoint positif h tel que $h^2 = u^*u$.

Si u est inversible, h l'est aussi, et on a nécessairement $v = uh^{-1}$. On vérifie que cet endomorphisme est orthogonal : $v^* = h^{*-1}u^* = h^{-1}u^*$ et $v^*v = h^{-1}u^*uh^{-1} = h^{-1}h^2h^{-1} = \mathrm{Id_E}$. Dans ce cas, le couple (v, h) existe et est unique.

Supposons maintenant que u ne soit pas inversible. L'égalité $u^*u = h^2 = h^*h$ implique encore l'existence de $v \in \mathrm{O(E)}$ tel que $u = vh$ (c'est le résultat de l'exercice 1.28). Lors de l'oral on reproduirait par exemple la preuve déjà donnée dans la solution de l'exercice 2.24 : soit $(e_1, \dots, e_n)$ une base orthonormale de vecteurs propres de u^*u, λ_i étant la valeur propre associée à e_i. On suppose que $\lambda_1, \dots, \lambda_r$ sont différents de 0 (où $r \in [\![0, n-1]\!]$ est le rang de u^*u) et que $\lambda_{r+1} = \cdots = \lambda_n = 0$. On a, pour tout $(i, j) \in [\![1, n]\!]^2$,

$$\langle u(e_i), u(e_j)\rangle = \langle u^*u(e_i), e_j\rangle = \langle \lambda_i e_i, e_j\rangle = \begin{cases} 0 & \text{si} \quad i \neq j \\ \lambda_i & \text{si} \quad i = j. \end{cases}$$

Pour $1 \leqslant i \leqslant r$, on pose $f_i = \dfrac{1}{\sqrt{\lambda_i}} u(e_i)$; la famille $(f_1, \dots, f_r)$ est orthonormale. On la complète en une base orthonormale $(f_1, \dots, f_n)$ de E. On a, pour tout $i \in [\![1, n]\!]$, $u(e_i) = \sqrt{\lambda_i} f_i$, par définition si $i \leqslant r$ et parce que les deux termes sont nuls sinon. Soit $v \in \mathcal{L}(\mathrm{E})$ tel que, pour tout $i \in [\![1, n]\!]$, $v(e_i) = f_i$. L'endomorphisme v est orthogonal et, pour tout $i \in [\![1, n]\!]$, $u(e_i) = \sqrt{\lambda_i} v(e_i) = v(\sqrt{\lambda_i} e_i) = v(h(e_i))$. On a donc $u = vh$. On voit que dans ce cas v n'est pas unique, car on peut choisir arbitrairement la base $(f_{r+1}, \dots, f_n)$ de $(\operatorname{Im} u)^\perp$.

2. Avec les notations de la question précédente, on a $u^*u = h^2$ et $uu^* = vh^2v^{-1}$. On en déduit que

$$u^*u = uu^* \iff vh^2 = h^2v.$$

Si v et h commutent, il est clair que v commute également avec h^2. Si

réciproquement v commute avec h^2, alors v commute aussi avec h qui est un polynôme en h^2 (en vertu de l'exercice 2.12).

3. On a $\mathrm{Tr}(uu) = \mathrm{Tr}(uvh)$. Quand u décrit $\mathrm{O(E)}$, uv décrit aussi $\mathrm{O(E)}$. Il s'agit donc de prouver que $|\mathrm{Tr}(wh)| \leqslant \mathrm{Tr}(h)$ pour tout w de $\mathrm{O(E)}$. En gardant les notations précédentes, on a

$$\mathrm{Tr}(wh) = \sum_{i=1}^{n} \langle (wh)(e_i), e_i \rangle = \sqrt{\lambda_i} \langle w(e_i), e_i \rangle.$$

D'après l'inégalité de Cauchy-Schwarz, on a $|\langle w(e_i), e_i \rangle| \leqslant 1$, car w est orthogonal. On en déduit que $|\mathrm{Tr}(wh)| \leqslant \sum_{i=1} \sqrt{\lambda_i} = \mathrm{Tr}(h)$.

4. Soit $u \in \mathcal{L}(\mathrm{E})$ et l'endomorphisme h qui lui est associé comme à la question 1. Comme h est symétrique positif, on a $\mathrm{Tr}\, h \geqslant 0$ et $\mathrm{N}(u)$ appartient à $\mathbb{R}_+$. Si $\mathrm{N}(u) = 0$, les valeurs propres de h sont nulles donc $h = 0$ et comme u s'écrit $u = vh$ avec $v \in \mathrm{O(E)}$, on a $u = 0$.

La question précédente prouve que $\mathrm{N}(u) = \sup_{w \in \mathrm{O(E)}} |\mathrm{Tr}(wu)|$. Si $\alpha \in \mathbb{R}$, on a

$$\begin{aligned} \mathrm{N}(\alpha u) &= \sup_{w \in \mathrm{O(E)}} |\mathrm{Tr}(w(\alpha u))| = \sup_{w \in \mathrm{O(E)}} |\alpha| |\mathrm{Tr}(wu)| \\ &= |\alpha| \sup_{w \in \mathrm{O(E)}} |\mathrm{Tr}(wu)| = |\alpha| \mathrm{N}(u). \end{aligned}$$

Enfin, si $u, u' \in \mathcal{L}(\mathrm{E})$, pour tout $w \in \mathrm{O(E)}$, on a

$$\begin{aligned} |\mathrm{Tr}(w(u+u'))| &= |\mathrm{Tr}(wu) + \mathrm{Tr}(wu')| \leqslant |\mathrm{Tr}(wu)| + |\mathrm{Tr}(wu')| \\ &\leqslant \mathrm{N}(u) + \mathrm{N}(u'). \end{aligned}$$

Il s'ensuit que $\mathrm{N}(u+u') \leqslant \mathrm{N}(u) + \mathrm{N}(u')$. On conclut que N est bien une norme sur $\mathcal{L}(\mathrm{E})$. ◁

La décomposition polaire est un résultat important qui est souvent posé à l'oral de manière matricielle. Voici un second énoncé, que nous rajoutons parce que la solution que nous donnons pour passer du cas inversible au cas général est différente.

2.27. Décomposition polaire (2)

Soit $A \in \mathcal{M}_n(\mathbb{R})$.

1. On suppose A inversible. Montrer qu'il existe un unique couple (O, S) telle que $A = OS$ avec O orthogonale et S symétrique définie positive.

2. On ne suppose plus que A est inversible. Vérifier que A s'écrit $A = OS$ avec O orthogonale et S symétrique positive.

(**École polytechnique**)

Solution.

1. Il ne s'agit là que de la traduction matricielle de la première question de l'exercice précédent. Supposons que (O, S) réponde au problème, on a ${}^tAA = {}^tS\,{}^tOOS = S^2$ et S est donc l'unique racine symétrique définie positive de la matrice définie positive ${}^tAA$ (voir exercices 2.14 et 2.12). L'unicité de O en découle. Pour l'existence, S désignant cette racine carrée, on pose $O = AS^{-1}$ et on vérifie qu'il s'agit d'une matrice orthogonale :

$$ {}^tOO = {}^tS^{-1}\,{}^tAAS^{-1} = S^{-1}S^2S^{-1} = I_n. $$

Le couple (O, S) répond bien au problème.

2. Là encore, on peut adapter la première question de l'exercice précédent. On peut également utiliser un argument de densité : comme $GL_n(\mathbb{R})$ est dense dans $\mathcal{M}_n(\mathbb{R})$, on peut considérer une suite $(A_p)_{p\in\mathbb{N}}$ de matrices inversibles qui converge vers A. On écrit chaque A_p sous la forme $A_p = O_pS_p$ avec O_p orthogonale et S_p symétrique définie positive. De la suite $(O_p)_{p\in\mathbb{N}}$ contenue dans le compact $O_n(\mathbb{R})$, on peut extraire une sous-suite convergente vers une matrice $O \in O_n(\mathbb{R})$. Quitte à considérer cette extraction, on peut supposer que O_p converge vers O. Dans ces conditions, il en va de même $S_p = {O_p}^{-1}A$ qui converge vers une matrice S. Bien évidemment, par passage à la limite, S est symétrique et $S = O^{-1}A$. De plus, S est également positive, car pour tout $X \in \mathbb{R}^n$,

$$ 0 \leqslant {}^tXS_pS \xrightarrow[p\to\infty]{} {}^tXSX, $$

et ${}^tXSX \geqslant 0$. On obtient donc $A = OS$ avec O orthogonale et S symétrique positive. ◁

La grande similitude qu'il y a entre les solutions des deux exercices qui précèdent et celle de l'exercice 2.24 n'est pas un hasard. Il est effectivement très simple de passer de la décomposition polaire à la décomposition en valeurs singulières et vice-versa. Expliquons-le matriciellement. Si $A = UDV$ *avec* U, V *orthogonales et* D *diagonale positive il suffit d'écrire* $A = (U\,{}^tV^{-1})({}^tVDV)$ *pour avoir une décomposition polaire. Réciproquement, si* $A = OS$ *avec* O *orthogonale et* S *symétrique positive, il suffit de dire que* S *est orthogonalement semblable à une matrice diagonale positive pour obtenir une décomposition* UDV.

Voici une jolie application géométrique de la décomposition polaire : il s'agit de déterminer les points extrémaux de la boule unité de $\mathcal{L}(E)$ *pour la triple norme associée à la norme euclidienne de* E. *On retiendra que de manière générale la décomposition polaire est très utile pour situer des matrices relativement au groupe orthogonal (voir par exemple le calcul de* $\max_{P\in O_n(\mathbb{R})} \mathrm{Tr}(PA)$ *dans la question 3 de l'exercice 2.26).*

2.28. Points extrémaux de la boule unité de $\mathcal{L}(E)$

Soit E un espace euclidien et $B = \{u \in \mathcal{L}(E),\ \|u\| \leqslant 1\}$. Montrer que les points extrémaux de B (*i.e.* les $u \in B$ tels que $B - \{u\}$ est convexe) sont exactement les éléments de O(E).

(École polytechnique)

▷ **Solution.**

La norme triple de l'énoncé est bien entendu la norme subordonnée à la norme euclidienne de E qui sera notée $\|\ \|$. Un point A d'un convexe X est extrémal si $X - \{A\}$ est encore convexe, ce qui équivaut à dire que A ne peut pas s'écrire comme milieu de deux points distincts de X. Par exemple, les points extrémaux d'un carré plein de $\mathbb{R}^2$ sont ses sommets.

Il est possible d'avoir une petite intuition du résultat en regardant la dimension 1. Dans ce cas, E s'identifie à $\mathbb{R}$ de même que $\mathcal{L}(E)$ et B s'identifie au segment $[-1, 1]$. Ses points extrémaux sont 1 et -1 ce qui correspond bien à $O(E) = \{\pm \mathrm{Id}_E\}$. Traitons maintenant le cas général.

• Commençons par prouver qu'un élément $u \in O(E)$ est extrémal. Bien entendu $u \in B$ puisque $\|u\| = 1$. Supposons pour cela que $u = \frac{1}{2}(v + w)$ avec $(v, w) \in B^2$. Soit $x \in E$ unitaire. On a

$$\begin{aligned}\|x\| = 1 = \|u(x)\| &= \frac{1}{2}\|v(x) + w(x)\| \leqslant \frac{1}{2}\|v(x)\| + \frac{1}{2}\|w(x)\| \\ &\leqslant \frac{1}{2}(\|v\| + \|w\|) \leqslant 1.\end{aligned}$$

La première inégalité provient de l'inégalité triangulaire, la seconde de la définition de la norme triple et la dernière résulte du fait que v et w sont dans B. Toutes ces inégalités sont en fait des égalités. On a donc $\|v\| = \|w\| = 1$, $\|v(x)\| = \|w(x)\| = 1$ et de plus les vecteurs $v(x)$ et $w(x)$ sont positivement liés (cas d'égalité dans l'inégalité triangulaire pour une norme euclidienne). On a donc nécessairement $v(x) = w(x)$. Comme cela vaut pour tout vecteur unitaire x, on a par linéarité $v = w$. Donc u est bien un point extrémal de B.

• Inversement, soit $u \in B$ n'appartenant pas à O(E). On va montrer que u n'est pas extrémal. On va travailler matriciellement en considérant la matrice A de u dans une base orthonormée de E et utiliser la décomposition polaire de A (cf. exercice 2.26 et 2.27) : A s'écrit $A = OS$ avec O orthogonale et S symétrique positive. La matrice S est orthogonalement semblable à une matrice diagonale $D = \mathrm{Diag}(d_1, \ldots, d_n)$, où l'on suppose $d_1 \leqslant \cdots \leqslant d_n$: $S = {}^tPDP$ avec $P \in O_n(\mathbb{R})$ (et $n = \dim E$). Comme $\|S\| = \|A\| \leqslant 1$, on a pour tout $k \in [\![1, n]\!]$, $d_k \in [0, 1]$. Par hypothèse, A n'est pas orthogonale donc d_1 est stric-

lement inférieur à 1. On peut alors écrire d_1 sous la forme $d_1 = \dfrac{\alpha+\beta}{2}$ avec $-1 \leqslant \alpha < \beta \leqslant 1$ (car d_1 n'est pas un point extrémal du segment $[-1, 1]$). Posons $\mathrm{D}' = \mathrm{Diag}(\alpha, d_2, \ldots, d_n)$ et $\mathrm{D}'' = \mathrm{Diag}(\beta, d_2, \ldots, d_n)$. On a $\dfrac{1}{2}(\mathrm{D}' + \mathrm{D}'') = \mathrm{D}$ avec $\mathrm{D}' \neq \mathrm{D}''$ et donc

$$\mathrm{A} = \frac{1}{2}(\mathrm{O}\,{}^t\mathrm{PD}'\mathrm{P} + \mathrm{O}\,{}^t\mathrm{PD}''\mathrm{P}).$$

Les matrices $\mathrm{O}\,{}^t\mathrm{PD}'\mathrm{P}$ et $\mathrm{O}\,{}^t\mathrm{PD}''\mathrm{P}$ sont de norme 1. En effet, pour tout vecteur X de norme 1, on a

$$\|\mathrm{O}\,{}^t\mathrm{PD}'\mathrm{PX}\|^2 = {}^t\mathrm{X}\,{}^t\mathrm{PD}'\mathrm{P}\,{}^t\mathrm{OO}\,{}^t\mathrm{PD}'\mathrm{PX} = {}^t(\mathrm{PX})\mathrm{D}'^2(\mathrm{PX}) \leqslant 1,$$

car $\|\mathrm{PX}\| = 1$ et les coefficients diagonaux de D'^2 sont compris entre 0 et 1. On a donc $\|\mathrm{O}\,{}^t\mathrm{PD}'\mathrm{P}\| \leqslant 1$. Il en est de même de $\mathrm{O}\,{}^t\mathrm{PD}''\mathrm{P}$. En repassant aux endomorphismes, u peut donc s'écrire comme milieu de deux éléments distincts de B : ce n'est pas un point extrémal. ◁

2.29. Décomposition de Cholesky. Décomposition QR

1. *Décomposition de Cholesky.* Soit $\mathrm{A} = (a_{ij})_{1\leqslant i,j\leqslant n}$ une matrice symétrique définie positive. Démontrer l'existence d'une unique matrice P de taille n, triangulaire supérieure à coefficients diagonaux strictement positifs telle que $\mathrm{A} = {}^t\mathrm{PP}$.

2. *Décomposition QR.* Soit $\mathrm{A} \in \mathrm{GL}_n(\mathbb{R})$. Montrer qu'il existe un unique couple (Q, R) avec Q orthogonale de taille n et $\mathrm{R} \in \mathcal{M}_n(\mathbb{R})$ triangulaire supérieure à coefficients diagonaux strictement positifs telles que $\mathrm{A} = \mathrm{QR}$.

(**École normale supérieure**)

Solution.

1. Considérons $\varphi : (\mathrm{X}, \mathrm{Y}) \in (\mathbb{R}^n)^2 \longmapsto {}^t\mathrm{XAY}$. Il s'agit là d'un produit scalaire. Il est possible d'orthonormaliser au sens de Gram-Schmidt la base canonique $(e_1, \ldots, e_n)$ en une base orthonormée $(\varepsilon_1, \ldots, \varepsilon_n)$ pour φ. Si l'on note $\mathrm{P} = (p_{ij})_{1\leqslant i,j\leqslant n}$ la matrice de passage des ε_i aux e_i, comme I_n est la matrice de φ dans la base des ε_i, on a $\mathrm{A} = {}^t\mathrm{PI}_n\mathrm{P} = {}^t\mathrm{PP}$. La matrice P est triangulaire supérieure et ses coefficients diagonaux sont strictement positifs : en effet, ce sont les produits scalaires $\varphi(e_i, \varepsilon_i)$ et ils sont strictement positifs d'après les règles de l'orthonormalisation de Gram-Schmidt.

Si Q est une autre matrice triangulaire supérieure à coefficients diagonaux strictement positifs vérifiant $\mathrm{A} = {}^t\mathrm{QQ} = {}^t\mathrm{PP}$, la matrice

$D = {}^tP^{-1}\,{}^tQ = PQ^{-1}$ est à la fois triangulaire supérieure et triangulaire inférieure. Elle est donc diagonale et ${}^tD = QP^{-1} = D^{-1}$, ce qui signifie qu'elle est orthogonale. Comme elle est à coefficients strictement positifs, $D = I_n$ et $P = Q$. L'unicité est prouvée.

On peut aussi justifier l'unicité de la manière suivante : si $A = {}^tQQ$, *avec* Q *triangulaire supérieure à coefficients diagonaux strictement positifs,* Q *peut être vue comme la matrice de passage d'une base* $\mathcal{C}$ *à la base canonique. Comme* $I_n = {}^tQ^{-1}AQ^{-1}$, *la base* $\mathcal{C}$ *est orthonormée pour* φ. *Le fait que* Q *soit triangulaire supérieure à coefficients diagonaux strictement positifs entraîne que* $\mathcal{C}$ *est l'orthonormalisé au sens de Gram-Schmidt de la base canonique. Donc* $\mathcal{C} = (\varepsilon_1, \ldots, \varepsilon_n)$ *et* $P = Q$.

2. Si le couple (Q, R) convient, ${}^tAA = {}^tR\,{}^tQQR = {}^tRR$. La première question nous garantit l'unicité de R et donc celle de Q.

Passons à l'existence : la matrice ${}^tAA$ est définie positive (voir par exemple l'exercice 2.14). D'après la première question, elle s'écrit ${}^tAA = {}^tRR$ avec R triangulaire supérieure à coefficients diagonaux strictement positifs. On pose $Q = AR^{-1}$. Calculons ${}^tQQ$:

$$ {}^tQQ = {}^tR^{-1}\,{}^tAAR^{-1} = {}^tR^{-1}\,{}^tRRR^{-1} = I_n. $$

La matrice Q est bien orthogonale et l'existence du couple (Q, R) est prouvée. $\lhd$

Cette décomposition avait déjà fait l'objet de l'exercice 1.21. On avait alors interprété le produit QR *comme un produit de matrices de passage faisant intervenir l'orthonormalisé au sens de Gram-Schmidt des colonnes de* A. *Cela fournit d'ailleurs un algorithme de calcul effectif de* Q *et* R. *Notons qu'une autre méthode due à Householder permet aussi, par multiplications successives par des matrices de réflexions, de déterminer le couple* QR. *Signalons que si on écrit* $R = DT$ *avec* D *diagonale à coefficients diagonaux strictement positifs (ceux de* R*) et* T *matrice triangulaire supérieure avec des* 1 *sur la diagonale, on obtient la décomposition d'Iwasawa* $A = QDT$. *Notons enfin qu'on peut poser l'exercice à l'envers et déduire la décomposition de Choleski à partir de la décomposition QR. Si* A *est symétrique définie positive on sait qu'elle s'écrit* $A = B^2$ *avec* B *symétrique définie positive (voir l'exercice 2.12). On écrit alors* $B = QR$ *et il vient* $A = B^2 = {}^tBB = {}^tRR$.

Ces décompositions vont assurer la transition avec le vaste sujet suivant qui porte sur les inégalités autour du déterminant et des valeurs propres de matrices réelles. En l'occurrence, l'inégalité d'Hadamard donne une majoration du déterminant d'une matrice symétrique réelle à l'aide de ses coefficients diagonaux.

2.30. Inégalité d'Hadamard

1. Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique positive. Démontrer que

$$\det A \leqslant \prod_{i=1}^{n} a_{ii}.$$

2. Soit E un espace euclidien de dimension n, f un endomorphisme auto-adjoint de E défini positif. On désigne par $\lambda_1, \ldots, \lambda_n$ les valeurs propres de f et $\mathcal{A}$ l'ensemble des bases orthonormales de E. Montrer que

$$\prod_{i=1}^{n} \lambda_i = \inf_{(e_1,\ldots,e_n)\in\mathcal{A}} \prod_{i=1}^{n} \langle f(e_i), e_i \rangle.$$

(École polytechnique)

Solution.

1. Si on note $\varphi : (X, Y) \in (\mathbb{R}^n)^2 \longmapsto {}^tXAY$ la forme bilinéaire symétrique associée à A et $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$, on a $a_{ii} = \varphi(e_i, e_i) \geqslant 0$ pour tout $1 \leqslant i \leqslant n$ puisque A est supposée positive. Si A n'est pas définie positive, on a donc $0 = \det A \leqslant a_{11} \ldots a_{nn}$. Supposons désormais A définie positive. Nous allons proposer trois argumentations.

Première méthode (à partir de la décomposition de Choleski) : on écrit comme dans l'exercice précédent $A = {}^tPP$ avec $P = (p_{ij})_{1 \leqslant i,j \leqslant n}$ triangulaire supérieure, à coefficients diagonaux strictement positifs. En passant au déterminant, il vient $\det A = \det {}^tP \det P = (\det P)^2 = \prod_{i=1}^{n} p_{ii}^2$ puisque P est triangulaire. Or pour tout $i \in [\![1, n]\!]$, $a_{ii} = \sum_{k=1}^{n} p_{ki}^2 \geqslant p_{ii}^2$ et finalement

$$\det A = \prod_{i=1}^{n} p_{ii}^2 \leqslant \prod_{i=1}^{n} a_{ii}.$$

Deuxième méthode (à partir de la diagonalisation de A*) :* il existe des matrices $O = (b_{ij})_{1 \leqslant i,j \leqslant n}$ orthogonale et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, à coefficients diagonaux strictement positifs tels que $A = OD\,{}^tO$. On en déduit, pour tout $i \in [\![1, n]\!]$, $a_{ii} = \sum_{j=1}^{n} b_{ij}^2 \lambda_j$. Comme $\sum_{j=1}^{n} b_{ij}^2 = 1$, puisque O est orthogonale, on obtient, par concavité de la fonction ln, $\ln a_{ii} \geqslant \sum_{j=1}^{n} b_{ij}^2 \ln \lambda_j$, puis

$$\ln \prod_{i=1}^{n} a_{ii} \geqslant \sum_{1 \leqslant i,j \leqslant n} b_{ij}^2 \ln \lambda_j = \sum_{j=1}^{n} \ln \lambda_j \left(\sum_{i=1}^{n} b_{ij}^2 \right) = \sum_{j=1}^{n} \ln \lambda_j,$$

car O est orthogonale. On a donc

$$\ln \prod_{i=1}^{n} a_{ii} \geqslant \ln \det \mathrm{A},$$

d'où l'inégalité recherchée.

Troisième méthode (à partir de la décomposition QR*) :* c'est la preuve qui a été présentée dans l'exercice 1.21. Soit B l'unique matrice symétrique définie positive telle que $\mathrm{A} = \mathrm{B}^2 = {}^t\mathrm{BB}$. Si $\mathrm{C}_1, \ldots, \mathrm{C}_n$ sont les colonnes de B on a d'après l'exercice mentionné $\det \mathrm{B} \leqslant \|\mathrm{C}_1\| \ldots \|\mathrm{C}_n\|$. Or

$$\det \mathrm{A} = (\det \mathrm{B})^2 \leqslant \|\mathrm{C}_1\|^2 \ldots \|\mathrm{C}_n\|^2 = a_{11} \ldots a_{nn}.$$

2. Les valeurs propres de f sont réelles positives et f est diagonalisable dans une base orthonormale $\mathcal{B}_0 = (\varepsilon_1, \ldots, \varepsilon_n)$, chaque ε_i étant vecteur propre relatif à la valeur propre λ_i. Le produit des λ_i est le déterminant de f.

Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base orthonormale quelconque de E et $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n}$ la matrice de f dans cette base. On a $\prod_{i=1}^{n} \langle f(e_i), e_i \rangle = \prod_{i=1}^{n} a_{ii}$ puisque $\mathcal{B}$ est orthonormale et d'après la question précédente,

$$\prod_{i=1}^{n} \lambda_i = \det f = \det \mathrm{A} \leqslant \prod_{i=1}^{n} \langle f(e_i), e_i \rangle.$$

On en déduit $\prod_{i=1}^{n} \lambda_i \leqslant \inf_{(e_1, \ldots, e_n) \in \mathcal{A}} \prod_{i=1}^{n} \langle f(e_i), e_i \rangle$ et en considérant la base $\mathcal{B}_0 \in \mathcal{A}$, on conclut qu'il y a en fait égalité. ◁

Le résultat suivant se déduit aisément de l'inégalité d'Hadamard.

2.31. Inégalité de Fischer

Soit $\mathrm{A} = \begin{pmatrix} \mathrm{A}_1 & \mathrm{B} \\ {}^t\mathrm{B} & \mathrm{A}_2 \end{pmatrix}$ une matrice réelle symétrique positive écrite en blocs, les blocs A_1 et A_2 étant carrés.

Montrer que $\det \mathrm{A} \leqslant \det \mathrm{A}_1 \det \mathrm{A}_2$.

(École polytechnique)

1 - **Solution.**

On note p et q les tailles respectives des matrices A_1 et A_2. Nous allons donner deux solutions.

• *Puisque* A_1 et A_2 sont symétriques réelles, il existe des matrices $P \in O_p(\mathbb{R})$ et $Q \in O_q(\mathbb{R})$ et des matrices diagonales D_1 et D_2 telles que $D_1 = {}^tPD_1P$ et $D_2 = {}^tQD_2Q$. On pose $S = \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix}$. On remarque que $S \in O_n(\mathbb{R})$ et on note $A' = {}^tSAS$. On obtient

$$A' = \begin{pmatrix} {}^tPA_1 & {}^tPB \\ {}^tQ\,{}^tB & {}^tQA_2 \end{pmatrix} \begin{pmatrix} P & 0 \\ 0 & Q \end{pmatrix} = \begin{pmatrix} {}^tPA_1P & {}^tPBQ \\ {}^tQ\,{}^tBP & {}^tQA_2Q \end{pmatrix} = \begin{pmatrix} D_1 & B' \\ {}^tB' & D_2 \end{pmatrix},$$

où $B' = {}^tPBQ$. La matrice $A' = (a_{ij})_{1 \leqslant i,j \leqslant n}$ est encore symétrique positive et l'on a

$$\det A = \det A', \quad \det A_1 \det A_2 = \det D_1 \det D_2 = \prod_{i=1}^{n} a_{ii}.$$

Cela revient donc à démontrer que $\det A' \leqslant \prod_{i=1}^{n} a_{ii}$ pour une matrice A' symétrique positive, ce qui est fait dans l'exercice 2.30.

• Voici une seconde solution qui n'utilise pas l'inégalité d'Hadamard, mais la décomposition de Choleski. On peut se restreindre au cas où A est définie positive, car sinon $\det A = 0$ et $\det A_1 \det A_2 \geqslant 0$, puisque A_1 et A_2 sont aussi positives. On peut alors écrire $A = {}^tPP$ avec P triangulaire supérieure (cf. exercice 2.29). Posons $P = \begin{pmatrix} U & V \\ 0 & W \end{pmatrix}$ où les blocs ont la même taille que dans A. On a alors $A_1 = {}^tUU$ et $A_2 = {}^tVV + {}^tWW$. Comme $\det A = (\det P)^2 = (\det U)^2(\det W)^2$ et que $\det A_1 = (\det U)^2$ le problème se ramène à montrer que

$$(\det W)^2 = \det({}^tWW) \leqslant \det A_2 = \det({}^tWW + {}^tVV).$$

Cela découle de l'inégalité classique suivante : si M, N sont symétriques positives, $\det(M+N) \geqslant \det M + \det N \geqslant \det M$. Le lecteur en trouvera la preuve dans l'exercice 3.29 du chapitre 3. ◁

Le résultat se généralise sans mal avec plus de blocs.

On trouvera encore dans l'exercice suivant une application de l'inégalité d'Hadamard.

2.32. Recherche d'un minimum

Soit A une matrice symétrique définie positive et α un réel strictement positif. On note $\mathcal{S}$ l'ensemble des matrices symétriques définies positives de déterminant supérieur ou égal à α.

Montrer que $\inf_{S\in\mathcal{S}} \operatorname{Tr}(AS) = n(\alpha \det A)^{1/n}$.

(École polytechnique)

▷ **Solution.**

Il existe $D = \operatorname{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale à termes diagonaux strictement positifs, et P orthogonale telles que $A = PD\,{}^tP$. On en déduit que si $S \in \mathcal{S}$, $\operatorname{Tr}(AS) = \operatorname{Tr}(PD\,{}^tPS) = \operatorname{Tr}(D\,{}^tPSP)$. La matrice $S' = {}^tPSP = (s_{ij})_{1\leqslant i,j\leqslant n}$ est symétrique définie positive, $\det S' = \det S$ et

$$\operatorname{Tr}(AS) = \operatorname{Tr}(DS') = \sum_{i=1}^{n} \lambda_i s_{ii}.$$

Par concavité du logarithme on obtient

$$\ln\left(\frac{1}{n}\operatorname{Tr}(AS)\right) \geqslant \frac{1}{n}\sum_{i=1}^{n} \ln(\lambda_i s_{ii}) \geqslant \frac{1}{n}\ln\left(\prod_{i=1}^{n} \lambda_i s_{ii}\right)$$
$$\geqslant \frac{1}{n}\ln\left(\det A \prod_{i=1}^{n} s_{ii}\right).$$

Il a été démontré dans l'exercice 2.30 que si S' est symétrique définie positive, on a $\prod_{i=1}^{n} s_{ii} \geqslant \det S'$. On en déduit

$$\ln\left(\frac{1}{n}\operatorname{Tr}(AS)\right) \geqslant \frac{1}{n}\ln(\det A \det S') \geqslant \frac{1}{n}\ln(\alpha \det A)$$

et donc

$$\frac{1}{n}\operatorname{Tr}(AS) \geqslant (\alpha \det A)^{1/n}.$$

On obtient finalement $\operatorname{Tr}(AS) \geqslant n(\alpha \det A)^{1/n}$ pour tout $S \in \mathcal{S}$.

Montrons qu'on peut choisir S pour qu'il y ait égalité, ce qui terminera la démonstration. On choisit S telle que S' soit diagonale avec $s_{ii} = \dfrac{\lambda}{\lambda_i}$ ($\lambda > 0$). On a alors $\operatorname{Tr}(AS) = \operatorname{Tr}(DS') = n\lambda$. On choisit λ pour que $\det S = \det S' = \alpha$. Il faut $\dfrac{\lambda^n}{\det A} = \alpha$ soit $\lambda = (\alpha \det A)^{1/n}$. La matrice $S = {}^tPS'P$ appartient alors à $\mathcal{S}$ et $\operatorname{Tr}(AS) = n(\alpha \det A)^{1/n}$. ◁

On peut en déduire l'inégalité de Minkowski suivante : si A, B *sont symétriques définies positives alors*

$$(\det \mathrm{A})^{1/n} + (\det \mathrm{B})^{1/n} \leqslant (\det(\mathrm{A}+\mathrm{B}))^{1/n}.$$

En effet, prenons $\alpha = 1$ *pour simplifier. Pour* $\mathrm{S} \in \mathcal{S}$ *on a*

$$(\det \mathrm{A})^{1/n} + (\det \mathrm{B})^{1/n} \leqslant \frac{1}{n}\Big(\mathrm{Tr}(\mathrm{AS}) + \mathrm{Tr}(\mathrm{BS})\Big) = \frac{1}{n}\mathrm{Tr}(\mathrm{A}+\mathrm{B})\mathrm{S}$$

et il suffit de passer à la borne inférieure pour conclure. Une autre démonstration de cette inégalité est donnée dans l'exercice 3.32.

2.33. Image de S^2 par une fonction, où S est la boule unité

L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique. Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive. Pour X et Y dans $\mathbb{R}^n$, on pose

$$\varphi(\mathrm{X},\mathrm{Y}) = ({}^t\mathrm{XAX})({}^t\mathrm{YAY}) - ({}^t\mathrm{XAY})^2.$$

Déterminer l'image par φ de $\mathrm{B} = \{(\mathrm{X},\mathrm{Y}),\ \|\mathrm{X}\| \leqslant 1,\ \|\mathrm{Y}\| \leqslant 1\}$.

(École normale supérieure)

Solution.

On note S la boule unité fermée de $\mathbb{R}^n$. On remarque que si $n = 1$, φ est la fonction nulle et $\varphi(\mathrm{B}) = \{0\}$. Nous écartons ce cas pour la suite. Puisque φ est continue et que $\mathrm{B} = \mathrm{S}^2$ est connexe, l'image de B par φ est connexe. C'est donc un intervalle de $\mathbb{R}$. La forme bilinéaire définie sur $\mathbb{R}^n$ par $(\mathrm{X},\mathrm{Y}) = {}^t\mathrm{XAY}$ est un produit scalaire et on a $\varphi(\mathrm{X},\mathrm{Y}) = \langle \mathrm{X},\mathrm{X}\rangle\langle \mathrm{Y},\mathrm{Y}\rangle - \langle \mathrm{X},\mathrm{Y}\rangle^2 \geqslant 0$, d'après l'inégalité de Cauchy-Schwarz, la valeur 0 étant obtenue pour $\mathrm{X} = 0$. On a donc la borne inférieure de l'intervalle image.

On note $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ les valeurs propres de A. On se place dans une base orthonormale (pour le produit scalaire canonique) de vecteurs propres, disons $(e_1,\ldots,e_n)$ avec $\mathrm{A}e_i = \lambda_i e_i$. On note $(x_1,\ldots,x_n)$ et $(y_1,\ldots,y_n)$ les coordonnées de X et Y dans cette base. On obtient

$$\begin{aligned}\varphi(\mathrm{X},\mathrm{Y}) &= \left(\sum_{i=1}^n \lambda_i x_i^2\right)\left(\sum_{j=1}^n \lambda_j y_j^2\right) - \left(\sum_{i=1}^n \lambda_i x_i y_i\right)^2 \\ &= \sum_{1\leqslant i,j\leqslant n} \lambda_i\lambda_j(x_i^2 y_j^2 - x_i x_j y_i y_j)\end{aligned}$$

$$= \sum_{i<j} \lambda_i \lambda_j (x_i y_j - x_j y_i)^2.$$

Puisque les valeurs propres de A sont positives et rangées par ordre décroissant, on a $\varphi(X, Y) \leqslant \lambda_1 \lambda_2 \sum_{i<j} (x_i y_j - x_j y_i)^2$ et en reprenant le calcul précédent dans l'autre sens avec $A = I_n$, on obtient

$$\sum_{i<j} (x_i y_j - x_j y_i)^2 = \|X\|^2 \|Y\|^2 - \left({}^t XY\right)^2 \leqslant 1,$$

avec égalité si X et Y sont orthogonaux et de norme 1. Ainsi, on a la majoration

$$\varphi(X, Y) \leqslant \lambda_1 \lambda_2 \sum_{i<j} (x_i y_j - x_j y_i)^2 \leqslant \lambda_1 \lambda_2.$$

Comme $\varphi(e_1, e_2) = \lambda_1 \lambda_2$, on conclut que $\boxed{\varphi(B) = [0, \lambda_1 \lambda_2]}$. ◁

On peut considérer $f(X) = \varphi(X, X) = \|AX\|^2 - ({}^t XAX)^2$ et étudier les bornes d'une telle fonction sur $\mathbb{R}^n$. C'est ce que propose l'exercice suivant à ceci près que l'on considère un endomorphisme symétrique plutôt qu'une matrice.

2.34. Majoration et minoration d'une fonction

Soit E un espace vectoriel euclidien et u un endomorphisme auto-adjoint. Pour $x \in E$, on pose

$$f(x) = \|u(x)\|^2 - (u(x), x)^2.$$

1. La fonction f est-elle minorée ?

2. À quelle condition est-elle majorée ? Calculer alors $\sup_E f$.

(École polytechnique)

▷ **Solution.**

1. L'endomorphisme u est diagonalisable dans une base orthonormale. On note $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ ses valeurs propres et $\mathcal{B} = (e_1, \ldots, e_n)$ une base orthonormale de vecteurs propres, e_i étant associé à la valeur propre λ_i.

Étudions l'image par f d'un vecteur propre de u. Pour tout $t \in \mathbb{R}$, $f(te_i) = \lambda_i^2 (t^2 - t^4)$. Si $\lambda_i \neq 0$, $f(te_i)$ tend vers $-\infty$ quand t tend vers

$+\infty$ et f n'est pas minorée. La fonction f n'est minorée que si tous les λ_i sont nuls, *i.e.* si $u = 0$. On a alors $f = 0$.

2. Si $x = \sum_{i=1}^{n} x_i e_i$, avec $(x_1, \ldots, x_n) \in \mathbb{R}^n$, alors

$$f(x) = \sum_{i=1}^{n} \lambda_i^2 x_i^2 - \left(\sum_{i=1}^{n} \lambda_i x_i^2 \right)^2 .$$

S'il existe i et j tels que $\lambda_i \lambda_j < 0$, on considère x tel que $\langle u(x), x \rangle = 0$. On prend $x = \frac{t}{\sqrt{|\lambda_i|}} e_i + \frac{t}{\sqrt{|\lambda_j|}} e_j$ $(t \in \mathbb{R})$. On obtient $f(x) = t^2(|\lambda_i| + |\lambda_j|)$, qui tend vers $+\infty$ quand t tend vers $+\infty$. Ainsi f n'est pas majorée.

On suppose désormais que tous les λ_i ont le même signe. Quitte à changer u en $-u$, on peut supposer qu'ils sont positifs. En isolant les termes contenant λ_1, on obtient

$$\begin{aligned} f(x) &= -\sum_{i=1}^{n} \lambda_i(\lambda_1 - \lambda_i)x_i^2 + \lambda_1 \sum_{i=1}^{n} \lambda_i x_i^2 - \left(\sum_{i=1}^{n} \lambda_i x_i^2 \right)^2 \\ &= \frac{\lambda_1^2}{4} - \sum_{i=1}^{n} \lambda_i(\lambda_1 - \lambda_i)x_i^2 - \left(\frac{\lambda_1}{2} - \sum_{i=1}^{n} \lambda_i x_i^2 \right)^2 \leqslant \frac{\lambda_1^2}{4}, \end{aligned}$$

car pour tout i, $\lambda_1 - \lambda_i \geqslant 0$ et $\lambda_i \geqslant 0$. Puisque $f\left(\frac{e_1}{\sqrt{2}} \right) = \frac{\lambda_1^2}{4}$, la borne supérieure cherchée est $\frac{\lambda_1^2}{4}$. Si les valeurs propres de u sont toutes de même signe, on obtient donc

$$\boxed{\sup_E f = \frac{1}{4} \max_{\lambda \in \mathrm{Sp}\, u} \lambda^2}. \quad \lhd$$

2.35. Inégalité de Kantorovich

Soit A une matrice symétrique définie positive d'ordre n. On note λ_{min} (resp. λ_{max}) la plus petite (resp. grande) valeur propre de A. Montrer que si $\mathrm{X} \in \mathbb{R}^n$,

$$({}^t\mathrm{XX})^2 \leqslant ({}^t\mathrm{XAX})({}^t\mathrm{XA}^{-1}\mathrm{X}) \leqslant \frac{1}{4} \left(\sqrt{\frac{\lambda_{max}}{\lambda_{min}}} + \sqrt{\frac{\lambda_{min}}{\lambda_{max}}} \right)^2 ({}^t\mathrm{XX})^2.$$

(**École normale supérieure**)

▷ **Solution.**

Notons $0 < \lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres distinctes ou confondues de A. Cela revient à démontrer que pour $X \in \mathbb{R}^n$ de norme 1,

$$1 \leqslant ({}^tXAX)({}^tXA^{-1}X) \leqslant \frac{1}{4}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)^2.$$

Compte-tenu de l'inégalité de Cauchy-Schwarz, on peut écrire, on notant $(x_1, \ldots, x_n)$ les coordonnées de X dans une base orthonormale de vecteurs propres

$$\begin{aligned}({}^tXAX)({}^tXA^{-1}X) &= \left(\sum_{i=1}^n \lambda_i x_i^2\right)\left(\sum_{i=1}^n \frac{1}{\lambda_i} x_i^2\right) \\ &\geqslant \left(\sum_{i=1}^n \frac{\sqrt{\lambda_i}}{\sqrt{\lambda_i}} x_i^2\right)^2 = \left(\sum_{i=1}^n x_i^2\right)^2 = 1.\end{aligned}$$

Passons à la majoration. Comme $\sqrt{ab} \leqslant \frac{1}{2}(a+b)$ pour $a, b \geqslant 0$, on a, de manière astucieuse,

$$\begin{aligned}\sqrt{({}^tXAX)({}^tXA^{-1}X)} &= \sqrt{\frac{\lambda_1}{\lambda_n}\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} x_i^2\right)\left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} x_i^2\right)} \\ &\leqslant \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\left(\sum_{i=1}^n \frac{\lambda_i}{\lambda_1} x_i^2\right) + \left(\sum_{i=1}^n \frac{\lambda_n}{\lambda_i} x_i^2\right)\right) \\ &\leqslant \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(\frac{\lambda_i}{\lambda_1} + \frac{\lambda_n}{\lambda_i}\right) x_i^2\right)\end{aligned}$$

L'étude rapide de la fonction $f : x \longmapsto \dfrac{x}{\lambda_1} + \dfrac{\lambda_n}{x}$ sur l'intervalle $[\lambda_1, \lambda_n]$ montre que cette fonction est décroissante sur $[\lambda_1, \sqrt{\lambda_1\lambda_n}]$ et croissante sur $[\sqrt{\lambda_1\lambda_n}, \lambda_n]$. Elle admet donc un maximum en λ_1 ou en λ_n. Comme $f(\lambda_1) = f(\lambda_n) = 1 + \dfrac{\lambda_n}{\lambda_1}$, on peut majorer ainsi

$$\begin{aligned}\sqrt{({}^tXAX)({}^tXA^{-1}X)} &\leqslant \frac{1}{2}\sqrt{\frac{\lambda_1}{\lambda_n}}\left(\sum_{i=1}^n \left(1 + \frac{\lambda_n}{\lambda_1}\right) x_i^2\right) \\ &\leqslant \frac{1}{2}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right)\sum_{i=1}^n x_i^2 \\ &\leqslant \frac{1}{2}\left(\sqrt{\frac{\lambda_1}{\lambda_n}} + \sqrt{\frac{\lambda_n}{\lambda_1}}\right).\end{aligned}$$

En élevant au carré, on obtient l'inégalité souhaitée. ◁

2.36. Inégalité de convexité

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique définie positive de valeurs propres $\lambda_1, \ldots, \lambda_n$. Montrer que pour $p > 0$ on a

$$\left(\sum_{k=1}^{n} \lambda_k^p \right)^{1/p} \geqslant \left(\sum_{k=1}^{n} a_{kk}^p \right)^{1/p}$$

lorsque $p \geqslant 1$ et

$$\left(\sum_{k=1}^{n} \lambda_k^p \right)^{1/p} \leqslant \left(\sum_{k=1}^{n} a_{kk}^p \right)^{1/p}$$

lorsque $p \leqslant 1$.

(**École normale supérieure**)

Solution.

Notons que pour $p \geqslant 1$, la fonction $t \longmapsto t^p$ est convexe sur $\mathbb{R}_+$. Par hypothèse, on peut écrire $A = PD\,{}^tP$ où $P = (p_{ij})_{1 \leqslant i,j \leqslant n} \in O(n)$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. Pour tout $k \in [\![1, n]\!]$, on a $a_{kk} = \sum_{j=1}^{n} \lambda_j p_{kj}^2$. Comme les p_{kj}^2 sont positifs ou nuls de somme égale à 1, la convexité de $t \longmapsto t^p$ nous assure que si $p \geqslant 1$,

$$a_{kk}^p \leqslant \sum_{j=1}^{n} \lambda_j^p p_{kj}^2.$$

En sommant de $k = 1$ à n, on obtient

$$\sum_{k=1}^{n} a_{kk}^p \leqslant \sum_{k=1}^{n} \sum_{j=1}^{n} \lambda_j^p p_{kj}^2 = \sum_{j=1}^{n} \lambda_j^p \sum_{k=1}^{n} p_{kj}^2 = \sum_{j=1}^{n} \lambda_j^p.$$

En passant à la puissance $\dfrac{1}{p}$, on obtient la première inégalité. La seconde s'établit de même en observant que pour $p \leqslant 1$, $t \longmapsto t^p$ est concave. ◁

Nous avons déjà observé que si u est un endomorphisme auto-adjoint d'un espace euclidien E alors la plus grande (resp. petite) valeur propre de u est égale à $\max\limits_{\|x\|=1} \langle u(x), x \rangle$ *(resp.* $\min\limits_{\|x\|=1} \langle u(x), x \rangle$*). Le principe du minimax de Courant-Fischer-Weyl, que dans la suite nous baptiserons plus simplement théorème du minimax, donne une expression variationnelle pour les autres valeurs propres de u. Plus précisément, si on numérote*

dans l'ordre croissant les valeurs propres distinctes ou confondues de u, la p-ième valeur propre n'est autre que le minimum pris sur tous les sous-espaces F *de dimension p du maximum du produit scalaire $\langle u(x), x\rangle$ sur la sphère unité de* F. *Nous verrons ensuite plusieurs inégalités qui sont des conséquences directes de ce résultat et qui, à l'oral, peuvent être proposées à la suite de l'exercice suivant ou posées directement.*

2.37. Théorème du minimax

Soit E un espace euclidien de dimension n, f un endomorphisme auto-adjoint de E, $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres de f, $(e_i)_{1\leqslant i\leqslant n}$ une base orthonormale de vecteurs propres de f, e_i étant un vecteur propre associé à la valeur propre λ_i.

1. Pour tout $h \in [\![1, n]\!]$, on pose $\mathrm{E}_h = \mathrm{Vect}(e_1, e_2, \ldots, e_h)$ et $\mathrm{F}_h = \mathrm{Vect}(e_h, \ldots, e_n)$ et on note f_1 et f_2 les endomorphismes de E_h et F_h respectivement, induits par f. Caractériser λ_h connaissant f_1 et f_2.

2. *Théorème du minimax.* Soit $\mathcal{A}_h$ l'ensemble des sous-espaces de E de dimension h. Démontrer que

$$\lambda_h = \min_{\mathrm{F}\in\mathcal{A}_h} \left[\max_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle \right] = \max_{\mathrm{F}\in\mathcal{A}_{n-h+1}} \left[\min_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle \right].$$

(École polytechnique)

▷ **Solution.**

Remarquons que pour tout sous-espace F de E, les bornes de la fonction continue $x \longmapsto \langle f(x), x\rangle$ définie sur la sphère unité de F sont atteintes, puisque cette sphère est compacte car F est de dimension finie. Les quantités $\max\limits_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle$ et $\min\limits_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle$ sont donc bien définies.

1. Si $x \in \mathrm{E}_h$ s'écrit $x = \sum_{i=1}^{h} x_i e_i$, avec $(x_1, \ldots, x_h) \in \mathbb{R}^h$, alors on a

$$f_1(x) = \sum_{i=1}^{h} x_i \lambda_i e_i \quad \text{et} \quad \langle f_1(x), x\rangle = \sum_{i=1}^{h} x_i^2 \lambda_i \leqslant \lambda_h \sum_{i=1}^{h} x_i^2 \leqslant \lambda_h \|x\|^2.$$

Si $\|x\| = 1$, alors $\langle f(x), x\rangle \leqslant \lambda_h$. Comme de plus, $\langle f_1(e_h), e_h\rangle = \lambda_h$, on obtient $\lambda_h = \max\limits_{\substack{x\in\mathrm{E}_h\\ \|x\|=1}} \langle f_1(x), x\rangle$.

De même, pour $x \in \mathrm{F}_h$, $x = \sum_{i=h}^{n} x_i e_i$, avec $(x_h, \ldots, x_n) \in \mathbb{R}^{n-h+1}$, on

a $\langle f_2(x), x\rangle = \sum_{i=h}^{n} x_i^2 \lambda_i \geqslant \lambda_h \sum_{i=h}^{n} x_i^2 \geqslant \lambda_h$ si $\|x\| = 1$. Sachant que, de plus $\langle f_2(e_h), e_h\rangle = \lambda_h$, on conclut que $\lambda_h = \min_{\substack{x\in F_h \\ \|x\|=1}} \langle f_2(x), x\rangle$.

2. Pour $F \in \mathcal{A}_h$. F et F_h ne peuvent être en somme directe puisque $\dim F + \dim F_h > n$. Ainsi, on a $F \cap F_h \neq \emptyset$. Si x est un élément de $F \cap F_h$ de norme 1, on a d'après la question 1, $\langle f(x), x\rangle \geqslant \lambda_h$. On en déduit que $\max_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \geqslant \lambda_h$ et cela étant vrai pour tout $F \in \mathcal{A}_h$,

$$\inf_{F\in\mathcal{A}_h} \left[\max_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \right] \geqslant \lambda_h.$$

Sachant que, de plus, $\max_{\substack{x\in E_h \\ \|x\|=1}} \langle f(x), x\rangle = \lambda_h$, on obtient

$$\min_{F\in\mathcal{A}_h} \left[\max_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \right] = \lambda_h.$$

De même, pour $F \in \mathcal{A}_{n-h+1}$, on montre que $\dim(F \cap E_h) \geqslant 1$; si x est un élément de $F \cap E_h$ de norme 1, alors $\langle f(x), x\rangle \leqslant \lambda_h$ et donc $\max_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \leqslant \lambda_h$. Ceci étant vrai pour tout $F \in \mathcal{A}_{n-h+1}$, on en déduit $\sup_{F\in\mathcal{A}_{n-h+1}} \left[\min_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \right] \leqslant \lambda_h$. Sachant que $\min_{\substack{x\in F_h \\ \|x\|=1}} \langle f(x), x\rangle = \lambda_h$, on obtient $\max_{F\in\mathcal{A}_{n-h+1}} \left[\min_{\substack{x\in F \\ \|x\|=1}} \langle f(x), x\rangle \right] = \lambda_h$. ◁

Parmi les questions proposées à l'oral à la suite de ce résultat en voici une qui est immédiate : si a et b sont deux endomorphismes auto-adjoints de spectre $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ pour a (resp. $\mu_1 \leqslant \cdots \leqslant \mu_n$ pour b) et si pour tout $x \in E$, $\langle a(x), x\rangle \leqslant \langle b(x), x\rangle$ alors $\lambda_k \leqslant \mu_k$ pour tout $1 \leqslant k \leqslant n$.

L'application donnée par l'exercice suivant est plus intéressante.

2.38. Théorème d'entrelacement de Cauchy

Soit E un espace euclidien de dimension n, f un endomorphisme auto-adjoint de E, G un hyperplan de E, p la projection orthogonale sur G et g l'endomorphisme induit sur G par $p \circ f$.

1. Montrer que g est auto-adjoint.

2. Soit $\lambda'_1 \leqslant \cdots \leqslant \lambda'_{n-1}$ les valeurs propres de g. Montrer que

$$\lambda_1 \leqslant \lambda'_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \lambda'_{n-1} \leqslant \lambda_n.$$

(École polytechnique, École normale supérieure)

▷ **Solution.**

1. On a, pour $(x, y) \in \mathrm{G}^2$,

$$\langle g(x), y\rangle = \langle p \circ f(x), y\rangle = \langle f(x), p(y)\rangle = \langle f(x), y\rangle,$$

puisque p est auto-adjoint (c'est un projecteur orthogonal), et de même $\langle x, g(y)\rangle = \langle x, f(y)\rangle$. Sachant que $\langle f(x), y\rangle = \langle x, f(y)\rangle$ puisque f est auto-adjoint, on en déduit que $\langle g(x), y\rangle = \langle x, g(y)\rangle$: g est donc auto-adjoint.

2. Pour $1 \leqslant h \leqslant n-1$, notons $\mathcal{A}_h$ (resp. $\mathcal{A}'_h$) l'ensemble des sous-espaces de E (resp. de G) de dimension h. Pour tout $\mathrm{F} \in \mathcal{A}'_h$, on a

$$\max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle g(x), x\rangle = \max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle f(x), x\rangle.$$

En appliquant le théorème du minimax à l'endomorphisme auto-adjoint g de G, on obtient, pour $1 \leqslant h \leqslant n-1$,

$$\lambda'_h = \min_{\mathrm{F} \in \mathcal{A}'_h} \left[\max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle g(x), x\rangle \right] = \min_{\mathrm{F} \in \mathcal{A}'_h} \left[\max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle f(x), x\rangle \right].$$

De l'inclusion $\mathcal{A}'_h \subset \mathcal{A}_h$, on déduit

$$\min_{\mathrm{F} \in \mathcal{A}'_h} \left[\max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle f(x), x\rangle \right] \geqslant \min_{\mathrm{F} \in \mathcal{A}_h} \left[\max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle f(x), x\rangle \right],$$

c'est-à-dire $\lambda'_h \geqslant \lambda_h$.

On obtient de même,

$$\lambda'_h = \max_{\mathrm{F} \in \mathcal{A}'_{n-h}} \left[\min_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle f(x), x\rangle \right]$$

et puisque $\mathcal{A}'_{n-h} \subset \mathcal{A}_{n-h}$,

$$\max_{\mathrm{F}\in\mathcal{A}'_{n-h}} \left[\min_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle \right] \leqslant \max_{\mathrm{F}\in\mathcal{A}_{n-h}} \left[\min_{\substack{x\in\mathrm{F}\\ \|x\|=1}} \langle f(x), x\rangle \right],$$

c'est-à-dire $\lambda'_h \leqslant \lambda_{h+1}$. On a finalement, pour $1 \leqslant h \leqslant n-1$,

$$\lambda_h \leqslant \lambda'_h \leqslant \lambda_{h+1} \quad \triangleleft.$$

Cet exercice est souvent posé de manière matricielle : si A *est une matrice symétrique réelle de taille* n *et si* B *est la sous-matrice de* A *obtenue en ôtant la ligne et la colonne d'indice* k *fixé, alors on a*

$$\lambda_1 \leqslant \mu_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \mu_{n-1} \leqslant \lambda_n$$

où $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ *sont les valeurs propres de* A *et* $\mu_1 \leqslant \cdots \leqslant \mu_{n-1}$ *sont celles de* B. *Il s'agit évidemment du résultat de l'exercice appliqué à l'endomorphisme* f *de* $\mathbb{R}^n$ *canoniquement associé à* A *et en prenant pour* G *l'hyperplan* $\mathrm{Vect}(e_i)_{i\neq k}$ *où* $(e_1, \ldots, e_n)$ *est la base canonique de* $\mathbb{R}^n$.

On peut en déduire une preuve très courte du théorème de Sylvester (voir exercices 2.18 et 3.15) qui caractérise les matrices symétriques définies positives par la stricte positivité de tous les mineurs principaux. Le sens direct est aisé et on prouve la réciproque par récurrence sur n *le cas* $n = 1$ *étant immédiat. Soit donc* A *symétrique de taille* n *dont les mineurs principaux sont strictement positifs. Soit* B *extraite de* A *en ôtant la dernière ligne et la dernière colonne : l'hypothèse de récurrence s'applique à* B *qui est donc définie positive. Or les valeurs propres* $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ *de* A *vérifient*

$$\lambda_1 \leqslant \mu_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_{n-1} \leqslant \mu_{n-1} \leqslant \lambda_n$$

les μ_i *étant les valeurs propres de* B *qui sont donc strictement positives. On en déduit que* $\lambda_2, \ldots, \lambda_n$ *sont strictement positifs. Mais comme le déterminant de* A *est strictement positif il en est forcément de même de la dernière valeur propre* λ_1 *et cela termine la récurrence :* A *est définie positive.*

Formulé matriciellement, l'exercice suivant apparaît comme une réciproque de l'exercice 2.38.

2.39. Une réciproque au théorème d'entrelacement de Cauchy

Soit $\lambda_1, \dots, \lambda_n, \mu_1, \dots, \mu_{n-1}$ des réels tels que

$$\lambda_1 < \mu_1 < \lambda_2 < \mu_2 \dots < \lambda_{n-1} < \mu_{n-1} < \lambda_n.$$

Prouver l'existence d'une matrice symétrique d'ordre n de bloc supérieur gauche $\mathrm{Diag}(\mu_1, \dots, \mu_{n-1})$ et de spectre $\{\lambda_1, \dots, \lambda_n\}$.

(École normale supérieure)

▷ **Solution.**

Il existe des réels $a_1, \dots, a_n$ tels que la matrice cherchée M s'écrive

$$\mathrm{M} = \begin{pmatrix} \mu_1 & 0 & \dots & 0 & a_1 \\ 0 & \mu_2 & & & a_2 \\ \vdots & & \ddots & & \vdots \\ 0 & & & \mu_{n-1} & a_{n-1} \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix}.$$

On calcule $\det(\mathrm{X}\mathrm{I}_n - \mathrm{M})$. En multipliant par $\dfrac{a_i}{\mathrm{X} - \mu_i}$ la i-ième ligne de la matrice et en l'ajoutant à la dernière, pour $1 \leqslant i \leqslant n$, on obtient une matrice triangulaire d'où l'on déduit

$$\det(\mathrm{X}\mathrm{I}_n - \mathrm{M}) = (\mathrm{X} - \mu_1) \cdots (\mathrm{X} - \mu_{n-1}) \left(\mathrm{X} - a_n - \sum_{i=1}^{n-1} \frac{a_i^2}{\mathrm{X} - \mu_i} \right).$$

On veut que la fraction rationnelle $\mathrm{F} = \mathrm{X} - a_n - \sum_{i=1}^{n-1} \dfrac{a_i^2}{\mathrm{X} - \mu_i}$ s'annule en $\lambda_1, \dots, \lambda_n$. Comme F est une fraction rationnelle de degré 1 et de coefficient dominant 1, on voit que cela est réalisé si, et seulement si, $\mathrm{F} = \dfrac{\mathrm{P}}{\mathrm{Q}}$, où $\mathrm{P} = \prod_{i=1}^{n} (\mathrm{X} - \lambda_i)$ et $\mathrm{Q} = \prod_{i=1}^{n-1} (\mathrm{X} - \mu_i)$.

Décomposons $\dfrac{\mathrm{P}}{\mathrm{Q}}$ en éléments simples. La partie entière est du premier degré avec un coefficient dominant 1. Il existe donc $(\alpha_1, \dots, \alpha_n) \in \mathbb{R}^n$ tel que $\dfrac{\mathrm{P}}{\mathrm{Q}} = \mathrm{X} + \alpha_n + \sum_{i=1}^{n-1} \dfrac{\alpha_i}{\mathrm{X} - \mu_i}$. Il faut choisir $a_1, \dots, a_n$ tel que $a_n = -\alpha_n$ et $a_i^2 = -\alpha_i$ pour $1 \leqslant i \leqslant n-1$. On a, pour $1 \leqslant j \leqslant n-1$,

$$\alpha_j = \frac{\mathrm{P}(\mu_j)}{\mathrm{Q}'(\mu_j)} = \frac{\displaystyle\prod_{i=1}^{n} (\mu_j - \lambda_i)}{\displaystyle\prod_{\substack{1 \leqslant i \leqslant n-1 \\ i \neq j}} (\mu_j - \mu_i)}.$$

Etant données les conditions sur $\lambda_1, \ldots, \lambda_n, \mu_1, \ldots, \mu_{n-1}$, on a $\alpha_j < 0$, car le numérateur de cette fraction a le signe de $(-1)^{n-j}$ (en effet, $\mu_j - \lambda_i < 0$ si $j+1 \leqslant i \leqslant n$) et son dénominateur a le signe de $(-1)^{n-1-j}$ ($\mu_j - \mu_i < 0$ si $j+1 \leqslant i \leqslant n-1$). En prenant $a_n = -\alpha_n$ et $a_j = \sqrt{-\alpha_j}$, on obtient une matrice M dont les valeurs propres sont $\lambda_1, \ldots, \lambda_n$. ◁

Le lecteur pourra généraliser le résultat au cas où les inégalités sont larges.

Voici une autre application importante du théorème du minimax.

2.40. Théorème de perturbation de Weyl

Soit A, B deux matrices symétriques réelles de taille n de spectres respectifs $\lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_n$ et $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_n$. Démontrer que pour tout $p \in [\![1, n]\!]$,

$$|\lambda_p - \mu_p| \leqslant |\!|\!|A - B|\!|\!|,$$

la triple norme étant associée à la norme euclidienne canonique de $\mathbb{R}^n$.

(École polytechnique)

Solution.

On considère une base orthonormale $(e_1, \ldots, e_n)$ (resp. $(\varepsilon_1, \ldots, \varepsilon_p)$) de $\mathbb{R}^n$ de vecteurs propres pour A (resp. pour B). Fixons $p \in [\![1, n]\!]$ et supposons par exemple $\lambda_p \geqslant \mu_p$. Si on note $F = \mathrm{Vect}(e_1, \ldots, e_p)$ et $G = \mathrm{Vect}(\varepsilon_p, \ldots, \varepsilon_n)$, on a (voir la première question de l'exercice 2.37)

$$\lambda_p = \min_{\substack{x \in G \\ \|x\| = 1}} \langle Ax, x \rangle \quad \text{et} \quad \mu_p = \max_{\substack{x \in F \\ \|x\| = 1}} \langle Bx, x \rangle.$$

Comme $\dim F + \dim G > n$, $F \cap G$ n'est pas réduit à $\{0\}$. En prenant x_0 de norme 1 dans cette intersection, il vient

$$|\lambda_p - \mu_p| = \lambda_p - \mu_p \leqslant \langle Ax_0, x_0 \rangle - \langle Bx_0, x_0 \rangle = \langle (A - B)x_0, x_0 \rangle,$$

et donc $|\lambda_p - \mu_p| \leqslant \|(A - B)x_0\| . \|x_0\| \leqslant |\!|\!|A - B|\!|\!|$ d'après l'inégalité de Cauchy-Schwarz et la définition de la norme triple. ◁

On en déduit que l'application $A \longmapsto \lambda_p$ qui à une matrice symétrique réelle associe sa p-ième valeur propre (par ordre croissant) est 1-lipschitzienne donc continue.

La première question de l'énoncé suivant étudie les liens qu'il y a entre la diagonale et le spectre d'une matrice symétrique. Nous en donnons ici

une solution utilisant le théorème du minimax. Le lecteur en trouvera une autre preuve basée sur le principe du minimax de Ky-Fan dans l'exercice 2.42.

2.41. Théorème de majoration de Schur

1. Soit E un espace euclidien de dimension n, $(e_1, \dots, e_n)$ une base orthonormale de E, u un endomorphisme auto-adjoint de E, $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ les valeurs propres de u. On note $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ la matrice de u dans la base $(e_1, \dots, e_n)$. Démontrer que pour $1 \leqslant p \leqslant n$, on a

$$\lambda_1 + \lambda_2 + \cdots + \lambda_p \leqslant a_{11} + a_{22} + \cdots + a_{pp}.$$

2. Soit a et b deux endomorphismes auto-adjoints d'un espace euclidien E. Démontrer que $\mathrm{Tr}(e^b(b-a) - e^b + e^a) \geqslant 0$.

(École normale supérieure)

▷ **Solution.**

1. Démontrons le résultat par récurrence sur p. Pour $p = 1$, on sait que

$$\lambda_1 = \inf_{\|x\|=1} \langle u(x), x \rangle \leqslant \langle u(e_1), e_1 \rangle = a_{11}.$$

Supposons $p \geqslant 2$. On note $\mathrm{F} = \mathrm{Vect}(e_1, \dots, e_p)$, $\widetilde{u}$ l'endomorphisme auto-adjoint de F dont la matrice dans $(e_1, \dots, e_p)$ est $(a_{ij})_{1 \leqslant i,j \leqslant p}$. D'après le théorème du minimax, on a $\lambda_p \leqslant \max_{\substack{x \in \mathrm{F} \\ \|x\|=1}} \langle u(x), x \rangle$. Il existe donc un vecteur $\varepsilon_p \in \mathrm{F}$ de norme 1 tel que $\lambda_p \leqslant \langle u(\varepsilon_p), \varepsilon_p \rangle$. On choisit $\varepsilon_1, \dots, \varepsilon_{p-1}$ de telle manière que $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p)$ soit une base orthonormée de F. Le système $\mathcal{B}' = (\varepsilon_1, \dots, \varepsilon_p, e_{p+1}, \dots, e_n)$ est une base orthonormale de E. Si $(a'_{ij})_{1 \leqslant i,j \leqslant n}$ désigne la matrice de u dans cette base, l'hypothèse de récurrence nous donne $\lambda_1 + \cdots + \lambda_{p-1} \leqslant a'_{11} + \cdots + a'_{p-1,p-1}$. Comme $\lambda_p \leqslant \langle u(\varepsilon_p), \varepsilon_p \rangle = a'_{pp}$, on obtient

$$\lambda_1 + \cdots + \lambda_{p-1} + \lambda_p \leqslant a'_{11} + \cdots + a'_{p-1,p-1} + a'_{pp} = \mathrm{Tr}(\widetilde{u}).$$

Comme on a également $\mathrm{Tr}(\widetilde{u}) = a_{11} + \cdots + a_{p-1,p-1} + a_{pp}$, on peut conclure

$$\lambda_1 + \cdots + \lambda_p \leqslant a_{11} + \cdots + a_{pp}.$$

2. Pour tout $(x, y) \in \mathbb{R}^2$, on a

$$e^y(y-x) - e^y + e^x = e^y(y - x - 1 + e^{x-y}) \geqslant 0,$$

car pour tout réel t, $e^t - 1 - t \geqslant 0$.

On prend $(e_1, \dots, e_n)$ une base orthonormale diagonalisant b ; on note $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$ et $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_n$ les valeurs propres de a et b respectivement et $(a_{ij})_{1 \leqslant i,j \leqslant n}$ la matrice de a dans la base $(e_1, \dots, e_n)$. On a alors

$$\mathrm{Tr}(e^b(b-a)) = \sum_{i=1}^{n} e^{\mu_i}(\mu_i - a_{ii}) \quad \text{et} \quad \mathrm{Tr}(e^b - e^a) = \sum_{i=1}^{n} (e^{\mu_i} - e^{\lambda_i})$$

et donc

$$\begin{aligned} \mathrm{Tr}(e^b(b-a) - e^b + e^a) &= \sum_{i=1}^{n} (e^{\mu_i}(\mu_i - a_{ii}) - e^{\mu_i} + e^{\lambda_i}) \\ &= \underbrace{\sum_{i=1}^{n} (e^{\mu_i}(\mu_i - \lambda_i) - e^{\mu_i} + e^{\lambda_i})}_{\geqslant 0} + \sum_{i=1}^{n} e^{\mu_i}(\lambda_i - a_{ii}). \end{aligned}$$

Il suffit donc de montrer que $\sum_{i=1}^{n} e^{\mu_i}(\lambda_i - a_{ii}) \geqslant 0$. Pour ce faire, on va appliquer une transformation d'Abel en faisant intervenir, pour $1 \leqslant i \leqslant n$, la quantité $\mathrm{D}_i = (a_{11} + \cdots + a_{ii}) - (\lambda_1 + \cdots + \lambda_i)$. D'après la première question, $\mathrm{D}_i \geqslant 0$ et d'autre part, $\mathrm{D}_n = \mathrm{Tr}(a) - \mathrm{Tr}(a) = 0$. On peut écrire, avec la convention $\mathrm{D}_0 = 0$,

$$\begin{aligned} \sum_{i=1}^{n} e^{\mu_i}(\lambda_i - a_{ii}) &= \sum_{i=1}^{n} e^{\mu_i}(\mathrm{D}_{i-1} - \mathrm{D}_i) = \sum_{i=0}^{n-1} e^{\mu_{i+1}} \mathrm{D}_i - \sum_{i=1}^{n} e^{\mu_i} \mathrm{D}_i \\ &= -e^{\mu_n} \mathrm{D}_n + \sum_{i=1}^{n-1} (e^{\mu_{i+1}} - e^{\mu_i}) \mathrm{D}_i \\ &= \sum_{i=1}^{n-1} (e^{\mu_{i+1}} - e^{\mu_i}) \mathrm{D}_i \geqslant 0, \end{aligned}$$

car les différences $e^{\mu_{i+1}} - e^{\mu_i}$ et les D_i sont positifs. L'inégalité est prouvée. ◁

La première question a montré le théorème de majoration de Schur. Une autre manière de le prouver consiste à établir que

$$\lambda_1 + \cdots + \lambda_p = \inf_{(x_1, \dots, x_p)} \sum_{i=1}^{p} (u(x_i), x_i)$$

où la borne inférieure porte sur toutes les familles orthonormales de p vecteurs (voir 2.42). Le résultat en découle directement en prenant la famille $(e_1, \dots, e_p)$.

En fait le théorème de Schur élucide complètement les liens qu'il y a entre la diagonale et le spectre d'une matrice symétrique : si on se donne $\lambda_1 \leqslant \cdots \leqslant \lambda_n$ *et* $a_1 \leqslant \cdots \leqslant a_n$ *deux* n*-uplets de réels vérifiant* $\lambda_1 + \cdots + \lambda_p \leqslant a_1 + \cdots + a_p$ *pour tout* p *et* $\sum_{i=1}^{n} \lambda_i = \sum_{i=1}^{n} a_i$ *alors il existe une matrice symétrique* A *ayant* $a_1, \ldots, a_n$ *comme coefficients diagonaux et* $\lambda_1, \ldots, \lambda_n$ *pour valeurs propres.*

2.42. Inégalités de Ky-Fan

Soit A et B deux matrices symétriques réelles de taille n. On range leurs valeurs propres par ordre décroissant

$$\lambda_1(\mathrm{A}) \geqslant \cdots \geqslant \lambda_n(\mathrm{A}).$$

1. Démontrer que $\lambda_1(\mathrm{A}+\mathrm{B}) \leqslant \lambda_1(\mathrm{A}) + \lambda_1(\mathrm{B})$.

2. Montrer que, pour tout $k \in [\![1, n]\!]$,

$$\sum_{i=1}^{k} \lambda_i(\mathrm{A}+\mathrm{B}) \leqslant \sum_{i=1}^{k} (\lambda_i(\mathrm{A}) + \lambda_i(\mathrm{B})).$$

(**École normale supérieure, École polytechnique**)

▷ **Solution.**

1. Soit $(e_1, \ldots, e_n)$ une base orthonormale de $\mathbb{R}^n$ composée de vecteurs propres de A, e_i étant relatif à la valeur propre $\lambda_i(\mathrm{A})$. Pour tout vecteur $\mathrm{X} = \sum_{i=1}^{n} x_i e_i$. de norme 1, on a

$$\langle \mathrm{X}, \mathrm{AX} \rangle = \sum_{i=1}^{n} \lambda_i(\mathrm{A}) x_i^2 \leqslant \lambda_1(\mathrm{A}) \sum_{i=1}^{n} x_i^2 \leqslant \lambda_1(\mathrm{A}).$$

Comme pour $\mathrm{X} = e_1$, on a égalité, on en déduit que

$$\lambda_1(\mathrm{A}) = \sup_{\|\mathrm{X}\|=1} \langle \mathrm{X}, \mathrm{AX} \rangle.$$

On a évidemment le même résultat pour B et A + B. On en déduit que, pour tout X de norme 1,

$$\langle \mathrm{X}, (\mathrm{A}+\mathrm{B})\mathrm{X} \rangle = \langle \mathrm{X}, \mathrm{AX} \rangle + \langle \mathrm{X}, \mathrm{BX} \rangle \leqslant \lambda_1(\mathrm{A}) + \lambda_1(\mathrm{B}).$$

En prenant la borne supérieure du membre de gauche, on obtient

$$\lambda_1(\mathrm{A}+\mathrm{B}) \leqslant \lambda_1(\mathrm{A}) + \lambda_1(\mathrm{B}).$$

2. On note E_k l'ensemble des familles $(\mathrm{X}_1, \ldots, \mathrm{X}_k)$ d'éléments de $\mathbb{R}^n$ de norme 1, deux à deux orthogonaux. Démontrons que

$$\sup_{(\mathrm{X}_1,\ldots,\mathrm{X}_k)\in\mathrm{E}_k} \sum_{j=1}^{k} \langle \mathrm{X}_j, \mathrm{A}\mathrm{X}_j\rangle = \sum_{i=1}^{k} \lambda_i(\mathrm{A}).$$

Soit $(\mathrm{X}_1, \ldots, \mathrm{X}_k) \in \mathrm{E}_k$ et $\mathrm{M} = (x_{ij}) \in \mathcal{M}_{n,k}(\mathbb{R})$ la matrice des coordonnées des vecteurs $\mathrm{X}_1, \ldots, \mathrm{X}_k$ dans la base $(e_1, \ldots, e_n)$. On a

$$\sum_{j=1}^{k} \langle \mathrm{X}_j, \mathrm{A}\mathrm{X}_j\rangle = \sum_{j=1}^{k}\sum_{i=1}^{n} \lambda_i(\mathrm{A}) x_{ij}^2 = \sum_{i=1}^{n} \lambda_i(\mathrm{A}) \sum_{j=1}^{k} x_{ij}^2.$$

On note $s_i = \sum_{j=1}^{k} x_{ij}^2$. On majore les valeurs propres $\lambda_i(\mathrm{A})$ avec $i \geqslant k$ par $\lambda_k(\mathrm{A})$. On obtient

$$\sum_{j=1}^{k} \langle \mathrm{X}_j, \mathrm{A}\mathrm{X}_j\rangle \leqslant \sum_{i=1}^{k-1} \lambda_i(\mathrm{A}) s_i + \lambda_k(\mathrm{A}) \sum_{i=k}^{n} s_i.$$

Comme $\sum_{i=1}^{n} s_i = \sum_{j=1}^{k}\sum_{i=1}^{n} x_{ij}^2 = k$, car les vecteurs X_j sont unitaires, on obtient

$$\begin{aligned}\sum_{j=1}^{k} (\mathrm{X}_j, \mathrm{A}\mathrm{X}_j\rangle &\leqslant \sum_{i=1}^{k-1} \lambda_i(\mathrm{A}) s_i + \lambda_k(\mathrm{A}) \left(k - \sum_{i=1}^{k-1} s_i\right) \\ &\leqslant \sum_{i=1}^{k-1} (\lambda_i(\mathrm{A}) - \lambda_k(\mathrm{A})) s_i + k\lambda_k(\mathrm{A}).\end{aligned}$$

La famille $(\mathrm{X}_1, \ldots, \mathrm{X}_k)$ étant orthonormale, on peut obtenir une matrice N orthogonale en ajoutant $n - k$ colonnes à la matrice M. On en déduit que $s_i \leqslant \sum_{j=1}^{n} n_{ij}^2 = 1$, d'où l'on tire

$$\sum_{j=1}^{k} \langle \mathrm{X}_j, \mathrm{A}\mathrm{X}_j\rangle \leqslant \sum_{i=1}^{k-1} (\lambda_i(\mathrm{A}) - \lambda_k(\mathrm{A})) + k\lambda_k(\mathrm{A}) \leqslant \sum_{i=1}^{k} \lambda_i(\mathrm{A}).$$

Comme il y a égalité pour $(\mathrm{X}_1, \ldots, \mathrm{X}_k) = (e_1, \ldots, e_k) \in \mathrm{E}_k$, on a le résultat annoncé.

On en déduit que, pour tout $(\mathrm{X}_1, \ldots, \mathrm{X}_k) \in \mathrm{E}_k$,

$$\sum_{j=1}^{k}\langle X_j,(A+B)X_j\rangle = \sum_{j=1}^{k}\langle X_j,AX_j\rangle + \sum_{j=1}^{k}\langle X_j,BX_j\rangle$$
$$\leqslant \sum_{i=1}^{k}(\lambda_i(A)+\lambda_i(B)).$$

En prenant la borne supérieure du membre de gauche, on obtient

$$\sum_{i=1}^{k}\lambda_i(A+B) \leqslant \sum_{i=1}^{k}(\lambda_i(A)+\lambda_i(B)). \quad \triangleleft$$

Un théorème de Ludskii généralise l'inégalité de la question 2 à une somme de valeurs propres d'indices $i_1 < i_2 < \cdots < i_k$ quelconques (au lieu des entiers de 1 à k).

Voici encore un exercice qui se ramène à démontrer le théorème de majoration de Schur.

2.43. Majoration de Tr(AB)

Soit A et B deux matrices symétriques réelles positives de $\mathcal{M}_n(\mathbb{R})$. On ordonne leurs valeurs propres : $0 \leqslant \lambda_1 \leqslant \cdots \leqslant \lambda_n$ pour A et $0 \leqslant \mu_1 \leqslant \cdots \leqslant \mu_n$ pour B. Montrer que

$$\mathrm{Tr}(AB) \leqslant \lambda_1\mu_1 + \cdots + \lambda_n\mu_n.$$

(École polytechnique)

▷ **Solution.**

On se ramène au cas où A est diagonale en écrivant $A = {}^tPDP$, avec $P \in O_n(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1,\ldots,\lambda_n)$. On a alors

$$\mathrm{Tr}(AB) = \mathrm{Tr}({}^tPDPB) = \mathrm{Tr}(DPB\,{}^tP) = \mathrm{Tr}(DB') = \sum_{i=1}^{n}\lambda_i b_{ii},$$

où $B' = PB\,{}^tP = (b_{ij})_{1\leqslant i,j\leqslant n}$. Pour utiliser la croissance de la suite $(\lambda_i)_{1\leqslant i\leqslant n}$, on fait une transformation d'Abel. On pose, pour $k \in [\![1,n]\!]$, $s_k = \sum_{i=k}^{n} b_{ii}$. L'expression de Tr(AB) devient

$$\mathrm{Tr}(AB) = \sum_{i=1}^{n-1}\lambda_i(s_i - s_{i+1}) + \lambda_n s_n = \sum_{i=1}^{n}(\lambda_i - \lambda_{i-1})s_i,$$

en posant $\lambda_0 = 0$. On obtient de même $\sum_{i=1}^{n} \lambda_i \mu_i = \sum_{i=1}^{n} (\lambda_i - \lambda_{i-1}) t_i$, en posant, pour tout $k \in [\![1, n]\!]$, $t_k = \sum_{i=k}^{n} \mu_i$ et $\mu_0 = 0$. Puisque, pour tout i, $\lambda_i - \lambda_{i-1} \geqslant 0$, il suffit de démontrer que, pour tout $k \in [\![1, n]\!]$, on a $s_k \leqslant t_k$. C'est ce qui a été vu par deux méthodes différentes dans les deux exercices précédents. ◁

L'énoncé suivant donne une expression variationnelle de l'écart maximal entre deux valeurs propres d'une matrice symétrique. Comme dans l'exercice précédent, on utilise une transformation d'Abel. C'est une idée souvent utile lorsque l'on travaille sur une somme et que l'on souhaite exploiter une suite d'inégalités du type $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$. On fait alors apparaître les différences $\lambda_i - \lambda_{i-1}$ dont on connaît le signe.

2.44. Écart maximal entre deux valeurs propres

Soit S une matrice symétrique réelle de taille n et de valeurs propres $\lambda_1, \ldots, \lambda_n$. On note

$$E = \{(X, Y) \in \mathbb{R}^n \times \mathbb{R}^n, \ \|X\| = \|Y\| = 1, \ \langle X, Y \rangle = 0\}.$$

Démontrer que $\max_{1 \leqslant i, j \leqslant n} |\lambda_i - \lambda_j| = 2 \sup_{(X,Y) \in E} |{}^t XSY|$.

(**École polytechnique**)

Solution.

On suppose que $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$. Soit $(V_1, \ldots, V_n)$ une base orthonormale de vecteurs propres de S, V_i étant associé à la valeur propre λ_i.

Soit $(X, Y) \in E$, qu'on écrit $X = \sum_{i=1}^{n} x_i V_i$ et $Y = \sum_{i=1}^{n} y_i V_i$. On a $\langle X, Y \rangle = \sum_{i=1}^{n} x_i y_i = 0$ et ${}^t XSY = \sum_{i=1}^{n} \lambda_i x_i y_i$. Pour utiliser la décroissance de la suite $(\lambda_i)_{1 \leqslant i \leqslant n}$, on fait une transformation d'Abel. On pose, pour $1 \leqslant k \leqslant n$, $s_k = \sum_{i=1}^{k} x_i y_i$ et $s_0 = 0$. On a alors

$$ {}^t XSY = \sum_{i=1}^{n} \lambda_i (s_i - s_{i-1}) = \sum_{i=1}^{n-1} (\lambda_i - \lambda_{i+1}) s_i,$$

car $s_n = 0$. On en déduit que

$$2|{}^t\mathrm{XSY}| \leqslant 2\sum_{i=1}^{n-1}(\lambda_i - \lambda_{i+1})|s_i| \leqslant 2(\lambda_1 - \lambda_n)\max(|s_1|, \ldots, |s_{n-1}|).$$

Pour obtenir $2\sup_{(\mathrm{X},\mathrm{Y})\in\mathrm{E}} |{}^t\mathrm{XSY}| \leqslant (\lambda_1 - \lambda_n)$, il suffit de démontrer que $2|s_i| \leqslant 1$ pour $1 \leqslant i \leqslant n-1$. On utilise $s_n = 0$. On en déduit que, pour $1 \leqslant i \leqslant n-1$, $s_i = \sum_{j=1}^{i} x_j y_j = -\sum_{j=i+1}^{n} x_j y_j$ et donc

$$2s_i = \sum_{j=1}^{i} x_j y_j + \sum_{j=i+1}^{n} x_j(-y_j).$$

En appliquant l'inégalité de Cauchy-Schwarz aux vecteurs X et Z, où $\mathrm{Z} = \sum_{j=1}^{i} y_j \mathrm{V}_j - \sum_{j=i+1}^{n} y_j \mathrm{V}_j$, on obtient

$$2|s_i| \leqslant \|\mathrm{X}\|\|\mathrm{Z}\| \leqslant \|\mathrm{X}\|\|\mathrm{Y}\| \leqslant 1,$$

ce qui $2\sup_{(\mathrm{X},\mathrm{Y})\in\mathrm{E}} |{}^t\mathrm{XSY}| \leqslant \lambda_1 - \lambda_n$. On prend $\mathrm{X} = \frac{1}{\sqrt{2}}(\mathrm{V}_1 + \mathrm{V}_n)$ et $\mathrm{Y} = \frac{1}{\sqrt{2}}(\mathrm{V}_1 - \mathrm{V}_n)$. On vérifie que $(\mathrm{X}, \mathrm{Y}) \in \mathrm{E}$ et que $2\,{}^t\mathrm{XSY} = \lambda_1 - \lambda_n$, d'où l'égalité souhaitée. ◁

Le dernier exercice sur ce thème très riche porte encore sur l'étude des valeurs propres d'une somme d'endomorphismes (ou de matrices réelles) symétriques.

2.45. Valeurs propres d'endomorphismes dont la somme est positive

Soit V un espace euclidien de dimension r et E l'espace vectoriel des endomorphismes symétriques de V. Pour tout $a \in \mathrm{E}$, on note $\lambda_1(a) \geqslant \lambda_2(a) \geqslant \cdots \geqslant \lambda_r(a)$ ses valeurs propres.

1. Soit $(a_1, a_2, \ldots, a_k) \in \mathrm{E}^k$ tel que $a_1 + \cdots + a_k$ soit positif. Montrer que

$$(i_1 + i_2 + \cdots + i_k < k + r) \Longrightarrow (\lambda_{i_1}(a_1) + \cdots + \lambda_{i_k}(a_k) \geqslant 0).$$

Considérer pour $j \in [\![1, k]\!]$ le sous-espace V_j engendré par les vecteurs propres de a_j associés aux valeurs propres $\lambda_{i_j}(a_j)$, $\lambda_{i_j+1}(a_j), \ldots, \lambda_r(a_j)$.

2. Soit $(a, b) \in \mathrm{E}^2$ tel que b soit positif. Montrer que

$$\forall i \in [\![1, r]\!] \quad \lambda_i(a) \leqslant \lambda_i(a + b).$$

(École polytechnique)

Solution.

1. Soit $j \in [\![1,k]\!]$. Considérons une base orthonormale $(e_1, e_2, \dots, e_r)$ de V, constituée de vecteurs propres de a_j tels que e_i soit associé à la valeur propre $\lambda_i(a_j)$ et posons $V_j = \mathrm{Vect}(e_{i_j}, e_{i_j+1}, \dots, e_r)$. On a donc $\dim V_j = r - i_j + 1$. D'autre part, par définition, les valeurs propres de la restriction de a_j à V_j sont inférieures ou égales à $\lambda_{i_j}(a_j)$. On en déduit que, pour tout $x \in V_j$, $\langle a_j(x), x\rangle \leqslant \lambda_{i_j}(a_j)\|x\|^2$.

Si un vecteur x non nul appartient à $\bigcap_{j=1}^{k} V_j$, on a

$$\sum_{j=1}^{k} \lambda_{i_j}(a_j)\|x\|^2 \geqslant \sum_{j=1}^{k} \langle a_j(x), x\rangle \geqslant \left\langle \left(\sum_{j=1}^{k} a_j\right)(x), x \right\rangle \geqslant 0,$$

car $\sum_{j=1}^{k} a_j$ est positif. On en déduit que $\sum_{j=1}^{k} \lambda_{i_j}(a_j) \geqslant 0$.

Pour conclure, il reste à démontrer que $\bigcap_{j=1}^{k} V_j$ n'est pas réduit au vecteur nul. On sait que

$$\dim(V_1 \cap V_2) = \dim V_1 + \dim V_2 - \dim(V_1 + V_2) \geqslant \dim V_1 + \dim V_2 - r.$$

On montre facilement, par récurrence sur k, que

$$\dim\left(\bigcap_{j=1}^{k} V_j\right) \geqslant \sum_{j=1}^{k} \dim V_j - (k-1)r.$$

Compte tenu de la valeur de $\dim V_j$, on obtient

$$\dim\left(\bigcap_{j=1}^{k} V_j\right) \geqslant \sum_{j=1}^{k} (r - i_j + 1) - (k-1)r \geqslant r + k - \sum_{j=1}^{k} i_j > 0.$$

2. Posons $a_1 = -a$ et $a_2 = a + b$. On a donc $a_1 + a_2 \geqslant 0$. Comme $\lambda_r(a) \geqslant \cdots \geqslant -\lambda_2(a) \geqslant -\lambda_1(a)$ sont les valeurs propres de $-a$, on a pour $i \in [\![1,r]\!]$, $-\lambda_i(a) = \lambda_{r+1-i}(-a) = \lambda_{r+1-i}(a_1)$. On peut donc appliquer la question 1 avec $k = 2$, $i_1 = r + 1 - i$ et $i_2 = i$. On obtient $\lambda_{r+1-i}(-a) + \lambda_i(a+b) = -\lambda_i(a) + \lambda_i(a+b) \geqslant 0$, ce qui est le résultat voulu. ◁

La seconde question découle aussi immédiatement du théorème du minimax, puisque pour tout x on a $\langle a(x), x\rangle \leqslant \langle (a+b)(x), x\rangle$ (voir la remarque qui suit l'exercice 2.37).

La série d'exercices à venir porte sur les matrices antisymétriques et les endomorphismes antisymétriques d'un espace euclidien E. *On dit qu'un endomorphisme* u *de* E *est antisymétrique si* $u^* = -u$ *c'est-à-dire si l'on a, pour tout* $(x, y) \in \mathrm{E}^2$, $\langle u(x), y\rangle = -\langle x, u(y)\rangle$. *On vérifie aisément que* u *est antisymétrique si, et seulement si sa matrice dans une base orthonormale est antisymétrique.*

2.46. Étude de $\mathrm{Tr}(\mathrm{AB} - \mathrm{BA})^4$, pour A et B antisymétriques

Soit A et B deux matrices antisymétriques réelles. Montrer que $\mathrm{Tr}(\mathrm{AB} - \mathrm{BA})^4 \geqslant 0$. À quelle condition cette trace est-elle nulle ?

(École polytechnique)

▷ **Solution.**

Posons $\mathrm{C} = \mathrm{AB} - \mathrm{BA}$. On a

$$^t\mathrm{C} = {}^t\mathrm{B}\,{}^t\mathrm{A} - {}^t\mathrm{A}\,{}^t\mathrm{B} = (-\mathrm{B})(-\mathrm{A}) - (-\mathrm{A})(-\mathrm{B}) = -\mathrm{C}.$$

La matrice C est antisymétrique, donc $\mathrm{D} = \mathrm{C}^2 = (d_{ij})_{1\leqslant i,j\leqslant n}$ est symétrique. On en déduit que

$$\mathrm{Tr}\,\mathrm{C}^4 = \mathrm{Tr}\,\mathrm{D}^2 = \sum_{1\leqslant i,j\leqslant n} d_{ij}d_{ji} = \sum_{1\leqslant i,j\leqslant n} d_{ij}^2 \geqslant 0.$$

Si $\mathrm{Tr}\,\mathrm{C}^4 = 0$, on a, pour tout $(i, j) \in [\![1, n]\!]^2$, $d_{ij} = 0$. Mais par définition,

$$d_{ii} = \sum_{k=1}^{n} c_{ik}c_{ki} = -\sum_{k=1}^{n} c_{ik}^2.$$

On a donc $c_{ik} = 0$, pour tout couple (i, k), $\mathrm{C} = 0$ et donc $\mathrm{AB} = \mathrm{BA}$. ◁

L'exercice suivant donne un résultat élémentaire très classique : il s'agit de constater que l'endomorphisme u *est antisymétrique dès que, pour tout vecteur* x, $\langle u(x), x\rangle = 0$ *et qu'alors son rang est pair.*

2.47. Parité du rang d'un endomorphisme antisymétrique

Soit E un espace euclidien de dimension $n \geqslant 1$ et u un endomorphisme de E tel que, pour tout $x \in \mathrm{E}$, $\langle u(x), x\rangle = 0$.

Montrer que $\mathrm{E} = \mathrm{Im}\,u \overset{\perp}{\oplus} \mathrm{Ker}\,u$ et que le rang de u est pair.

(École polytechnique)

Solution.

On déduit de l'hypothèse que, pour tout couple $(x, y) \in \mathrm{E}^2$, on a

$$\begin{aligned} 0 &= \langle u(x+y), x+y\rangle = (u(x), x\rangle + \langle u(y), y\rangle + (u(x), y\rangle + (u(y), x\rangle \\ &= \langle u(x), y\rangle + (u(y), x\rangle, \end{aligned}$$

et donc $(u(y), x\rangle = -(y, u(x)\rangle$: u est antisymétrique, *i.e.* $u^* = -u$.

Si $x \in \operatorname{Ker} u$, on a, pour tout $y \in \mathrm{E}$, $(u(y), x\rangle = -(y, u(x)\rangle = 0$ et $x \in (\operatorname{Im} u)^\perp$. Ainsi $\operatorname{Ker} u \subset (\operatorname{Im} u)^\perp$ et par égalité des dimensions, $\operatorname{Ker} u = (\operatorname{Im} u)^\perp$. Soit v la restriction de u à $\operatorname{Im} u$, considérée comme à valeurs dans $\operatorname{Im} u$. Comme u, l'endomorphisme v est antisymétrique. De plus, il est injectif, puisque $\operatorname{Im} u \cap \operatorname{Ker} u = \{0\}$, et il est donc bijectif. En conséquence, on a

$$\det v^* = \det v = \det(-v^*) = (-1)^r \det(v^*).$$

où r est le rang de u. Comme $\det v^* \neq 0$, on obtient $(-1)^r = 1$ et le rang de u est pair. ◁

L'exercice suivant fait observer qu'en dimension 3 *les endomorphismes antisymétriques se décrivent aisément à l'aide du produit vectoriel.*

2.48. Endomorphismes antisymétriques en dimension 3

Soit E un espace vectoriel orienté de dimension 3 et $f \in \mathcal{L}(\mathrm{E})$. Montrer l'équivalence entre

(i) il existe un unique $a \in \mathrm{E}$ tel que, pour tout $x \in \mathrm{E}$,

$$f(x) = a \wedge x\ ;$$

(ii) pour tout $x \in \mathrm{E}$, on a $\langle f(x), x\rangle = 0$.

(École normale supérieure)

Solution.

L'implication $(i) \Longrightarrow (ii)$ est claire car, pour tout $(x, y) \in \mathrm{E}^2$, le produit vectoriel $x \wedge y$ est orthogonal à x et à y. Supposons donc que (ii) est réalisé. On a alors, pour tout $(x, y) \in \mathrm{E}^2$,

$$\begin{aligned} 0 = (f(x+y), x+y\rangle &= (f(x), x\rangle + (f(y), y\rangle + \langle f(x), y\rangle + (f(y), x\rangle \\ &= \langle f(x), y\rangle + \langle x, f(y)\rangle. \end{aligned}$$

Ainsi, f est antisymétrique et sa matrice A dans une base orthonormale directe est antisymétrique. Il existe $(p, q, r) \in \mathbb{R}^3$ tel que

$$A = \begin{pmatrix} 0 & -r & q \\ r & 0 & -p \\ -q & p & 0 \end{pmatrix}.$$

Le vecteur a de coordonnées (p, q, r) convient. Si un autre vecteur b convient, on a, pour tout $x \in E$,

$$(a - b) \wedge x = a \wedge x - b \wedge x = f(x) - f(x) = 0.$$

Le vecteur $a-b$ est colinéaire à tout vecteur de E donc est nul et $a = b$. ◁

L'exercice qui suit est l'occasion d'un premier pas vers la réduction des matrices antisymétriques. Nous allons constater que le spectre d'une matrice réelle antisymétrique est composé d'imaginaires purs.

2.49. Carré d'une matrice antisymétrique

Soit S une matrice symétrique réelle. Donner une condition nécessaire et suffisante sur le spectre de S pour qu'il existe une matrice antisymétrique réelle telle que $A^2 = S$.

(École polytechnique)

▷ **Solution.**

Pour commencer, montrons que les valeurs propres dans $\mathbb{C}$ d'une matrice antisymétrique réelle A sont imaginaires pures. Soit $\lambda \in \operatorname{Sp} A$ et $X \in \mathbb{C}^n$ un vecteur propre associé. On a $AX = \lambda X$. On en déduit que ${}^t\overline{X}\,{}^tA = -{}^t\overline{X}A = \overline{\lambda}\,{}^t\overline{X}$, puis que ${}^t\overline{X}AX = -\lambda\,{}^t\overline{X}X = \overline{\lambda}\,{}^t\overline{X}X$. Puisque ${}^t\overline{X}X \neq 0$ (il s'agit du produit scalaire canonique sur $\mathbb{C}^n$), on a donc $\lambda = -\overline{\lambda}$ et λ est imaginaire pur.

Si $S = A^2$, avec A antisymétrique, les valeurs propres de S sont les carrés de celles de A. Ce sont donc des réels négatifs. De plus, les valeurs propres non réelles, c'est-à-dire non nulles, de A sont deux à deux conjuguées, c'est-à-dire opposées. On en déduit que les valeurs propres non nulles de S ont toutes une multiplicité paire.

Supposons réciproquement que les valeurs propres de S soient négatives ou nulles et que les valeurs propres non nulles aient une multiplicité paire. Il existe alors un entier $r \leqslant \dfrac{n}{2}$ et des réels non nuls $\alpha_1, \ldots, \alpha_r$ tels que les valeurs propres de S comptées avec multiplicité soient $-\alpha_1^2, -\alpha_1^2, \ldots, -\alpha_r^2, -\alpha_r^2, 0, \ldots, 0$. Il existe $P \in O_n(\mathbb{R})$ et $D = \operatorname{Diag}(-\alpha_1^2, -\alpha_1^2, \ldots, -\alpha_r^2, -\alpha_r^2, 0, \ldots, 0)$ telles que $S = {}^tPDP$. On remarque que, pour tout $\alpha \in \mathbb{R}$, si l'on pose $R_\alpha = \begin{pmatrix} 0 & -\alpha \\ \alpha & 0 \end{pmatrix}$, on

obtient $R_\alpha^2 = -\alpha^2 I_2$. On considère la matrice diagonale par blocs $\Delta = \begin{pmatrix} R_{\alpha_1} & & & & \\ & \ddots & & & \\ & & R_{\alpha_r} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0 \end{pmatrix}$, qui est clairement antisymétrique et vérifie $\Delta^2 = D$. On pose $A = {}^tP\Delta P$; la matrice A est encore antisymétrique et $A^2 = S$. La matrice S est le carré d'une matrice antisymétrique. ◁

Nous terminons ce thème en donnant un résultat complet sur la réduction des endomorphismes antisymétriques.

2.50. Réduction des endomorphismes antisymétriques

Soit f un endomorphisme antisymétrique bijectif d'un espace euclidien E.

1. Montrer que $\dim E$ est un entier pair. On le note $2p$.

2. Montrer qu'il existe une base orthonormale $\mathcal{B} = (e_1, \ldots, e_{2p})$ de E et des réels $a_1, \ldots, a_p$ tels que, pour $i \in [\![1, p]\!]$,

$$f(e_{2i}) = -a_i e_{2i-1} \quad \text{et} \quad f(e_{2i-1}) = a_i e_{2i}.$$

(École polytechnique)

Solution.

1. On a

$$\det f = \det f^* = \det(-f) = (-1)^{\dim E} \det f.$$

Comme $\det f \neq 0$, on en déduit que $(-1)^{\dim E} = 1$ et donc que $\dim E$ est un entier pair.

2. On remarque que si la base $\mathcal{B} = (e_1, \ldots, e_{2p})$ convient les plans $\mathrm{Vect}(e_{2i-1}, e_{2i})$ doivent être stables par f pour tout $i \in [\![1, p]\!]$. On va donc montrer l'existence d'un plan stable par f qui convient et faire une récurrence sur la dimension ensuite.

Comme f est inversible, il n'a aucune valeur propre réelle : en effet, si $f(x) = \lambda x$ on a $0 = \langle f(x), x\rangle = \lambda \|x\|^2$ donc $x = 0$, car $\lambda \neq 0$ (on a d'ailleurs vu dans l'exercice précédent que les valeurs propres de f sont imaginaires pures). En revanche f^2, qui est auto-adjoint, est diagonalisable. Soit e_1 un vecteur propre unitaire de f^2 associé à une valeur propre

λ_1. Le vecteur $f(e_1)$ est orthogonal à e_1 et non nul : choisissons e_2 unitaire positivement colinéaire à $f(e_1)$. Il existe donc a_1 strictement positif tel que $f(e_1) = a_1 e_2$. Comme $f(e_2)$ est colinéaire à $f^2(e_1) = \lambda_1 e_1$ le plan $\mathrm{Vect}(e_1, e_2)$ est stable par f et la matrice de l'endomorphisme induit par f dans la base (e_1, e_2) est de la forme $\begin{pmatrix} 0 & a_2 \\ a_1 & 0 \end{pmatrix}$. Mais comme cet endomorphisme induit est antisymétrique et la base (e_1, e_2) orthonormée, on a forcément $a_2 = -a_1$. Cela règle déjà le cas $p = 1$.

On démontre alors la propriété par récurrence sur p. On suppose la propriété vraie au rang $p-1$ et on considère un espace E de dimension $2p$. On construit un plan $\mathrm{P} = \mathrm{Vect}(e_1, e_2)$ comme ci-dessus. On montre que $\mathrm{P}^\perp$ est également stable par f. On a en effet, pour $x \in \mathrm{P}$ et $y \in \mathrm{P}^\perp$,

$$\langle x, f(y)\rangle = \langle f(x), y\rangle = 0.$$

En appliquant l'hypothèse de récurrence à $\mathrm{P}^\perp$, on montre l'existence d'une base orthonormale $(e_3, \dots, e_{2p})$ de $\mathrm{P}^\perp$ et de réels $a_2, \dots, a_p$ (qu'on peut imposer strictement positifs) tels que, pour $2 \leqslant i \leqslant p$, on ait

$$f(e_{2i}) = -a_i e_{2i-1} \quad \text{et} \quad f(e_{2i-1}) = a_i e_{2i}.$$

La base $\mathcal{B} = (e_1, \dots, e_{2p})$ a les propriétés voulues. ◁

Cela clôt la question de la réduction des endomorphismes antisymétriques puisque, si f n'est pas bijectif, E *est somme directe de* $\mathrm{Ker}\, f$ *et de son orthogonal* $\mathrm{F} = \mathrm{Im}\, f$ *(voir exercice 2.47). Or, f induit sur* F *un endomorphisme antisymétrique sur lequel on peut donc appliquer ce qui précède. On en déduit qu'il existe une base orthonormale $\mathcal{B}$ de* E *telle que la matrice de f dans cette base soit de la forme*

$$\begin{pmatrix} 0 & \dots & \dots & \dots & \dots & 0 \\ \vdots & \ddots & & & & \vdots \\ \vdots & & 0 & & & \vdots \\ \vdots & & & \boxed{\mathrm{A}_1} & & \vdots \\ \vdots & & & & \ddots & 0 \\ 0 & \dots & \dots & \dots & 0 & \boxed{\mathrm{A}_p} \end{pmatrix} \quad \text{avec } \mathrm{A}_i = \begin{pmatrix} 0 & -a_i \\ a_i & 0 \end{pmatrix}.$$

Matriciellement, cela signifie que toute matrice antisymétrique réelle est orthogonalement semblable à une matrice de ce type.

Les deux exercices suivants servent de transition avec le cadre hermitien. En effet, il sera utile de se placer sur l'espace hermitien $\mathbb{C}^n$ muni de son produit hermitien canonique : pour $\mathrm{X} = (x_1, \dots, x_n)$, $\mathrm{Y} = (y_1, \dots, y_n) \in \mathbb{C}^n$, on a $\langle \mathrm{X}, \mathrm{Y}\rangle = \mathrm{X}^* \mathrm{Y} = \sum_{k=1}^{n} \overline{x_k} y_k$.

2.51. Matrices strictement positives

Une matrice réelle A est dite *strictement positive* si ${}^tA + A$ est symétrique définie positive.

1. Montrer que A est strictement positive si, et seulement si, pour tout $X \in \mathbb{R}^n$, non nul, on a ${}^tXAX > 0$.

2. Montrer que si A est strictement positive et si S est symétrique définie positive, alors les valeurs propres de SA et de AS sont de parties réelles strictement positives.

(**École polytechnique**)

Solution.

1. La matrice $A + {}^tA$ est symétrique. Soit $X \in \mathbb{R}^n$. On a

$$ {}^t({}^tXAX) = {}^tX{}^tA{}^t({}^tX) = {}^tX{}^tAX $$

et donc ${}^tX({}^tA + A)X = 2{}^tXAX$. On en déduit que

$$\begin{aligned} A \text{ strictement positive} &\iff A + {}^tA \text{ définie positive} \\ &\iff \forall X \in \mathbb{R}^n \backslash \{0\},\ {}^tX(A + {}^tA)X > 0 \\ &\iff \forall X \in \mathbb{R}^n \backslash \{0\},\ {}^tXAX > 0. \end{aligned}$$

2. Munissons $\mathbb{C}^n$ de son produit hermitien canonique $\langle\ ,\ \rangle$. Soit λ une valeur propre de SA et $X \in \mathbb{C}^n$ un vecteur propre associé. On a $SAX = \lambda X$ et donc $AX = \lambda S^{-1}X$. On calcule le produit scalaire $\langle X, AX\rangle$. On a

$$\langle X, AX\rangle = \langle X, \lambda S^{-1}X\rangle = \lambda\langle X, S^{-1}X\rangle,$$

c'est-à-dire $X^*AX = \lambda X^*S^{-1}X$. Si on pose $X = X_1 + iX_2$ avec X_1 et X_2 dans $\mathbb{R}^n$, on obtient

$$\begin{aligned} X^*S^{-1}X &= {}^t(X_1 - iX_2)S^{-1}(X_1 + iX_2) \\ &= \underbrace{{}^tX_1S^{-1}X_1 + {}^tX_2S^{-1}X_2}_{>0} + i\underbrace{({}^tX_1S^{-1}X_2 - {}^tX_2S^{-1}X_1)}_{=0} \in \mathbb{R}_+^*, \end{aligned}$$

car S^{-1} est symétrique définie positive et X_1 et X_2 ne sont pas tous les deux nuls. En prenant la partie réelle de l'égalité précédente, il vient

$$\mathrm{Re}(\lambda)X^*S^{-1}X = \mathrm{Re}(X^*AX) = {}^tX_1AX_1 + {}^tX_1AX_1 > 0,$$

d'après la première question, car X_1 et X_2 ne sont pas tous les deux nuls. On a donc $\mathrm{Re}(\lambda) > 0$.

La matrice ${}^tA$ est également strictement positive et on a de plus $\mathrm{Sp}(AS) = \mathrm{Sp}({}^t(AS)) = \mathrm{Sp}({}^tS{}^tA) = \mathrm{Sp}(S{}^tA)$. Les valeurs propres de

$S\,{}^tA$, et donc celles de AS ont une partie réelle strictement positive (on pourrait également invoquer le résultat suivant : si $(A,B) \in \mathcal{M}_n(\mathbb{C})^2$, $\mathrm{Sp}\,AB = \mathrm{Sp}\,BA$, car $\chi_{AB} = \chi_{BA}$). $\lhd$

Les matrices de partie symétrique positive ont fait l'objet de la seconde épreuve écrite du concours de l'école Polytechnique en 2006.

2.52. Théorème de Liapounov

Soit $A \in \mathcal{M}_n(\mathbb{R})$. On dit que A est *stable* si ses valeurs propres sont de partie réelle strictement négative. Soit $W \in \mathcal{S}_n^{++}(\mathbb{R})$ (le cône des matrices symétriques définies positives).

Montrer que A est stable si et seulement s'il existe $V \in \mathcal{S}_n^{++}(\mathbb{R})$ telle que ${}^tAV + VA = -W$.

(École normale supérieure)

▷ **Solution.**

• Supposons qu'il existe $V \in \mathcal{S}_n^{++}(\mathbb{R})$ telle que ${}^tAV + VA = -W$. Soit $\lambda \in \mathrm{Sp}\,A$, X un vecteur propre associé. On a $AX = \lambda X$ et donc, puisque A est réelle, $X^*\,{}^tA = \overline{\lambda}X^*$. On en déduit que

$$-X^*WX = X^*\,{}^tAVX + X^*VAX = \overline{\lambda}X^*VX + \lambda X^*VX = 2\,\mathrm{Re}(\lambda)X^*VX.$$

Si $X = X_1 + iX_2$, avec X_1 et X_2 dans $\mathbb{R}^n$, on trouve, comme dans l'exercice précédent, puisque V est symétrique réelle, $X^*VX = X_1^*VX_1 + X_2^*VX_2$ et donc $X^*VX > 0$, car V est définie positive et X_1 et X_2 ne sont pas tous les deux nuls. On a de même $X^*WX > 0$. On en déduit que $\mathrm{Re}(\lambda) < 0$. Les valeurs propres de A, ont une partie réelle strictement négative.

Les lecteurs qui connaissent la notion de matrice hermitienne pourront dire directement que V *et* W *étant symétriques réelles définies positives, elles sont hermitiennes définies positives, ce qui justifie* $X^*VX > 0$ *et* $X^*WX > 0$, *pour tout vecteur* X *de* $\mathbb{C}^n$ *non nul.*

• Supposons réciproquement que les valeurs propres de A aient une partie réelle strictement négative. Pour toute matrice symétrique réelle V, la matrice ${}^tAV + VA$ est encore symétrique. On définit donc une application linéaire φ de $\mathcal{S}_n(\mathbb{R})$ dans $\mathcal{S}_n(\mathbb{R})$ en posant $\varphi(V) = {}^tAV + VA$. Pour montrer que $-W$ a un antécédent par φ, nous allons démontrer que φ est surjective. Comme $\mathcal{S}_n(\mathbb{R})$ est de dimension finie, il suffit de démontrer qu'elle est injective.

Soit $V \in \mathrm{Ker}\,\varphi$. On a donc ${}^tAV = -VA$. Par une démonstration par récurrence évidente, on montre que, pour tout entier naturel n, on a $({}^tA)^nV = V(-A)^n$ et donc $(t\,{}^tA)^nV = V(-tA)^n$, pour tout réel t. On en déduit, en divisant par $n!$, puis sommant sur n que $e^{t\,{}^tA}V = Ve^{-tA}$ et comme $e^{-tA} = (e^{tA})^{-1}$, $e^{t\,{}^tA}Ve^{tA} = V$. Pour démontrer que $V = 0$, nous utiliserons le lemme suivant.

Lemme. *Si $A \in \mathcal{M}_n(\mathbb{R})$ est une matrice stable, on a* $\lim_{t \to +\infty} e^{tA} = 0$.

Démonstration. On utilise la décomposition de Dunford (voir l'exercice 2.30 du tome 2 d'algèbre). Il existe D diagonalisable et N nilpotente dans $\mathcal{M}_n(\mathbb{C})$ qui commutent telles que $A = D + N$. Comme elles commutent, on a, pour tout réel t, $e^{tA} = e^{tD}e^{tN}$. Si D est semblable à $\mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, alors e^{tD} est semblable à $\mathrm{Diag}(e^{\lambda_1 t}, \ldots, e^{\lambda_n t})$. Si p est l'indice de nilpotence de N, $e^{tN} = \sum_{k=0}^{p} \frac{1}{k!} t^k N^k$. Les coefficients de e^{tN} sont des polynômes en t. On en déduit que ceux de e^{tA} sont des sommes (finies) de termes de la forme $e^{t\lambda_i} t^l$. Comme $\mathrm{Re}(\lambda_i) < 0$, on a $\lim_{t \to +\infty} e^{t\lambda_i} = 0$ et par croissance comparée, $\lim_{t \to +\infty} e^{t\lambda_i} t^l = 0$. Le résultat en découle. ◇

Du lemme, on déduit que $\lim_{t \to +\infty} e^{tA} = 0$ et $\lim_{t \to +\infty} e^{t\,{}^tA} = 0$, car ${}^tA$ a les mêmes valeurs propres que A. On obtient $V = \lim_{t \to +\infty} e^{t\,{}^tA} W e^{tA} = 0$. Ainsi $\mathrm{Ker}\,\varphi = \{0\}$ et φ est bijective. Soit V l'antécédent de $-W$ par φ. Afin de démontrer que V est définie positive, nous allons donner une expression explicite de V.

Posons, pour tout réel t, $C(t) = e^{t\,{}^tA} W e^{tA}$. La fonction C est de classe C^1 et $C'(t) = {}^tA e^{t\,{}^tA} W e^{tA} + e^{t\,{}^tA} W e^{tA} A$. D'après la démonstration du lemme, les coefficients de $C'(t)$ sont des sommes de produits de fonctions polynomiales par des $e^{t\lambda_i}$, où les λ_i ont une partie réelle strictement négative. On en déduit que C' est intégrable sur $\mathbb{R}_+$. Toujours d'après le lemme, on a $\lim_{t \to +\infty} C(t) = 0$. On en déduit que

$$\begin{aligned}\int_0^{+\infty} C'(t)\mathrm{d}t &= {}^tA\left(\int_0^{+\infty} e^{t\,{}^tA} W e^{tA}\mathrm{d}t\right) + \left(\int_0^{+\infty} e^{t\,{}^tA} W e^{tA}\mathrm{d}t\right) A \\ &= -C(0) = -W.\end{aligned}$$

De l'unicité de V, on déduit $V = \int_0^{+\infty} e^{t\,{}^tA} W e^{tA}\mathrm{d}t$. On a donc, pour tout $X \in \mathbb{R}^n$,

$$\begin{aligned}{}^tXVX &= {}^tX\left(\int_0^{+\infty} e^{t\,{}^tA} W e^{tA}\mathrm{d}t\right)X = \int_0^{+\infty} {}^tXe^{t\,{}^tA} W e^{tA} X\mathrm{d}t \\ &= \int_0^{+\infty} {}^t(e^{tA}X) W (e^{tA}X)\mathrm{d}t.\end{aligned}$$

Si $X \neq 0$, alors pour tout $t \geqslant 0$, $e^{tA}X \neq 0$, car e^{tA} est inversible, donc ${}^t(e^{tA}X)W(e^{tA}X) > 0$, car W est symétrique définie positive. On en déduit que ${}^tXVX > 0$. Ainsi V est symétrique définie positive et l'équivalence est démontrée. ◁

Le fin du chapitre est consacrée à l'étude des endomorphismes hermitiens et des matrices hermitiennes, section actuellement hors programme des classes préparatoires mais qui est dans le prolongement naturel de l'étude des endomorphismes symétriques et qui intéressera les candidats aux concours de recrutement de professeurs.

Considérons E *un espace hermitien (voir page 84). Un endomorphisme* u *de* E *possède un unique endomorphisme adjoint* u^* *vérifiant pour tout* $(x,y)\in \mathrm{E}^2$, $(u(x),y)=\langle x,u^*(y)\rangle$. *On dit que* u *est hermitien ou auto-adjoint si* $u=u^*$. *Dans ces conditions, si* $\mathrm{A}=(a_{ij})_{1\leqslant i,j\leqslant n}$ *désigne la matrice de* u *dans une base orthonormale de* E, *la matrice* A^* *de coefficient* (i,j) *égal à* $\overline{a_{j,i}}$ *est la matrice de* u^* *dans la même base. La matrice* A *est dite hermitienne si* $\mathrm{A}=\mathrm{A}^*$. *Dans ces conditions,* u *est hermitien si, et seulement si,* A *est hermitienne.*

La théorie des endomorphismes hermitiens se calque sur celles des endomorphismes symétriques et le théorème spectral s'exprime de la sorte : un endomorphisme hermitien est diagonalisable en base orthonormale et son spectre est réel; une matrice hermitienne A *est unitairement semblable à une matrice diagonale réelle, autrement dit, il existe* $\mathrm{U}\in \mathrm{U}_n(\mathbb{C})$ *et* D *matrice diagonale réelle telles que*

$$\mathrm{A}=\mathrm{UDU}^{-1}=\mathrm{UDU}^*.$$

Enfin, relevons la correspondance bijective entre les formes sesquilinéaires hermitiennes de $\mathbb{C}^n$ *et les matrices hermitiennes de taille* n *: si on note* $(e_1,\dots,e_n)$ *est la base canonique de* $\mathbb{C}^n$, *à toute forme hermitienne* φ, *on fait correspondre la matrice hermitienne* $\mathrm{A}=(a_{ij})_{1\leqslant i,j\leqslant n}$ *où* $a_{ij}=\varphi(e_i,e_j)$. *Dans ces conditions, pour tout* $\mathrm{X},\mathrm{Y}\in\mathbb{C}^n$, *on a* $\varphi(\mathrm{X},\mathrm{Y})=\mathrm{X}^*\mathrm{AY}$. *Si* P *est la matrice de passage de la base canonique à une base* $\mathcal{B}$, $\mathrm{P}^*\mathrm{AP}$ *est la matrice de* φ *dans la nouvelle base* $\mathcal{B}$. *Le théorème spectral affirme donc pour toute forme hermitienne* φ *l'existence d'une base orthonormale qui est orthogonale pour* φ.

Enfin, une matrice hermitienne A *sera dite positive (resp. définie positive) si la forme sesquilinéaire hermitienne associée est positive (resp. définie positive) :*

$$\forall \mathrm{X}\in\mathbb{C}^n\backslash\{0\},\quad \mathrm{X}^*\mathrm{AX}\geqslant 0\ \ \textit{(resp. } \mathrm{X}^*\mathrm{AX}>0\textit{)}.$$

Le premier exercice permet de travailler toutes ces notions.

2.53. Condition pour que $\mathrm{Im\,B}\subset\mathrm{Im\,A}$

On suppose $\mathbb{C}^n$ muni de sa structure hermitienne usuelle.

1. Soit $\mathrm{A}\in\mathcal{M}_n(\mathbb{C})$ telle que, pour tout $x\in\mathbb{C}^n$, $(x,\mathrm{A}x\rangle=0$. Que dire de A ?

2. Soit $A \in \mathcal{M}_n(\mathbb{R})$ telle que pour tout $x \in \mathbb{C}^n$, $\langle x, Ax \rangle \in \mathbb{R}_+$. Montrer que A est hermitienne positive.

3. Soit A une matrice hermitienne. Montrer que A est positive si, et seulement si, son spectre est contenu dans $\mathbb{R}_+$.

On note H_+ l'ensemble des matrices hermitiennes positives de taille n.

4. Soit $A \in H_+$ et $x \in \mathbb{C}^n$. Montrer que

$$\langle x, Ax \rangle = 0 \Longleftrightarrow x \in \operatorname{Ker} A.$$

5. Soit A et B dans H_+. On dit que $A \geqslant B$ si $A - B \in H_+$. Montrer que

$$A \geqslant B \Longrightarrow \operatorname{Im} B \subset \operatorname{Im} A.$$

6. Soit $A \in \mathcal{M}_n(\mathbb{C})$. Montrer que $\operatorname{Im}(AA^*) = \operatorname{Im} A$.

7. Soit A et B dans $\mathcal{M}_n(\mathbb{C})$. Étudier les implications entre

(*i*) $\exists \lambda \geqslant 0 \;\; BB^* \leqslant \lambda AA^*$;

(*ii*) $\operatorname{Im} B \subset \operatorname{Im} A$;

(*iii*) $\exists C \in \mathcal{M}_n(\mathbb{C}) \;\; B = AC$.

(École polytechnique)

Solution.

1. Pour x et y dans $\mathbb{C}^n$, on a

$$0 = \langle x + y, A(x + y) \rangle = \langle x, Ay \rangle + \langle y, Ax \rangle,$$
$$0 = \langle x + iy, A(x + iy) \rangle = i\langle x, Ay \rangle - i\langle y, Ax \rangle.$$

On en déduit que $\langle y, Ax \rangle = 0$, pour tous x et y de $\mathbb{C}^n$. Cela implique $Ax = 0$ pour tout x donc $A = 0$.

2. Pour tout $x \in \mathbb{C}^n$, on a

$$\langle x, Ax \rangle = \overline{\langle x, Ax \rangle} = \langle Ax, x \rangle = \langle x, A^*x \rangle$$

et donc $\langle x, (A - A^*)x \rangle = 0$. D'après la question 1, on en déduit $A = A^*$: la matrice A est hermitienne ; elle est évidemment positive.

3. Supposons A positive. Si λ est une valeur propre de A et x un vecteur propre associé, on a

$$\langle x, Ax \rangle = \langle x, \lambda x \rangle = \lambda \|x\|^2 \geqslant 0$$

et donc $\lambda \geqslant 0$. Réciproquement si $\operatorname{Sp} A \subset \mathbb{R}_+$, comme A est diagonalisable dans une base orthonormale $(e_1, \ldots, e_n)$, les valeurs propres associées $\lambda_1, \ldots, \lambda_n$ étant positives, on a pour $x = \sum_{i=1}^{n} x_i e_i$,

$$\langle x, Ax\rangle = \sum_{i=1}^{n} \lambda_i |x_i|^2 \geqslant 0.$$

De la même façon, pour qu'une matrice hermitienne A *soit définie positive, il faut et il suffit que son spectre soit contenu dans* $\mathbb{R}_+^*$.

4. Si $x \in \operatorname{Ker} A$, il est clair que $\langle x, Ax\rangle = 0$.

Supposons réciproquement que $\langle x, Ax\rangle = 0$. Avec les notations de la question 3, on a $\langle x, Ax\rangle = \sum_{i=1}^{n} \lambda_i |x_i|^2 = 0$. Puisque tous les termes de la somme sont positifs, on a, pour tout $i \in [\![1, n]\!]$, $\lambda_i |x_i|^2 = 0$ et donc $\lambda_i x_i = 0$. On en déduit que

$$Ax = \sum_{i=1}^{n} \lambda_i x_i e_i = 0.$$

5. Soit A et B dans H_+ telles que $A \geqslant B$. On a, pour tout $x \in \mathbb{C}^n$, $\langle x, (A - B)x\rangle \geqslant 0$ et donc $\langle x, Ax\rangle \geqslant \langle x, Bx\rangle$. Ainsi, si $x \in \operatorname{Ker} A$, on a $0 \leqslant \langle x, Bx\rangle \leqslant 0$ et d'après la question précédente, $x \in \operatorname{Ker} B$, ce qui montre que $\operatorname{Ker} A \subset \operatorname{Ker} B$.

Mais on sait que, pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, on a l'égalité $(\operatorname{Ker} A)^\perp = \operatorname{Im} A^*$. En effet pour $x \in \operatorname{Ker} A$ et $y \in \mathbb{C}^n$, on obtient

$$\langle x, A^* y\rangle = \langle Ax, y\rangle = 0,$$

ce qui montre que $\operatorname{Im} A^* \subset (\operatorname{Ker} A)^\perp$ et comme

$$\operatorname{rg} A^* = \operatorname{rg} A = n - \dim(\operatorname{Ker} A) = \dim(\operatorname{Ker} A)^\perp,$$

il y a égalité : $(\operatorname{Ker} A)^\perp = \operatorname{Im} A^*$. Comme A est hermitienne, on a en fait $(\operatorname{Ker} A)^\perp = \operatorname{Im} A$. On obtient de même $(\operatorname{Ker} B)^\perp = \operatorname{Im} B$.

La relation $\operatorname{Ker} A \subset \operatorname{Ker} B$ équivaut à $(\operatorname{Ker} B)^\perp \subset (\operatorname{Ker} A)^\perp$, c'est-à-dire à $\operatorname{Im} B \subset \operatorname{Im} A$.

6. Notons que la matrice AA^* est hermitienne positive, car pour tout $x \in \mathbb{C}^n$, $\langle x, AA^* x\rangle = \langle A^* x, A^* x\rangle = \|A^* x\|^2 \geqslant 0$.

Il est clair que $\operatorname{Im} AA^* \subset \operatorname{Im} A$. Montrons que ces deux sous-espaces ont même dimension, ce qui démontrera l'égalité demandée. Il suffit de montrer que $\dim \operatorname{Ker} AA^* = \dim \operatorname{Ker} A$. Montrons pour cela que $\operatorname{Ker} AA^* = \operatorname{Ker} A^*$. Cela résulte de la question 4, car

$$x \in \operatorname{Ker} AA^* \iff \langle x, AA^* x\rangle = 0 \iff \|A^* x\| = 0 \iff x \in \operatorname{Ker} A^*.$$

Comme A et A^* ont le même rang, $\operatorname{Ker} A$ et $\operatorname{Ker} A^*$ ont la même dimension et on a le résultat.

7. Montrons que les trois propriétés sont équivalentes.

• Montrons que (i) implique (ii). Les matrices AA^*, BB^* et donc λAA^* sont hermitiennes positives. D'après la question 5, $BB^* \leqslant \lambda AA^*$ implique $\operatorname{Im} BB^* \subset \operatorname{Im} \lambda AA^* = \operatorname{Im}(\sqrt{\lambda}A)(\sqrt{\lambda}A)^*$ et d'après la question 6. $\operatorname{Im} B \subset \operatorname{Im} \sqrt{\lambda}A \subset \operatorname{Im} A$.

• L'implication entre (ii) et (iii) résulte du lemme de factorisation (cf. exercice 6.4 du tome 1 d'algèbre).

• Montrons que (iii) implique (i). Si M est hermitienne positive de plus grande valeur propre λ, on obtient en se plaçant dans une base orthonormale de vecteurs propres $(e_1, \ldots, e_n)$, pour tout $x = \sum_{i=1}^{n} x_i e_i$,

$$\langle x, Mx \rangle = \sum_{i=1}^{n} \lambda_i |x_i|^2 \leqslant \lambda \sum_{i=1}^{n} |x_i|^2 \leqslant \lambda \|x\|^2.$$

On note λ ($\lambda \geqslant 0$) la plus grande valeur propre de $CC^* \in H_+$. On obtient, pour tout x,

$$\begin{aligned} \langle x, (\lambda AA^* - BB^*)x \rangle &= \lambda \|A^*x\|^2 - \langle x, ACC^*A^*x \rangle \\ &= \lambda \|A^*x\|^2 - \langle A^*x, CC^*A^*x \rangle \geqslant 0, \end{aligned}$$

ce qui, d'après la question 2, implique $\lambda AA^* - BB^* \in H_+$ et donc par définition $BB^* \leqslant \lambda AA^*$. ◁

De l'exercice précédent, on retiendra entre autres que pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$. A^*A *est hermitienne positive et* $\operatorname{Ker} A^*A = \operatorname{Ker} A$. *L'exercice suivant reprend la relation d'ordre introduite dans l'exercice précédent.*

2.54. Application convexe

On considère l'application

$$\begin{array}{rccc} \Phi : & \mathcal{M}_n(\mathbb{C}) & \longrightarrow & \mathcal{M}_n(\mathbb{C}) \\ & A & \longmapsto & A^*A. \end{array}$$

Étant donné deux matrices hermitiennes, on pose

$$H \leqslant K \Longleftrightarrow K - H \text{ est positive.}$$

Montrer que, pour cette relation d'ordre, Φ est convexe.

(École polytechnique)

▷ **Solution.**

Notons que, pour tout $A \in \mathcal{M}_n(\mathbb{C})$, la matrice A^*A est hermitienne et positive car, pour tout $X \in \mathcal{M}_{n,1}(\mathbb{C})$, on a $X^*A^*AX = \|AX\|^2 \geqslant 0$. Par ailleurs il est facile de voir que la relation $\leqslant$ définie dans l'énoncé est une relation d'ordre.

Soit A et B dans $\mathcal{M}_n(\mathbb{C})$ et $\lambda \in [0,1]$. Il nous faut démontrer que

$$\Phi(\lambda A + (1-\lambda)B) \leqslant \lambda\Phi(A) + (1-\lambda)\Phi(B).$$

En développant, on obtient

$$\begin{aligned}\lambda\Phi(A) + (1-\lambda)\Phi(B) - \Phi(\lambda A + (1-\lambda)B) &= \lambda(1-\lambda)(A^*A + B^*B - A^*B - B^*A)\\ &= \lambda(1-\lambda)(A-B)^*(A-B).\end{aligned}$$

D'après ce qui précède, la matrice $(A-B)^*(A-B)$ est hermitienne positive. Comme $\lambda(1-\lambda) \in \mathbb{R}_+$, $\lambda\Phi(A) + (1-\lambda)\Phi(B) - \Phi(\lambda A + (1-\lambda)B)$ est hermitienne positive. ◁

2.55. Condition pour qu'une matrice hermitienne soit définie positive

Soit M une matrice hermitienne écrite sous la forme $\begin{pmatrix} a & v^* \\ v & A \end{pmatrix}$, où A est une matrice carrée d'ordre $n-1$. Montrer que la matrice M est définie positive si, et seulement si, a est un réel strictement positif et $aA - vv^*$ est hermitienne définie positive.

(**École normale supérieure**)

▷ **Solution.**

Puisque M est hermitienne, a est réel et A est hermitienne.

Considérons un élément quelconque X de $\mathcal{M}_{n,1}(\mathbb{C})$ écrit sous la forme $\begin{pmatrix} y \\ Z \end{pmatrix}$, où $y \in \mathbb{C}$ et $Z \in \mathcal{M}_{n-1,1}(\mathbb{C})$. On a alors $MX = \begin{pmatrix} ay + v^*Z \\ yv + AZ \end{pmatrix}$ et

$$X^*MX = \overline{y}ay + \overline{y}v^*Z + Z^*yv + Z^*AZ = a|y|^2 + 2\,\mathrm{Re}(\overline{y}v^*Z) + Z^*AZ,$$

car $\overline{v^*Z} = Z^*v$ par symétrie hermitienne.

La matrice M est définie positive si, et seulement si, cette quantité est strictement positive pour tout X non nul. En prenant $Z = 0$, on voit qu'il faut $a > 0$, ce que l'on suppose réalisé par la suite. Fixons Z et faisons varier y. Il existe $t \in \mathbb{R}$ tel que $v^*Z = |v^*Z|e^{it}$. Prenons y sous la forme $\rho e^{i\theta}$ avec $\rho \in \mathbb{R}_+$ et $\theta \in \mathbb{R}$. On obtient

$$X^*MX = a\rho^2 + 2\rho|v^*Z|\cos(t-\theta) + Z^*AZ.$$

Quand θ varie, le minimum de X^*MX est $a\rho^2 - 2\rho|v^*Z| + Z^*AZ$. Cette quantité est minimale pour $\rho = \dfrac{|v^*Z|}{a}$ et la valeur minimale obtenue est $\dfrac{1}{a}|v^*Z|^2 + Z^*AZ$. Puisque $|v^*Z|^2 = \overline{v^*Z}v^*Z = Z^*vv^*Z$, on a

$$-\frac{1}{a}|v^*Z|^2 + Z^*AZ = \frac{1}{a}(aZ^*AZ - Z^*vv^*Z) = \frac{1}{a}(Z^*(aA - vv^*)Z).$$

Ainsi M est définie positive si, et seulement si, $a > 0$ et pour tout $Z \in \mathcal{M}_{n-1,1}(\mathbb{C})$, $Z^*(aA - vv^*)Z > 0$, c'est-à-dire si $aA - vv^*$ est définie positive. ◁

2.56. Calcul de l'inverse d'une matrice par une méthode itérative

Soit X_0 et C dans $\mathbb{C}^n$, $B \in \mathcal{M}_n(\mathbb{C})$ et la suite $(X_n)_{n\in\mathbb{N}}$ telle que

$$X_{n+1} = BX_n + C.$$

1. Que dire si la suite (X_n) converge ?

2. Donner une condition suffisante pour que (X_n) ait une limite indépendante de X_0.

3. Soit $A = M - N$ hermitienne, avec M inversible. On suppose que A et $M^* + N$ sont définies positives. On munit $\mathbb{C}^n$ de la norme définie par $\|X\| = \sqrt{\langle X, AX\rangle}$. Prouver que $\|M^{-1}N\| < 1$.

4. Donner un algorithme permettant d'inverser A.

(École normale supérieure)

▷ **Solution.**

1. Si l'on note L la limite de la suite (X_n), on a $L = BL + C$ et donc $(I_n - B)L = C$: le vecteur C appartient à $\mathrm{Im}(I_n - B)$.

2. On munit $\mathcal{M}_n(\mathbb{C})$ d'une norme triple quelconque. Montrons que la condition $|||B||| < 1$ suffit.

En effet la matrice $I_n - B$ est alors inversible, car si $X \in \mathrm{Ker}(I_n - B)$, alors $BX = X$ et $\|X\| \leqslant \|BX\| \leqslant |||B|||X\|$, ce qui implique $X = 0$. Il existe un unique $L \in \mathbb{C}^n$ tel que $(I_n - B)L = C$. Pour tout $n \in \mathbb{N}$, on a $X_{n+1} - L = B(X_n - L)$ et donc $X_n - L = B^n(X_0 - L)$. On en déduit que

$$\|X_n - L\| \leqslant |||B^n|||X_0 - L\| \leqslant |||B|||^n\|X_0 - L\|.$$

Comme $|||B||| < 1$, la suite (X_n) converge vers L, limite indépendante de X_0.

3. Posons $B = M^{-1}N$. On a

$$|||B|||^2 = \sup_{X\neq 0} \frac{\|BX\|^2}{\|X\|^2} = \sup_{X\neq 0} \frac{X^*B^*ABX}{X^*AX}.$$

Cette borne supérieure étant atteinte, on a $\|B\| < 1$ si, et seulement si, $X^*B^*ABX < X^*AX$ pour tout $X \neq 0$, c'est-à-dire $X^*(A - B^*AB)X > 0$ pour tout $X \neq 0$. Il faut donc démontrer que la matrice hermitienne $A - B^*AB$ est définie positive.

On a $B = M^{-1}N = M^{-1}(M - A) = I_n - M^{-1}A$. On en déduit que

$$\begin{aligned} A - B^*AB &= A - (I_n - AM^{*-1})A(I_n - M^{-1}A) \\ &= A - A + AM^{*-1}A + AM^{-1}A - AM^{*-1}AM^{-1}A \\ &= AM^{*-1}(M + M^* - A)M^{-1}A \\ &= (M^{-1}A)^*(M + M^* - A)(M^{-1}A). \end{aligned}$$

Comme $M + M^* - A = M^* + N$ est définie positive, la matrice $A - B^*AB$ qui lui est congruente est aussi définie positive.

4. Si M et N vérifient les conditions de la question précédente, une suite (X_n) qui vérifie la relation de récurrence $X_{n+1} = BX_n + C$, converge d'après le question 2 vers $(I_n - B)^{-1}C = (M^{-1}A)^{-1}C = A^{-1}(MC)$. Pour déterminer une valeur approchée de $A^{-1}X$, il faut donc prendre $C = M^{-1}X$. On peut ainsi calculer $A^{-1}X$ pour tout X et donc déterminer A^{-1}.

La méthode n'a d'intérêt que si l'on peut calculer M^{-1} facilement. On peut prendre par exemple $M = \lambda I_n$, avec $\lambda > 0$. On a alors $M^* + N = M + N = 2M - A$. Cette matrice est définie positive si 2λ est supérieur à la plus grande valeur propre de A. ◁

2.57. Encadrement des valeurs propres de AB

Pour $A \in \mathcal{M}_n(\mathbb{C})$, on note $\lambda_{\max}(A)$ et $\lambda_{\min}(A)$ deux éléments de Sp A tels que, pour tout $\lambda \in \operatorname{Sp} A$, $|\lambda_{\min}(A)| \leqslant |\lambda| \leqslant |\lambda_{\max}(A)|$.

Soit A et B deux éléments de $\mathcal{M}_n(\mathbb{C})$. Montrer que, pour tout $\mu \in \operatorname{Sp}(AB)$,

$$\lambda_{\min}(AA^*)\lambda_{\min}(BB^*) \leqslant |\mu|^2 \leqslant \lambda_{\max}(AA^*)\lambda_{\max}(BB^*).$$

(École polytechnique)

▷ **Solution.**

La matrice AA^* est hermitienne positive, car pour tout $X \in \mathcal{M}_{n,1}(\mathbb{C})$, on a $X^*AA^*X = \|A^*X\|^2 \geqslant 0$. Il existe une base orthonormale $(Y_1, \ldots, Y_n)$ de $\mathbb{C}^n$ formée de vecteurs propres de AA^*, avec les valeurs propres positives $\lambda_1, \ldots, \lambda_n$. Si $X = \sum_{i=1}^{n} x_i Y_i$, on a alors

$$AA^*X=\sum_{i=1}^{n}\lambda_i x_i Y_i \quad \text{et} \quad \|A^*X\|^2=X^*AA^*X=\langle X, AA^*X\rangle=\sum_{i=1}^{n}\lambda_i|x_i|^2.$$

On en déduit, par définition de $\lambda_{\min}$ et $\lambda_{\max}$, que

$$\lambda_{\min}(AA^*)\|X\|^2 \leqslant \|A^*X\|^2 \leqslant \lambda_{\max}(AA^*)\|X\|^2.$$

Le même raisonnement s'applique à la matrice B et on obtient, pour tout $X \in \mathbb{C}^n$,

$$\begin{aligned}\lambda_{\min}(BB^*)\lambda_{\min}(AA^*)\|X\|^2 &\leqslant \lambda_{\min}(BB^*)\|A^*X\|^2 \leqslant \|B^*A^*X\|^2\\ &\leqslant \lambda_{\max}(BB^*)\|A^*X\|^2 \leqslant \lambda_{\max}(BB^*)\lambda_{\max}(AA^*)\|X\|^2.\end{aligned}$$

Soit $\mu \in \mathrm{Sp}(AB)$. Alors $\overline{\mu} \in \mathrm{Sp}(AB)^* = \mathrm{Sp}(B^*A^*)$, car $\mathrm{Sp}(AB) = \mathrm{Sp}({}^t(AB))$. Si X est un vecteur propre associé, on a $B^*A^*X = \overline{\mu}X$ et $\|B^*A^*X\|^2 = \|\overline{\mu}X\|^2 = |\mu|^2\|X\|^2$ et l'inégalité précédente fournit l'encadrement voulu. ◁

L'exercice suivant est la version hermitienne de l'exercice 2.30. Toutes les solutions que nous avions données peuvent se généraliser.

2.58. Inégalité d'Hadamard pour une matrice hermitienne

Soit $H = (h_{ij}) \in \mathcal{M}_n(\mathbb{C})$ une matrice hermitienne positive.

1. Montrer que $0 \leqslant \det H \leqslant \prod_{j=1}^{n} h_{jj}$.

2. En désignant par R et S les parties réelle et imaginaire de H, montrer que $\det H \leqslant \det R$.

3. Soit $A = (a_{ij})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_n(\mathbb{C})$. Démontrer l'inégalité

$$|\det A| \leqslant \prod_{i=1}^{n}\sqrt{|a_{i1}|^2+|a_{i2}|^2+\cdots+|a_{in}|^2}.$$

(École polytechnique)

Solution.

1. L'inégalité $0 \leqslant \det H$ est vérifiée, car H est positive. Les coefficients h_{jj} sont positifs, car $h_{jj} = \Phi(e_j)$, où Φ est la forme hermitienne associée à H dans la base canonique $(e_1, \ldots, e_n)$ de $\mathbb{C}^n$. Si $\det H = 0$, il n'y a rien à démontrer. On suppose donc que H est définie positive : les h_{jj} sont alors strictement positifs.

Il existe $P = (p_{ij})$ unitaire et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale, à co-

efficients diagonaux strictement positifs, telles que $H = PDP^*$. On en déduit que, pour tout $j \in [\![1, n]\!]$, $h_{jj} = \sum_{k=1}^{n} p_{jk}\lambda_k\overline{p_{ik}} = \sum_{k=1}^{n} |p_{jk}|^2\lambda_k$. On remarque que, pour tout j, $\sum_{k=1}^{n} |p_{jk}|^2 = 1$, car P est unitaire. On en déduit, en utilisant la concavité de la fonction ln que

$$\ln h_{jj} = \ln\left(\sum_{k=1}^{n} |p_{jk}|^2\lambda_k\right) \geqslant \sum_{k=1}^{n} |p_{jk}|^2 \ln\lambda_k,$$

puis que

$$\sum_{j=1}^{n} \ln h_{jj} \geqslant \sum_{j=1}^{n}\sum_{k=1}^{n} |p_{jk}|^2 \ln\lambda_k \geqslant \sum_{k=1}^{n} \ln\lambda_k \left(\sum_{k=1}^{n} |p_{jk}|^2\right) \geqslant \sum_{k=1}^{n} \ln\lambda_k,$$

car P est unitaire.

On obtient $\prod_{j=1}^{n} \lambda_j \geqslant \prod_{j-1}^{n} h_{jj}$, ce qui est le résultat voulu.

2. On a par définition $R = \frac{1}{2}(H+\overline{H})$ et $S = \frac{1}{2i}(H-\overline{H})$. Puisque H est hermitienne, R est symétrique réelle et S antisymétrique réelle. Il existe $Q \in O_n(\mathbb{R})$ et Δ diagonale réelle telle ${}^tQRQ = \Delta = \mathrm{Diag}(\mu_1, \ldots, \mu_n)$. On pose $S' = {}^tQSQ$ et $H' = {}^tQHQ = \Delta + iS'$. La matrice S' est antisymétrique comme S et H' est hermitienne positive comme H. On a de plus $\det H' = \det H$ et $\det R = \det\Delta$, car $(\det Q)^2 = 1$. Comme les termes diagonaux de H' sont égaux à ceux de Δ, puisque S' est antisymétrique, la première question appliquée à H' donne

$$\det H = \det H' \leqslant \prod_{j=1}^{n} \mu_j = \det\Delta = \det R,$$

car Δ est diagonale.

3. La matrice AA^* est hermitienne positive, car pour $X \in \mathbb{C}^n$, $X^*AA^*X = \|A^*X\|^2 \geqslant 0$. Son i-ième coefficient diagonal ($1 \leqslant i \leqslant n$) est $|a_{i1}|^2 + |a_{i2}|^2 + \cdots + |a_{in}|^2$. En appliquant la première question, il vient

$$\det AA^* = |\det A|^2 \leqslant \prod_{i=1}^{n} \left(|a_{i1}|^2 + |a_{i2}|^2 + \cdots + |a_{in}|^2\right).$$

On obtient le résultat voulu en prenant la racine carrée. ◁

En considérant A^*A, *on aurait obtenu*

$$|\det A| \leqslant \prod_{j=1}^{n} \sqrt{|a_{1j}|^2 + |a_{2j}|^2 + \cdots + |a_{nj}|^2}.$$

Dans l'exercice suivant, on démontre qu'une matrice hermitienne positive possède une unique racine carrée hermitienne positive. C'est un résultat classique que l'on rapprochera de celui de l'exercice 2.12. La deuxième question a été également traitée dans le cas euclidien (exercice 2.13).

2.59. Produit de matrices hermitiennes positives

1. Soit H une matrice hermitienne positive. Montrer qu'il existe une unique matrice K hermitienne positive telle que $K^2 = H$.

2. Soit H et K deux matrices hermitiennes positives. Montrer que les valeurs propres de HK sont réelles positives.

3. Soit H, K, L trois matrices hermitiennes positives. On suppose HKL hermitienne. Montrer qu'elle est positive (on commencera par supposer H définie positive).

(École polytechnique)

Solution.

1. La matrice $H \in \mathcal{M}_n(\mathbb{C})$ est diagonalisable dans une base orthonormale : il existe $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale à termes diagonaux positifs et $P \in U_n(\mathbb{C})$ telles que $H = PDP^*$. On écrit $D = \Delta^2$, avec $\Delta = \mathrm{Diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ et $H = P\Delta^2P^* = (P\Delta P^*)^2$. La matrice $K = P\Delta P^*$ convient.

Si K' est une autre matrice répondant à la question, elle est diagonalisable et ses valeurs propres sont nécessairement les racines carrés de celle de K. Elle est donc semblable à Δ : il existe $Q \in U_n(\mathbb{C})$ telle que $K' = Q\Delta Q^*$. On en déduit que $H = PDP^* = K'^2 = Q\Delta^2Q^* = QDQ^*$, ce qui peut encore s'écrire $Q^*PD = DQ^*P$. Si on pose $Q^*P = (a_{ij})$, on a pour $1 \leqslant i, j \leqslant n$, $a_{ij}\lambda_j = \lambda_i a_{ij}$. Si $\lambda_i \neq \lambda_j$, alors $a_{ij} = 0$. Dans tous les cas, on a $a_{ij}\sqrt{\lambda_j} = \sqrt{\lambda_i}a_{ij}$ et donc $Q^*P\Delta = \Delta Q^*P$, ce qui implique $P\Delta P^* = Q\Delta Q^*$, soit $K = K'$, ce qui montre l'unicité.

2. Soit λ une valeur propre de HK et X un vecteur propre associé. On a

$$X^*KHKX = \lambda X^*KX,$$

et comme H est positive, $X^*KHKX = (KX)^*H(KX) \geqslant 0$. Ainsi, si $X^*KX > 0$, on obtient $\lambda \geqslant 0$. Sinon $X^*KX = 0$ et d'après la question 1, il existe une matrice hermitienne positive K' telle que $K = K'^2$. On peut alors écrire $0 = X^*KX = X^*K'^2X = \|K'X\|^2$ et il s'ensuit que $K'X = 0$ et donc $KX = 0$ et $HKX = 0$. Dans ce cas, λ est nul.

Cette technique a été vue dans l'exercice 2.13.

3. Supposons que H soit définie positive. Soit H' hermitienne positive telle que $H'^2 = H$. On a même H' définie positive. On écrit alors

$$HKL = H'(H'KLH'^{-1})H'.$$

Ainsi HKL et $H'KLH'^{-1}$ (cette dernière est aussi hermitienne, car égale à $H'^{-1}HKLH'^{-1}$) représentent la même forme hermitienne. Or le spectre de KL est contenu dans $\mathbb{R}_+$ d'après la question précédente. Donc celui de $H'KLH'^{-1}$ qui lui est semblable est aussi contenu dans $\mathbb{R}_+$. Donc, HKL est la matrice d'une forme hermitienne positive, puisque son spectre est contenu dans $\mathbb{R}_+$.

La situation est plus délicate si on ne suppose plus H définie. On essaie de se ramener au premier cas. On identifie matrice et endomorphisme canoniquement associé. Pour $\alpha > 0$, on pose

$$M(\alpha) = (H + \alpha L)KL = HKL + \alpha LKL,$$

qui est une matrice hermitienne, car somme de matrices hermitiennes. Considérons l'orthogonal F de $\operatorname{Ker} H \cap \operatorname{Ker} L$. C'est un sous-espace stable par $H + \alpha L$ (c'est l'orthogonal d'un sous-espace stable par un endomorphisme hermitien). La restriction de $H + \alpha L$ à F est définie positive. En effet, elle est clairement positive et si on suppose que $X \in F$ et $(H + \alpha L)X = 0$, on a

$$0 \leqslant X^*HX = -\alpha X^*LX \leqslant 0$$

et $X^*HX = X^*LX = 0$ et en faisant intervenir les racines carrées de H et L comme dans la question précédente, on en déduit $HX = LX = 0$ et $X \in \operatorname{Ker} H \cap \operatorname{Ker} L = F^\perp$. Finalement, on a bien $X = 0$.

En considérant une base orthonormale de $\mathbb{C}^n$ constituée d'une base de F et d'une base de $\operatorname{Ker} H \cap \operatorname{Ker} L$, on voit qu'il existe une matrice U unitaire telle que

$$UHU^* = \begin{pmatrix} H_1 & 0 \\ 0 & 0 \end{pmatrix}, \quad UKU^* = \begin{pmatrix} K_1 & K_2 \\ {}^tK_2 & K_3 \end{pmatrix}, \quad ULU^* = \begin{pmatrix} L_1 & 0 \\ 0 & 0 \end{pmatrix}$$

On obtient

$$UHKLU^* = \begin{pmatrix} H_1K_1L_1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{et} \quad UM(\alpha)U^* = \begin{pmatrix} (H_1 + \alpha L_1)K_1L_1 & 0 \\ 0 & 0 \end{pmatrix}$$

D'après ce qui précède, la matrice $H_1 + \alpha L_1$ est hermitienne définie positive. Comme les matrices K_1 et L_1 sont hermitiennes positives, l'étude du premier cas montre que $(H_1 + \alpha L_1)K_1L_1$ est positive. Il en est de même de $UM(\alpha)U^*$, car si $X \in \mathbb{C}^n$ est écrit $X \begin{pmatrix} Y \\ Z \end{pmatrix}$, où $Y \in \mathbb{C}^r$, où $r = \dim F$, on a

$$X^*UHKLUX = Y^*(H_1 + \alpha L_1)K_1L_1Y \geqslant 0.$$

La matrice $M(\alpha)$ qui lui est congruente est aussi positive. On a pour tout $X \in \mathbb{C}^n$, $X^*M(\alpha)X \geqslant 0$. En faisant tendre α vers 0, on obtient $X^*HKLX \geqslant 0$: la matrice HKL est bien positive. ◁

L'existence d'une racine carrée est encore utilisée dans les deux exercices suivants.

2.60. Produit de Schur de deux matrices hermitiennes positives

Soit $A = (a_{ij})$ et $B = (b_{ij})$ deux matrices de $\mathcal{M}_n(\mathbb{C})$ hermitiennes positives. Montrer que la matrice $C = (a_{ij}b_{ij})_{1 \leqslant i,j \leqslant n}$ est hermitienne positive.

(École normale supérieure)

Solution.

Il est clair que C est hermitienne : pour tout $(i, j) \in [\![1, n]\!]^2$,

$$\overline{c_{ij}} = \overline{a_{ij}}\,\overline{b_{ij}} = a_{ji}b_{ji} = c_{ji}.$$

Pour démontrer que C est positive, nous considérons une matrice H hermitienne positive telle que $A = H^2$ (cf. exercice 2.59). On a alors, pour tout $(i, j) \in [\![1, n]\!]^2$,

$$c_{ij} = \sum_{k=1}^{n} h_{ik}h_{kj}b_{ij} = \sum_{k=1}^{n} \overline{h_{ki}}h_{kj}b_{ij}$$

et pour ${}^tX = (x_1, \ldots, x_n) \in \mathbb{C}^n$,

$${}^t\overline{X}CX = \sum_{1 \leqslant i,j \leqslant n} c_{ij}\overline{x_i}x_j = \sum_{k=1}^{n} \left(\sum_{1 \leqslant i,j \leqslant n} b_{ij}\overline{h_{ki}x_i}h_{kj}x_j \right).$$

La matrice B étant positive, la somme $\sum_{1 \leqslant i,j \leqslant n} b_{ij}\overline{h_{ki}x_i}h_{kj}x_j$ est positive pour tout $k \in [\![1, n]\!]$. On en déduit que ${}^t\overline{X}CX \geqslant 0$: la matrice C est hermitienne positive. ◁

Si les matrices A et B sont définies positives, ${}^t\overline{X}CX = 0$ implique, pour tout $k \in [\![1, n]\!]$, $\sum_{1 \leqslant i,j \leqslant n} b_{ij}\overline{h_{ki}x_i}h_{kj}x_j = 0$ et puisque B est définie positive, $h_{ki}x_i = 0$ pour tout i. Comme A, H est inversible, donc aucune de ses colonnes n'est nulle et, pour tout $i \in [\![1, n]\!]$, il existe $k \in [\![1, n]\!]$ tel que $h_{ki} \neq 0$, ce qui implique $x_i = 0$ et finalement $X = 0$: la matrice C est définie positive.

2.61. Inégalités

On se place dans $\mathbb{C}^n$ muni de sa structure hermitienne canonique.

1. Soit L un endomorphisme hermitien défini positif et $p \in \mathbb{N}^*$. Montrer que, pour tout $(x, y) \in (\mathbb{C}^n)^2$, on a

$$\langle (\mathrm{Id} + \mathrm{L})^{-p}(x+y), x+y\rangle \leqslant \|x\|^2 + \langle \mathrm{L}^{-p}(y), y\rangle.$$

2. Soit H et K deux endomorphismes hermitiens définis positifs. Montrer que, pour tout $(x, y) \in (\mathbb{C}^n)^2$, on a

$$\langle (\mathrm{H} + \mathrm{K})^{-1}(x+y), x+y\rangle \leqslant \langle \mathrm{H}^{-1}(x), x\rangle + \langle \mathrm{K}^{-1}(y), y\rangle.$$

(**École polytechnique**)

▷ **Solution.**

1. On considère une base orthonormale $\mathcal{B} = (e_1, \ldots, e_n)$ de vecteurs propres de L ; on note $\lambda_1, \ldots, \lambda_n$ les valeurs propres (strictement positives) correspondantes. Si $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$, on a $\mathrm{L}(x) = \sum_{i=1}^{n} \lambda_i x_i e_i$ et $(\mathrm{Id} + \mathrm{L})(x) = \sum_{i=1}^{n} (1+\lambda_i) x_i e_i$. On en déduit que $\mathrm{L}^{-1}(x) = \sum_{i=1}^{n} \lambda_i^{-1} x_i e_i$ et $(\mathrm{Id} + \mathrm{L})^{-1} = \sum_{i=1}^{n} (1+\lambda_i)^{-1} x_i e_i$. Si l'on note $y = (y_1, \ldots, y_n)$, on obtient

$$\begin{aligned}\langle (\mathrm{Id} + \mathrm{L})^{-p}(x+y), x+y\rangle &= \left\langle \sum_{i=1}^{n} (1+\lambda_i)^{-p}(x_i+y_i)e_i, \sum_{i=1}^{n} (x_i+y_i)e_i \right\rangle \\ &= \sum_{i=1}^{n} (1+\lambda_i)^{-p} |x_i + y_i|^2,\end{aligned}$$

et de même

$$\|x\|^2 + \langle \mathrm{L}^{-p}(y), y\rangle = \sum_{i=1}^{n} |x_i|^2 + \sum_{i=1}^{n} \lambda_i^{-p} |y_i|^2.$$

Pour montrer l'inégalité voulue, il suffit de vérifier que, pour tout i,

$$(1+\lambda_i)^{-p} |x_i + y_i|^2 \leqslant |x_i|^2 + \lambda_i^{-p} |y_i|^2.$$

On étudie le signe de $\delta = (1+\lambda)^{-p}|u+v|^2 - |u|^2 - \lambda^{-p}|v|^2$, pour $(u, v) \in \mathbb{C}^2$ et $\lambda \in \mathbb{R}_+^*$. On a

$$\begin{aligned}\delta &\leqslant (1+\lambda)^{-p}(|u|^2 + |v|^2 + 2|u||v|) - |u|^2 - \lambda^{-p}|v|^2 \\ &\leqslant ((1+\lambda)^{-p} - 1)|u|^2 + 2(1+\lambda)^{-p}|u||v| + ((1+\lambda)^{-p} - \lambda^{-p})|v|^2.\end{aligned}$$

Le discriminant réduit de ce trinôme en $|u|$ vaut

$$\Delta = |v|^2 (1+\lambda)^{-p} \lambda^{-p} (1 + \lambda^p - (1+\lambda)^p).$$

Il est négatif car $(1+\lambda)^p \geqslant 1+\lambda^p$, comme le montre la formule du binôme. Comme le coefficient de $|u|^2$ est négatif, on a $\delta \leqslant 0$, pour tout $(u,v) \in \mathbb{C}^2$, ce qui permet de conclure.

2. On se ramène à la relation démontrée dans la question 1. Il existe un endomorphisme hermitien G défini positif tel que $H = G^2$ (voir exercice 2.59). Comme G^{-1} est aussi hermitien, on a

$$\langle H^{-1}(x), x\rangle = \langle G^{-1}(G^{-1}(x)), x\rangle = \|G^{-1}(x)\rangle\|^2.$$

Ceci nous pousse à poser $x' = G^{-1}(x)$ et $y' = G^{-1}(y)$. On a alors

$$\langle K^{-1}(y), y\rangle = \langle K^{-1}(G(y')), G(y')\rangle = \langle (G \circ K^{-1} \circ G)(y'), y'\rangle.$$

On pose $L = G^{-1} \circ K \circ G^{-1}$. L'endomorphisme L est hermitien et défini positif, car pour tout $z \in \mathbb{C}^n \setminus \{0\}$,

$$\langle L(z), z\rangle = \langle K(G^{-1}(z)), G^{-1}(z)\rangle > 0,$$

car K est défini positif.

En appliquant le résultat de la question 1, avec $p = 1$, on obtient

$$\begin{aligned}\langle H^{-1}(x), x\rangle + \langle K^{-1}(y), y\rangle &= \|x'\|^2 + \langle L^{-1}(y'), y'\rangle \\ &\geqslant \langle (\mathrm{Id}+L)^{-1}(x'+y'), x'+y'\rangle.\end{aligned}$$

Il ne reste plus qu'à transformer cette dernière expression. On peut écrire

$$\mathrm{Id}+L = \mathrm{Id}+G^{-1}\circ K\circ G^{-1} = G^{-1}\circ(G^2+K)\circ G^{-1} = G^{-1}\circ(H+K)\circ G^{-1}$$

et $(\mathrm{Id}+L)^{-1} = G \circ (H+K)^{-1} \circ G$. On en déduit que

$$\begin{aligned}((\mathrm{Id}+L)^{-1}(x'+y'), x'+y'\rangle &= \langle G \circ (H+K)^{-1} \circ G(x'+y'), x'+y'\rangle \\ &= \langle (H+K)^{-1}(x+y), x+y)\end{aligned}$$

et l'inégalité est démontrée. ◁

Voici maintenant la version hermitienne de la décomposition polaire. Elle a été présentée dans le cadre euclidien dans les exercices 2.26 et 2.27.

2.62. Décomposition polaire dans $\mathcal{M}_n(\mathbb{C})$

1. Montrer que pour tout $A \in \mathcal{M}_n(\mathbb{C})$, il existe une unique matrice hermitienne positive H telle que $A^*A = H^2$.

2. Montrer que, pour tout $A \in GL_n(\mathbb{C})$, il existe un unique couple (U, H), où $U \in U_n$ et H est hermitienne positive, tel que $A = UH$.

3. On considère sur $\mathcal{M}_n(\mathbb{C})$ la norme définie par $\|A\| = \sqrt{\mathrm{Tr}(A^*A)}$. Montrer que

$$\|A - U\| = \inf_{V \in U_n} \|A - V\| = d(A, U_n).$$

(École polytechnique)

▷ **Solution.**

1. Soit $M = A^*A$. La matrice M est hermitienne positive. La question a été traitée dans l'exercice 2.59. Nous présentons une autre version pour l'unicité : soit H' une autre matrice répondant à la question. Le spectre de H'^2 est celui de M et donc celui de H^2. Comme les valeurs propres de H' sont aussi positives, H et H' ont les mêmes valeurs propres.

Soit λ une valeur propre de H (et donc de H'). On sait déjà que $\mathrm{Ker}(H' - \lambda I_n) \subset \mathrm{Ker}(M - \lambda^2 I_n) = \mathrm{Ker}(H - \lambda I_n)$, la dernière égalité étant justifiée par construction de H. Comme H' est diagonalisable, la somme de ses sous-espaces propres vaut $\mathbb{C}^n = \bigoplus_{\lambda \in \mathrm{Sp}(H)} \mathrm{Ker}(H - \lambda I_n)$ et pour des raisons de dimension, $\mathrm{Ker}(H' - \lambda I_n) = \mathrm{Ker}(H - \lambda I_n)$. Il s'ensuit que $H = H'$ puisqu'elles ont les mêmes valeurs propres et les mêmes sous-espaces propres.

2. *Unicité.* Si $A = UH$, avec $U \in U_n$ et H hermitienne positive, alors $A^*A = HU^*UH = H^2$. D'après la question 1, H est unique. La matrice H étant inversible puisque A l'est, $U = AH^{-1}$ est uniquement déterminée.

Existence. Prenons H la racine carrée de $M = A^*A$ définie dans la question précédente. Ici, H est hermitienne définie positive, car inversible comme M. On peut poser $U = AH^{-1}$. Vérifions que U est unitaire :

$$U^*U = H^{-1}A^*AH^{-1} = H^{-1}H^2H^{-1} = I_n.$$

3. En gardant les notations précédentes, on écrit $H = P^*\Delta P$, où P est unitaire et $\Delta = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. On peut remarquer que si M et V sont deux matrices de taille n, V étant unitaire, on a $\|VM\| = \|MV\| = \|M\|$. On en déduit

$$\|A - U\|^2 = \|U(H - I_n)\|^2 = \|H - I_n\|^2 = \mathrm{Tr}(H - I_n)^2$$
$$= \mathrm{Tr}(\Delta - I)^2 = \sum_{i=1}^{n} (\lambda_i - 1)^2 .$$

Pour $V \in U_n$, on obtient

$$\begin{aligned}\|A - V\| &= \|UH - V\| = \|V^{-1}UH - I_n\| = \|V^{-1}UP^*\Delta P - I_n\| \\ &= \|P(V^{-1}UP^*\Delta P - I_n)P^*\| = \|PV^{-1}UP^*\Delta - I_n\|,\end{aligned}$$

car P est unitaire. On pose $W = PV^{-1}UP^*$ qui est une matrice unitaire. On obtient

$$\begin{aligned}\|A - V\|^2 &= \|W\Delta - I_n\|^2 = \mathrm{Tr}(\Delta W^* W\Delta - W\Delta - \Delta W^* + I_n) \\ &= \mathrm{Tr}(\Delta^2 - W\Delta - \Delta W^* + I_n) \\ &= \sum_{i=1}^{n} \left(\lambda_i^2 - (w_{ii} + \overline{w_{ii}})\lambda_i + 1\right) \\ &\geqslant \sum_{i=1}^{n} (\lambda_i^2 - 2\lambda_i + 1) = \|A - U\|^2,\end{aligned}$$

car $\sum_{j=1}^{n} |w_{ij}|^2 = 1$ implique $\mathrm{Re}(w_{ii}) \leqslant |w_{ii}| \leqslant 1$. L'égalité a lieu si $w_{ii} = 1$ pour $1 \leqslant i \leqslant n$ et donc $W = I_n$. On a alors $V^{-1}U = I_n$ et donc $V = U$. ◁

Beaucoup d'autres exercices du début de ce chapitre peuvent s'adapter en termes hermitiens. Sans faire un catalogue exhaustif, nous précisons que c'est notamment le cas pour les décompositions classiques et le théorème du minimax (voir exercices 2.29 et 2.37).

2.63. Écriture d'un endomorphisme comme combinaison linéaire d'endomorphismes unitaires

Soit E un espace vectoriel hermitien de dimension n. On note $\|\ \|$ la norme dans $\mathcal{L}(E)$ subordonnée à la norme de E.

1. Soit $f \in \mathcal{L}(E)$. Montrer l'équivalence de :

(i) f est hermitien et $\|f\| \leqslant 1$;

(ii) il existe un endomorphisme unitaire g de E tel que

$$f = \frac{1}{2}(g + g^*).$$

2. Montrer que tout $f \in \mathcal{L}(E)$ s'écrit comme combinaison linéaire d'au plus quatre endomorphismes unitaires de E.

(**École polytechnique**)

▷ **Solution.**

1. Supposons que $f = \frac{1}{2}(g + g^*)$. avec g unitaire. Alors f est auto-adjoint et d'autre part comme $\|g\| = \|g^*\| = 1$ on a $\|f\| \leqslant 1$. Donc (ii) implique (i).

Traitons la réciproque matriciellement. Soit A la matrice de f dans une base orthonormée B de E. La matrice A est hermitienne et donc unitairement semblable à une matrice diagonale réelle $D = \mathrm{Diag}(d_1, \ldots, d_n)$. Soit U unitaire telle que $A = UDU^*$. Par hypothèse, $\|f\| \leqslant 1$ de sorte que $\forall i \in [\![1, n]\!]$, $d_i \in [-1, 1]$. On peut donc poser $d_i = \cos\theta_i$ pour tout $i \in [\![1, n]\!]$. Posons alors $V = U\,\mathrm{Diag}(e^{i\theta_1}, \ldots, e^{i\theta_n})U^*$. La matrice V est unitaire et on a $A = \frac{1}{2}(V + V^*)$.

2. Le résultat est vérifié pour les endomorphismes hermitiens de norme inférieure à 1. Si f est hermitien de norme supérieure à 1, il suffit d'appliquer le résultat précédent à $\frac{f}{\|f\|}$. Pour terminer, il suffit de prouver que tout endomorphisme f de E est combinaison linéaire de deux endomorphismes hermitiens. Or, il suffit d'écrire

$$f = \frac{f + f^*}{2} + \frac{f - f^*}{2} = \underbrace{\frac{f + f^*}{2}}_{\text{hermitien}} + i\underbrace{\frac{i(f^* - f)}{2}}_{\text{hermitien}}. \quad \triangleleft$$

En fait, on peut avoir mieux avec la décomposition polaire (cf. exercice 2.62) : tout $f \in \mathcal{L}(E)$ peut s'écrire $f = uh$ avec u unitaire et h hermitien. Il en résulte que f est combinaison linéaire de deux endomorphismes unitaires.

À partir de là, on peut montrer que le centre du groupe unitaire de E correspond aux homothéties unitaires. En effet, si $u \in Z(U(E))$, u commute avec tout élément de $\mathcal{L}(E)$ d'après le résultat précédent. Donc u est une homothétie. Réciproquement, les homothéties unitaires c'est-à-dire $f = \lambda\,\mathrm{Id}_E$ avec $|\lambda| = 1$ sont dans le centre de $U(E)$. En particulier, $Z(U(E)) \simeq S^1$.

L'exercice suivant est l'occasion de démontrer un théorème général de réduction des endomorphismes normaux (ce sont les endomorphismes qui commutent avec leur adjoint). Nous allons constater qu'ils sont diagonalisables en base orthonormée. On peut démontrer à partir de là le théorème spectral pour les endomorphismes hermitiens et le fait qu'une matrice unitaire est diagonalisable.

2.64. Matrices normales

Soit $A \subset \mathcal{M}_n(\mathbb{C})$, $\lambda_1, \ldots, \lambda_n$ ses valeurs propres. Montrer que les trois conditions suivantes sont équivalentes :

(i) $AA^* = A^*A$ (on dit que A est *normale*) ;

(ii) A est diagonalisable en base orthonormale ;

(iii) $\mathrm{Tr}(AA^*) = \sum_{i=1}^{n} |\lambda_i|^2$.

(École polytechnique)

Solution.

• Montrons que si la condition $AA^* = A^*A$ est réalisée, alors A est diagonalisable dans une base orthonormale, i.e. il existe U unitaire telle que $A = UDU^*$, où $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$. Soit f l'endomorphisme de $\mathbb{C}^n$ dont la matrice dans la base canonique est A. L'hypothèse se traduit par $f^* \circ f = f \circ f^*$, où f^* est l'adjoint de f. Il faut démontrer que f est diagonalisable dans une base orthonormale.

On raisonne par récurrence sur n, le résultat étant évident pour $n = 1$. Supposons le résultat vrai jusqu'au rang $n-1$ et montrons-le au rang n. Soit λ une valeur propre de f, E_λ l'espace propre correspondant. Comme f^* et f commutent, E_λ est stable par f^*. Posons $F_\lambda = (E_\lambda)^\perp$. Montrons que F_λ est stable par f et f^*. Soit $y \in F_\lambda$. Pour tout $x \in E_\lambda$, on a

$$\begin{aligned} \langle x, f^*(y)\rangle &= \langle f(x), y\rangle = \langle \lambda x, y\rangle = \overline{\lambda}\langle x, y\rangle = 0 \\ \langle x, f(y)\rangle &= \langle f^*(x), y\rangle = 0. \end{aligned}$$

car $f^*(x) \in E_\lambda$. Ainsi $f^*(y)$ et $f(y)$ appartiennent à F_λ.

Comme f_{F_λ} et $(f_{|F_\lambda})^* = f^*_{|F_\lambda}$ commutent et que $\dim F_\lambda \leqslant n-1$, on peut appliquer l'hypothèse de récurrence à $f_{|F_\lambda}$. Il existe une base orthonormale de F_λ qui diagonalise $f_{|F_\lambda}$. En lui adjoignant une base orthonormale de E_λ, on obtient une base orthonormale de E qui diagonalise f. L'implication $(i) \Longrightarrow (ii)$ est prouvée.

• Montrons que $(ii) \Longrightarrow (iii)$: on a $A = UDU^*$, avec $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ et U unitaire. On en déduit que $A^* = U\overline{D}U^*$ et $AA^* = UD\overline{D}U^*$. On obtient alors

$$\mathrm{Tr}(AA^*) = \mathrm{Tr}(UD\overline{D}U^*) = \mathrm{Tr}(D\overline{D}U^*U) = \mathrm{Tr}(D\overline{D}) = \sum_{i=1}^{n} |\lambda_i|^2.$$

C'est ce qu'on voulait.

• Pour terminer il reste à prouver que $(iii) \Longrightarrow (i)$. Supposons que $\mathrm{Tr}(AA^*) = \sum_{i=1}^{n} |\lambda_i|^2$. *A priori*, on ne peut pas diagonaliser A dans une

base orthonormale, mais on peut la trigonaliser dans une telle base, *i.e.* il existe T matrice triangulaire supérieure et U unitaire telle que $A = UTU^*$.

En effet, puisque $A \in \mathcal{M}_n(\mathbb{C})$. A est trigonalisable et il existe T' triangulaire supérieure et $P \in GL_n(\mathbb{C})$ telles que $A = PT'P^{-1}$. On note $\mathcal{B}_0$ la base canonique de $\mathbb{C}^n$, $\mathcal{B}$ la base dont la matrice de passage par rapport à $\mathcal{B}_0$ est P et $\mathcal{B}'$ la base orthonormale obtenue à partir de $\mathcal{B}$ par le procédé d'orthonormalisation de Gram-Schmidt. La matrice de passage de T'' de $\mathcal{B}$ à $\mathcal{B}'$ est triangulaire supérieure et $U = PT''$, matrice de passage de $\mathcal{B}_0$ à $\mathcal{B}'$ est unitaire. En écrivant $P = UT''^{-1}$, on obtient enfin $A = UT''^{-1}T'T''U^*$. La matrice $T = T''^{-1}T'T''$ est triangulaire supérieure et a les propriétés voulues.

On a alors $A^* = UT^*U^*$, $AA^* = UTT^*U^*$ et

$$\mathrm{Tr}(AA^*) = \mathrm{Tr}(UTT^*U^*) = \mathrm{Tr}(TT^*) = \sum_{1 \leqslant i,j \leqslant n} |t_{ij}|^2.$$

Comme les termes diagonaux de T sont $\lambda_1, \ldots, \lambda_n$, puisque T est semblable à A, l'hypothèse s'écrit $\sum_{i \neq j} |t_{ij}|^2 = 0$. La matrice T est donc diagonale. Ainsi T commute avec T^* et $A^*A = UT^*TU^* = UTT^*U^* = AA^*$ ce qui prouve (i). $\lhd$

Dans la dernière implication on a prouvé le théorème de Schur : toute matrice de $\mathcal{M}_n(\mathbb{C})$ est trigonalisable dans une base orthonormée.

Nous terminons ce chapitre par deux énoncés délicats.

2.65. Théorème de Mirman (1968)

$\mathbb{C}^n$ est muni de sa structure hermitienne canonique et $\mathcal{M}_n(\mathbb{C})$ de la norme subordonnée.

1. Soit $A \in \mathcal{M}_n(\mathbb{C})$ et $B = \begin{pmatrix} I_n & A \\ A^* & I_n \end{pmatrix}$. Montrer que la matrice B est positive si. et seulement si, $\|A\| \leqslant 1$.

2. Soit $F = \{{}^t\overline{X}AX. \quad \|X\| = 1\}$ et (H) la propriété : il existe trois nombres complexes z_1, z_2, z_3 de module 1 tels que F soit contenu dans le triangle $z_1z_2z_3$.

Montrer que (H) implique $\|A\| \leqslant 1$. On pourra d'abord montrer que (H) équivaut à : il existe trois nombres complexes z_1, z_2, z_3 de module 1 et trois matrices hermitiennes positives A_1, A_2, A_3 tels que $A = z_1A_1 + z_2A_2 + z_3A_3$ et $I_n = A_1 + A_2 + A_3$.

(**École normale supérieure**)

▷ **Solution.**

1. La matrice B est clairement hermitienne. Considérons un élément quelconque $Z \in \mathcal{M}_{2n,1}(\mathbb{C})$ que nous écrirons $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$, où X et Y sont dans $\mathcal{M}_{n,1}(\mathbb{C})$. On a alors $BZ = \begin{pmatrix} X + AY \\ A^*X + Y \end{pmatrix}$ et

$$\begin{aligned} Z^*BZ &= X^*X + X^*AY + Y^*A^*X + Y^*Y \\ &= \|X\|^2 + \|Y\|^2 + X^*AY + \overline{X^*AY} \\ &= \|X\|^2 + \|Y\|^2 + 2\,\mathrm{Re}(X^*AY). \end{aligned}$$

On veut montrer que Z^*BZ est $\geqslant 0$ pour tout $Z \in \mathcal{M}_{2n,1}(\mathbb{C})$ si, et seulement si, $\|A\| \leqslant 1$.

Supposons $\|A\| \leqslant 1$. On obtient alors

$$|\,\mathrm{Re}(X^*AY)| \leqslant |X^*AY| \leqslant \|X\|\|AY\| \leqslant \|A\|\|X\|\|Y\| \leqslant \|X\|\|Y\|,$$

en utilisant l'inégalité de Cauchy-Schwarz et la définition de la norme subordonnée. On en déduit que, pour tout $Z \in \mathcal{M}_{2n,1}(\mathbb{C})$,

$$Z^*BZ \geqslant \|X\|^2 + \|Y\|^2 - 2\|X\|\|Y\| \geqslant (\|X\| - \|Y\|)^2 \geqslant 0,$$

donc B est positive.

Supposons $\|A\| > 1$. On a, par définition, $\|A\| = \sup_{\|X\|=1} \|AX\|$. Comme la sphère unité est compacte, cette borne supérieure est atteinte. Il existe $Y \in \mathcal{M}_{n,1}(\mathbb{C})$ tel que $\|Y\| = 1$ et $\|AY\| = \|A\|$. Prenons enfin $X = AY$ et $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$. On a alors $X^*AY = \|X\|^2 = \|AY\|^2 = \|A\|^2$. On en déduit que

$$Z^*BZ = \|A\|^2 + 1 - 2\|A\|^2 = 1 - \|A\|^2 < 0.$$

La matrice B n'est pas positive.

2. Pour commencer, démontrons l'équivalence suggérée par l'énoncé. Supposons qu'il existe trois nombres complexes de module 1, z_1, z_2 et z_3 et trois matrices hermitiennes positives A_1, A_2 et A_3 tels que

$$A = z_1A_1 + z_2A_2 + z_3A_3 \quad \text{et} \quad I_n = A_1 + A_2 + A_3.$$

On a alors pour tout $X \in \mathcal{M}_{n,1}(\mathbb{C})$, de norme 1,

$$X^*AX = \sum_{i=1}^{3} z_i X^*A_iX \quad \text{et} \quad 1 = \|X\|^2 = X^*I_nX = \sum_{i=1}^{3} X^*A_iX.$$

Les nombres complexes X^*A_iX sont des réels positifs, puisque A_i est hermitienne positive. Leur somme est égale à 1. L'égalité vérifiée par

X^*AX signifie qu'il appartient à l'enveloppe convexe de $\{z_1, z_2, z_3\}$, c'est-à-dire à l'intérieur du triangle $z_1z_2z_3$. Ceci étant vrai pour tout X de norme 1, F est inclus dans le triangle $z_1z_2z_3$.

Supposons réciproquement qu'il existe trois nombres de module 1 tels que F soit inclus dans le triangle $z_1z_2z_3$. On peut supposer que z_1, z_2 et z_3 ne sont pas alignés, *i.e.* qu'on a un vrai triangle. Sinon F est inclus dans un segment et on peut choisir un troisième nombre complexe de module 1 arbitraire non aligné avec les extrémités du segment. Déterminons trois matrices hermitiennes A_1, A_2, A_3 telles que

$$A_1 + A_2 + A_3 = I_n \text{ et } z_1A_1 + z_2A_2 + z_3A_3 = A.$$

Comme les matrices A_i doivent être hermitiennes, on doit avoir de plus $\overline{z_1}A_1 + \overline{z_2}A_2 + \overline{z_3}A_3 = A^*$. On résout le système constitué de ces trois équations. Les z_i étant distincts, on trouve une solution unique. On a par exemple

$$A_1 = \frac{1}{(z_1 - z_2)(z_1 - z_3)}(z_1A - z_1(z_2 + z_3)I_n + z_1z_2z_3A^*)$$

et des expressions similaires pour A_2 et A_3. On vérifie que A_1 est hermitienne. Les trois nombres z_i sont de module 1. On en déduit que $\overline{(z_1 - z_2)(z_1 - z_3)} = \left(\frac{1}{z_1} - \frac{1}{z_2}\right)\left(\frac{1}{z_1} - \frac{1}{z_3}\right) = \frac{(z_1 - z_2)(z_1 - z_3)}{z_1^2z_2z_3}$, puis que

$$A_1^* = \frac{z_1^2z_2z_3}{(z_1 - z_2)(z_1 - z_3)}\left(\frac{1}{z_1}A^* - \frac{z_2 + z_3}{z_1z_2z_3}I_n + \frac{1}{z_1z_2z_3}A\right) = A_1.$$

On montre de même que A_2 et A_3 sont hermitiennes. Reste à démontrer qu'elles sont positives. Pour tout $X \in \mathcal{M}_{n,1}(\mathbb{C})$ de norme 1, X^*AX est à l'intérieur du triangle $z_1z_2z_3$ donc il existe un triplet de réels positifs $(\lambda_1, \lambda_2, \lambda_3) \in \mathbb{R}_+^3$ tel que $\lambda_1 + \lambda_2 + \lambda_3 = 1$ et $X^*AX = \lambda_1z_1 + \lambda_2z_2 + \lambda_3z_3$. D'autre part, des relations vérifiées par les matrices A_i on tire

$$X^*AX = z_1X^*A_1X + z_2X^*A_2X + z_3X^*A_3X,\ X^*A_1X + X^*A_2X + X^*A_3X = 1.$$

Mais la famille (z_1, z_2, z_3) étant une base affine du plan, X^*AX possède un unique triplet de coordonnées barycentriques. On en déduit que pour $1 \leqslant i \leqslant 3$, on a $X^*A_iX = \lambda_i \geqslant 0$. Ceci est vrai pour tout X de norme 1 et donc en fait pour X quelconque, car si $X \neq 0$, on a $X^*A_iX = \|X\|^2Y^*A_iY$, avec $Y = \frac{1}{\|X\|}X$. Ceci montre que les matrices A_i sont positives.

D'après la question 1, pour démontrer que $\|A\| \leqslant 1$, il suffit de démontrer que B est positive. D'après l'équivalence démontrée précédemment, on a $A = z_1A_1 + z_2A_2 + z_3A_3$, $A^* = \overline{z_1}A_1 + \overline{z_2}A_2 + \overline{z_3}A_3$ et

$I_n = A_1 + A_2 + A_3$. On obtient

$$B = \begin{pmatrix} A_1 + A_2 + A_3 & z_1A_1 + z_2A_2 + z_3A_3 \\ \overline{z_1}A_1 + \overline{z_2}A_2 + \overline{z_3}A_3 & A_1 + A_2 + A_3 \end{pmatrix}$$
$$= \begin{pmatrix} A_1 & z_1A_1 \\ \overline{z_1}A_1 & A_1 \end{pmatrix} + \begin{pmatrix} A_2 & z_2A_2 \\ \overline{z_2}A_2 & A_2 \end{pmatrix} + \begin{pmatrix} A_3 & z_3A_3 \\ \overline{z_3}A_3 & A_3 \end{pmatrix}.$$

Pour démontrer que B est positive, il suffit de démontrer que chacune des matrices hermitiennes $B_i = \begin{pmatrix} A_i & z_iA_i \\ \overline{z_i}A_i & A_i \end{pmatrix}$ l'est. En écrivant comme dans la question 1, $Z = \begin{pmatrix} X \\ Y \end{pmatrix}$, on obtient

$$Z^*B_iZ = X^*A_iX + Y^*A_iY + (z_i + \overline{z_i})X^*A_iY.$$

La matrice A_i étant positive, on peut appliquer l'inégalité de Cauchy-Schwarz et on obtient, puisque z_i est de module 1,

$$|(z_i + \overline{z_i})X^*A_iY| \leqslant 2|X^*A_iY| \leqslant 2\sqrt{X^*A_iX \cdot Y^*A_iY}.$$

On en déduit que

$$Z^*B_iZ \geqslant X^*A_iX + Y^*A_iY - 2\sqrt{X^*A_iX \cdot Y^*A_iY}$$
$$\geqslant (\sqrt{X^*A_iX} - \sqrt{Y^*A_iY})^2 \geqslant 0.$$

La matrice B qui est une somme de trois matrices hermitiennes positives est positive et $\|A\| \leqslant 1$. ◁

2.66. Matrices unitairement congruentes à une matrice triangulaire

Soit $A \in \mathcal{M}_n(\mathbb{C})$. Montrer que cette matrice admet une décomposition $A = UT^tU$, avec U unitaire et T triangulaire supérieure si, et seulement si, le spectre de $A\overline{A}$ est inclus dans $\mathbb{R}_+$.

(École normale supérieure)

▷ **Solution.**

Si $A = UT^tU$, avec U unitaire et $T = (t_{ij})_{1 \leqslant i,j \leqslant n}$ triangulaire supérieure, on a

$$A\overline{A} = UT\,{}^tU\overline{U}\,\overline{T}\,{}^t\overline{U} = UT\overline{T}U^{-1},$$

car ${}^t\overline{U} = U^{-1}$. Les matrices T, $\overline{T}$ et donc $T\overline{T}$ sont triangulaires supérieures. Les termes diagonaux, c'est-à-dire les valeurs propres de $T\overline{T}$ sont

les produits $t_{ii}\overline{t_{ii}} = |t_{ii}|^2$ de ceux de T et $\overline{\mathrm{T}}$. Ils sont positifs. La matrice $\mathrm{A}\overline{\mathrm{A}}$ qui lui est semblable a aussi des valeurs propres réelles positives.

On démontre la réciproque par récurrence sur n. Il n'y a rien à démontrer pour $n = 1$. On suppose que la propriété est vraie au rang $n-1$ ($n \geqslant 2$) et on considère une matrice $\mathrm{A} \in \mathcal{M}_n(\mathbb{C})$ telle que $\mathrm{Sp}(\mathrm{A}\overline{\mathrm{A}}) \subset \mathbb{R}_+$.

Notons que si $\mathrm{A} = \mathrm{U}\,{}^t\mathrm{U} = \mathrm{U}\overline{\mathrm{U}}^{-1}$ et si X_1 est le premier vecteur de la base canonique de $\mathcal{M}_{n,1}(\mathbb{C})$, on a

$$\mathrm{A}(\overline{\mathrm{U}}\mathrm{X}_1) = \mathrm{U}\,{}^t\mathrm{U}\mathrm{X}_1 = t_{11}\mathrm{U}\mathrm{X}_1 = t_{11}\overline{\overline{\mathrm{U}}\mathrm{X}_1}.$$

Le vecteur $\mathrm{X} = \overline{\mathrm{U}}\mathrm{X}_1$ vérifie donc $\mathrm{AX} = t_{11}\overline{\mathrm{X}}$. Mais on a alors

$$\overline{\mathrm{A}}\mathrm{AX} = t_{11}\overline{\mathrm{AX}} = |t_{11}|^2\mathrm{X}.$$

Montrons, pour commencer l'existence d'un tel vecteur X.

Soit $\lambda \in \mathbb{R}_+$ une valeur propre de $\mathrm{A}\overline{\mathrm{A}}$, $\mathrm{X} \neq 0$ un vecteur propre associé. On pose $\mathrm{Y} = \mathrm{A}\overline{\mathrm{X}}$. On a alors

$$\mathrm{A}\overline{\mathrm{A}}\mathrm{Y} = \mathrm{A}\overline{\mathrm{A}}\mathrm{A}\overline{\mathrm{X}} = \mathrm{A}\overline{\mathrm{A}\overline{\mathrm{A}}\mathrm{X}} = \mathrm{A}\overline{\lambda\mathrm{X}} = \lambda\mathrm{A}\overline{\mathrm{X}} = \lambda\mathrm{Y},$$

car λ est réel. Ainsi Y est aussi un vecteur propre de $\mathrm{A}\overline{\mathrm{A}}$ pour la valeur propre λ. Montrons qu'on peut choisir X tel que $\mathrm{A}\overline{\mathrm{X}}$ soit colinéaire à X. Si le vecteur X ci-dessus ne convient pas, la famille (X, Y) est libre et on pose $\mathrm{Z} = \mathrm{Y} + \alpha\mathrm{X}$, où $\alpha = \sqrt{\lambda}$ (ceci est possible car $\lambda \geqslant 0$). On a alors

$$\mathrm{A}\overline{\mathrm{Z}} = \mathrm{A}\overline{\mathrm{Y}} + \alpha\mathrm{A}\overline{\mathrm{X}} = \mathrm{A}\overline{\mathrm{A}}\mathrm{X} + \alpha\mathrm{Y} = \lambda\mathrm{X} + \alpha\mathrm{Y} = \alpha(\mathrm{Y} + \alpha\mathrm{X}) = \alpha\mathrm{Z}.$$

Par ailleurs Z est un vecteur propre de $\mathrm{A}\overline{\mathrm{A}}$ pour la valeur propre λ, car combinaison linéaire de tels vecteurs, non nul car (X, Y) est libre.

En normant le vecteur obtenu précédemment, on trouve donc un vecteur unitaire X et $\mu \in \mathbb{C}$ tels que $\mathrm{A}\overline{\mathrm{X}} = \mu\mathrm{X}$. On considère une matrice unitaire U_1 dont le premier vecteur colonne est X et on pose $\mathrm{B} = \mathrm{U}_1^{-1}\mathrm{A}\,{}^t\mathrm{U}_1{}^{-1} = \mathrm{U}_1^{-1}\mathrm{A}\overline{\mathrm{U}_1}$. On a $\mathrm{U}_1\mathrm{X}_1 = \mathrm{X}$ et donc $\overline{\mathrm{U}_1}\mathrm{X}_1 = \overline{\mathrm{X}}$. On en déduit que $\mathrm{BX}_1 = \mathrm{U}_1^{-1}\mathrm{A}\overline{\mathrm{X}} = \mu\mathrm{U}_1^{-1}\mathrm{X} = \mu\mathrm{X}_1$. La première colonne de B est donc $\begin{pmatrix} \mu \\ 0 \\ \vdots \\ 0 \end{pmatrix}$. Il existe donc $\mathrm{C} \in \mathcal{M}_{n-1}(\mathbb{C})$ et $\mathrm{Y} \in \mathcal{M}_{n,1}(\mathbb{C})$ tels que $\mathrm{B} = \begin{pmatrix} \mu & {}^t\mathrm{Y} \\ 0 & \mathrm{C} \end{pmatrix}$. On en déduit que $\overline{\mathrm{B}} = \begin{pmatrix} \overline{\mu} & {}^t\overline{\mathrm{Y}} \\ 0 & \overline{\mathrm{C}} \end{pmatrix}$ et donc qu'il existe $\mathrm{Z} \in \mathcal{M}_{n,1}(\mathbb{C})$ tel que

$$\mathrm{B}\overline{\mathrm{B}} = \begin{pmatrix} |\mu|^2 & {}^t\mathrm{Z} \\ 0 & \mathrm{C}\overline{\mathrm{C}} \end{pmatrix}.$$

Comme $B\overline{B} = U_1^{-1}A\overline{A}U_1$, le spectre de $B\overline{B}$ est égal à celui de $A\overline{A}$, donc est inclus dans $\mathbb{R}_+$. Mais on a aussi $\mathrm{Sp}(B\overline{B}) = \mathrm{Sp}(C\overline{C}) \cup \{|\mu|^2\}$, donc le spectre de $C\overline{C}$ est inclus dans $\mathbb{R}_+$. D'après l'hypothèse de récurrence, il existe T triangulaire supérieure d'ordre $n-1$ et $V \in U_{n-1}(\mathbb{C})$ tels que $C = VT\,{}^tV$. On pose $U_2 = \begin{pmatrix} 1 & 0 \\ 0 & V \end{pmatrix} \in U_n(\mathbb{C})$. On a alors

$$U_2^{-1}B\,{}^tU_2^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & V^{-1} \end{pmatrix} \begin{pmatrix} \mu & {}^tY \\ 0 & C \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & {}^tV^{-1} \end{pmatrix} = \begin{pmatrix} \mu & {}^t(V^{-1}Y) \\ 0 & T \end{pmatrix}.$$

On note T′ cette matrice triangulaire supérieure. On obtient

$$A = U_1B\,{}^tU_1 = U_1U_2T'\,{}^tU_2\,{}^tU_1 = U_1U_2T'\,{}^t(U_1U_2),$$

ce qui est le résultat voulu puisque $U_1U_2 \in U_n(\mathbb{C})$. ◁

Chapitre 3

Formes quadratiques

Dans tout ce qui suit les espaces vectoriels seront toujours de dimension finie. Si rien n'est précisé, $\mathbb{K}$ *désigne un corps quelconque de caractéristique différente de* 2*, mais la majorité des exercices concerneront les formes quadratiques réelles.*

Nous commençons ce chapitre par quelques exercices sur les formes bilinéaires quelconques. Si b *est une forme bilinéaire symétrique sur un espace vectoriel* E *la relation d'orthogonalité associée à* b *qui est définie par* $x \perp y \iff b(x,y) = 0$ *est évidemment une relation symétrique. Dans l'exercice suivant on demande de trouver toutes les formes bilinéaires qui vérifient cette propriété.*

3.1. Formes bilinéaires réflexives

Déterminer les formes bilinéaires φ sur un espace vectoriel réel E de dimension finie qui vérifient :

$$\forall (x,y) \in \mathrm{E}^2, \quad \varphi(x,y) = 0 \Longrightarrow \varphi(y,x) = 0.$$

(École polytechnique)

Solution.

Les formes bilinéaires symétriques conviennent. Mais ce ne ne sont pas les seules, puisque les formes antisymétriques conviennent aussi. On va voir qu'en réalité ce sont les seules solutions.

Soit φ une forme bilinéaire sur E vérifiant l'hypothèse. Pour tout $x \in \mathrm{E}$ notons $f_x : \mathrm{E} \longrightarrow \mathbb{R}$ et $g_x : \mathrm{E} \longrightarrow \mathbb{R}$ les applications définies par $f_x(y) = \varphi(x,y)$ et $g_x(y) = \varphi(y,x)$ pour tout $y \in \mathrm{E}$. Il s'agit de formes linéaires sur E et par hypothèse $\operatorname{Ker} f_x = \operatorname{Ker} g_x$ (en effet, l'implication de l'hypothèse est en fait une équivalence par symétrie). Il en résulte que f_x et g_x sont proportionnelles : il existe $\lambda_x \in \mathbb{R}$ tel que $g_x = \lambda_x f_x$. Le réel λ_x est unique lorsque f_x est non nulle, mais peut être choisi quelconque lorsque f_x et g_x sont toutes les deux nulles. Autrement dit, si f (resp. g) désigne l'application linéaire de E dans E^* qui à x associe f_x (resp. g_x) on a

$$\forall x \in \mathrm{E}, \quad f(x) = \lambda_x g(x).$$

On va montrer que f et g sont proportionnelles. Pour cela, on utilise le lemme suivant.

Lemme. *Soit* E *et* F *deux* K*-espaces vectoriels, u et v deux éléments de* $\mathcal{L}(\mathrm{E},\mathrm{F})$. *Si pour tout* $x \in \mathrm{E}$ *il existe un scalaire* λ_x *tel que* $u(x) = \lambda_x v(x)$, *alors il existe un scalaire* λ *tel que* $u = \lambda v$.

Démonstration.

Dans le cas où v est bijective, on a alors $v^{-1} \circ u(x) = \lambda_x x$ pour $x \in \mathrm{E}$. Or, d'après un lemme classique (cf. par exemple page 247 du tome 1 d'algèbre pour une preuve), cela implique que $v^{-1} \circ u$ est une homothétie : il existe $\lambda \in \mathbb{K}$ tel que $v^{-1} \circ u = \lambda \mathrm{Id}_\mathrm{E}$ et donc $u = \lambda v$.

Traitons le cas général. La propriété vérifiée par u et v implique en particulier que $\operatorname{Ker} v \subset \operatorname{Ker} u$. Donc d'après le lemme de factorisation (voir l'exercice 6.4. du tome 1 d'algèbre), il existe $h \in \mathcal{L}(\operatorname{Im} v, \mathrm{F})$ telle que $u = h \circ v$. On a pour tout $x \in \mathrm{E}$, $u(x) = h(v(x)) = \lambda_x v(x)$. D'après le lemme classique rappelé ci-dessus, h est une homothétie : il existe $\lambda \in \mathbb{R}$ tel que $u = \lambda v$. $\diamondsuit$

En appliquant le lemme à f et g, on montre l'existence de $\lambda \in \mathbb{R}$ tel que pour tout $x \in \mathrm{E}$, $f_x = \lambda g_x$, c'est-à-dire

$$\forall\, (x,y) \in \mathbb{R}^2,\ \varphi(x,y) = \lambda \varphi(y,x).$$

Il ne reste plus qu'à voir que λ vaut forcément 1 ou -1. En itérant, on a $\lambda\varphi(x,y) = \lambda^2\varphi(x,y)$ pour tout couple (x,y). Si φ est nulle, elle est symétrique ; sinon, on a $\lambda^2 = 1$, d'où le résultat. $\lhd$

On peut noter qu'en fait le résultat reste valable sur un corps quelconque et sur un espace de dimension quelconque avec la même preuve.

L'exercice suivant apporte une généralisation du résultat de l'exercice précédent.

3.2. Caractérisation des applications bilinéaires symétriques ou antisymétriques

Soit E et F deux espaces vectoriels réels de dimension finie et $f : \mathrm{E}^2 \longrightarrow \mathrm{F}$ une application bilinéaire telle que $f(x,y)$ est colinéaire à $f(y,x)$ pour tout couple $(x,y) \in \mathrm{E}^2$. Montrer que f est symétrique ou antisymétrique.

(École normale supérieure)

$\rhd$ **Solution.**

Nous allons faire usage du lemme démontré dans l'exercice précédent.

Pour résoudre l'exercice proposé, il suffit de l'appliquer deux fois de suite. Pour x dans E notons g_x et d_x les applications linéaires de E dans F définies par $g_x(y) = f(x,y)$ et $d_x(y) = f(y,x)$ pour tout $y \in \mathrm{E}$. D'après l'hypothèse, il existe pour tout $y \in \mathrm{E}$ un réel λ_y tel que $g_x(y) = \lambda_y d_x(y)$. D'après le lemme cela impose l'existence d'un réel α_x indépendant de y (mais dépendant *a priori* de x) tel que $g_x = \alpha_x d_x$. Il est maintenant possible d'appliquer à nouveau le lemme mais aux applications linéaires $g : \mathrm{E} \longrightarrow \mathcal{L}(\mathrm{E},\mathrm{F})$ qui à x associe g_x et $d : \mathrm{E} \longrightarrow \mathcal{L}(\mathrm{E},\mathrm{F})$ qui à x associe d_x. Il existe donc un réel μ tel que

$$\forall (x,y) \in \mathrm{E}^2, \quad f(x,y) = \mu f(y,x).$$

On applique cette relation en prenant $y = x$ pour trouver μ. Deux cas se présentent.

- On a $f(x,x) = 0$ pour tout $x \in \mathrm{E}$. Alors le développement

$$f(x+y,x+y) = f(x,x) + f(x,y) + f(y,x) + f(y,y)$$

montre que $f(x,y) = -f(y,x)$ c'est-à-dire que f est antisymétrique.
- Il existe $a \in \mathrm{E}$ tel que $f(a,a) \neq 0$. On a alors forcément $\mu = 1$ et f est symétrique. ◁

Si q est une forme quadratique sur E, *un vecteur x tel que $q(x) = 0$ sera dit isotrope. L'ensemble des vecteurs isotropes,* $\mathrm{C}(q) = q^{-1}(0)$, *est un cône de* E *appelé cône isotrope de q. Bien entendu, pour une forme réelle définie positive (ou négative!) ce cône isotrope est réduit à* $\{0\}$. *Il est remarquable, et c'est l'objet de l'exercice suivant, que lorsque ce cône isotrope n'est pas trivial il contienne toujours une base de* E.

3.3. Base formée de vecteurs isotropes

Soit E un espace réel de dimension finie $n \geqslant 2$, q une forme quadratique non dégénérée sur E. On suppose qu'il existe $x \in \mathrm{E}$ non nul tel que $q(x) = 0$.

1. Montrer qu'il existe $y \in \mathrm{E}$ tel que (x,y) soit libre et $q(y) = 0$.

2. Montrer l'existence d'une base $(e_i)_{1 \leqslant i \leqslant n}$ de E telle que $q(e_i) = 0$ pour tout $i \in [\![1,n]\!]$.

(École polytechnique)

Solution.

1. Notons b la forme polaire de q. Comme q n'est pas dégénérée, il existe un vecteur $z \in \mathrm{E}$ tel que $b(x,z) \neq 0$. Comme $q(x) = b(x,x) = 0$ le vecteur z n'est pas colinéaire à x. On va chercher y dans le plan $\mathrm{Vect}(x,z)$

sous la forme $y = z + tx$. On a $q(y) = b(z + tx, z + tx) = b(z, z) + 2tb(x, z)$ de sorte qu'il suffit de prendre $t = -\dfrac{b(z,z)}{2b(x,z)}$ pour que $q(y) = 0$. Le vecteur y obtenu est clairement non colinéaire à x.

2. Prenons naturellement $e_1 = x$. On va utiliser la même idée que dans la question précédente : il suffit de prouver l'existence d'une base $(e_1, f_2, \ldots, f_n)$ de E telle que $b(e_1, f_k) \neq 0$ pour tout $k \in [\![2, n]\!]$. En effet, la question précédente montre l'existence dans chaque plan $\mathrm{Vect}(e_1, f_k)$ d'un vecteur isotrope e_k de la forme $f_k + t_k e_1$. Il est alors clair que $(e_1, \ldots, e_n)$ est une base de E formée de vecteurs isotropes. Passons à la preuve de l'existence d'une telle base. L'application $y \longmapsto b(e_1, y)$ est une forme linéaire non nulle sur E. Son noyau est un hyperplan de E (c'est l'orthogonal au sens de b de la droite $\mathbb{R}e_1$). Cet hyperplan contient la droite $\mathbb{R}e_1$, car e_1 est isotrope. Prenons alors une base de cet hyperplan de la forme $(e_1, g_2, \ldots, g_{n-1})$ et rajoutons le vecteur z (où z est, comme dans la question 1, un vecteur tel que $b(e_1, z) \neq 0$)) pour obtenir une base de E. Posons $f_k = g_k + z$ pour tout $k \in [\![2, n-1]\!]$ et $f_n = z$. Il est évident que la famille $(e_1, f_2, \ldots, f_n) = (e_1, g_2 + z, \ldots, g_{n-1} + z, z)$ reste une base de E. Et de plus $b(e_1, f_k) = b(e_1, z) \neq 0$ pour tout k. D'où le résultat. ◁

Dans l'exercice suivant, on regarde dans quelle mesure le cône isotrope caractérise une forme quadratique sur un espace vectoriel complexe.

3.4. Formes quadratiques ayant même cône isotrope

Soit q, q' deux formes quadratiques sur un espace vectoriel complexe E telles que $q(x) = 0$ si et seulement si $q'(x) = 0$. Que dire de q et q' ?

(École polytechnique)

▷ **Solution.**

Nous allons montrer que q' et q sont proportionnelles. Si q est nulle alors q' est aussi nulle et c'est terminé. Sinon il existe $x \in \mathrm{E}$ tel que $q(x) \neq 0$. Posons $\lambda = \dfrac{q'(x)}{q(x)}$ (λ est non nul) et considérons y un vecteur quelconque de E. Pour tout réel t on

$$q(y + tx) = q(y) + 2tb(x, y) + q(x)t^2$$

et

$$q'(y + tx) = q'(y) + 2tb'(x, y) + q'(x)t^2 = q'(y) + 2tb'(x, y) + \lambda q(x)t^2$$

où b et b' désignent les formes polaires de q et q' respectivement. Par hypothèse les deux trinômes ci-dessus ont les mêmes racines dans $\mathbb{C}$. Ils sont donc proportionnels et en particulier $q'(y) = \lambda q(y)$. On a donc $q' = \lambda q$ comme annoncé. La réciproque est évidente. ◁

Bien entendu le résultat ne reste pas vrai sur $\mathbb{R}$, car les trinômes n'y sont pas forcément scindés. Par exemple $q(x, y, z) = x^2 + y^2$ et $q'(x, y, z) = x^2 + 2y^2$ sont deux formes quadratiques sur $\mathbb{R}^3$ ayant toutes les deux la droite $x = y = 0$ pour cône isotrope mais elles ne sont pas proportionnelles. Toutefois si on suppose que les deux formes quadratiques sont non dégénérées et ont un même cône isotrope non nul alors on peut prouver qu'elles restent proportionnelles, et ce sur n'importe quel corps $\mathbb{K}$ de caractéristique différente de 2.

On déduit immédiatement de l'exercice que si u est un automorphisme de $\mathbb{E}$ qui conserve le cône isotrope de q alors u est une similitude, c'est-à-dire de la forme λv avec $\lambda \in \mathbb{C}^*$ et v dans le groupe orthogonal de q (voir page 226 la définition du groupe orthogonal).

3.5. Composantes connexes par arcs

Soit $n \in \mathbb{N}^*$ et q une forme quadratique sur $\mathbb{R}^n$. On considère $\Sigma_q = \{x \in \mathbb{R}^n,\ \|x\| = 1,\ q(x) = 0\}$ où $\|x\|$ est la norme euclidienne canonique de x.

1. Décrire Σ_q lorsque $n \in \{1, 2, 3\}$.

2. On suppose $n \geqslant 3$. Soient x, y dans Σ_q. Montrer que y est dans la composante connexe par arcs de x ou de $-x$ dans Σ_q.

(École normale supérieure)

Solution.

L'ensemble Σ_q est l'intersection du cône isotrope de q et de la sphère unité de $\mathbb{R}^n$. Il est symétrique par rapport à l'origine. Par homogénéité de q, il est évident que tout vecteur non nul du cône isotrope est relié à un point de Σ_q par un segment. On en déduit aisément que le nombre de composantes connexes par arcs[1] de $C(q) \setminus \{0\}$ est égal au nombre de composantes connexes par arcs de Σ_q. L'objet de l'exercice est de montrer que pour $n \geqslant 3$, l'ensemble Σ_q est connexe par arcs ou admet deux composantes connexes par arcs opposées.

1. Quelques remarques générales pour commencer.

Il existe une base orthonormale $(\varepsilon_1, \ldots, \varepsilon_n)$ de $\mathbb{R}^n$ dans laquelle la matrice A de q est diagonale. On obtient $q(x_1\varepsilon_1 + \cdots + x_n\varepsilon_n) = \sum_{i=1}^{n} a_i x_i^2$, où $(a_1, \ldots, a_n)$ sont les valeurs propres de A. Si r (resp. s) est le nombre

[1] La définition de cette notion est rappelée page 68

de valeurs propres strictement positives (resp. négatives) de A, il existe une base $(e_1, \ldots, e_n)$ de $\mathbb{R}^n$ dans laquelle, pour tout $(x_1, \ldots, x_n) \in \mathbb{R}^n$,

$$q(x_1e_1 + \cdots + x_ne_n) = \sum_{i=1}^{r} x_i^2 - \sum_{i=r+1}^{r+s} x_i^2$$

(on pose $e_i = \frac{1}{\sqrt{a_i}}\varepsilon_i$ si $a_i > 0$, $e_i = \frac{1}{\sqrt{-a_i}}\varepsilon_i$ si $a_i < 0$). Pour l'étude de Σ_q, on se placera dorénavant dans une telle base.

Le rang de q est alors $r + s$. Notons aussi que comme q et $-q$ ont le même cône isotrope on peut se limiter au cas $r \geqslant s$.

Lorsque q est nulle, son cône isotrope est égal à $\mathbb{R}^n$ tout entier et Σ_q est la sphère unité de $\mathbb{R}^n$ qui est connexe par arcs dès que $n \geqslant 2$. Lorsque q est définie positive ou négative Σ_q est vide. Cela recouvre tous les cas lorsque $n = 1$. Étudions les cas restants pour $n = 2$ ou 3.

• Supposons $n = 2$ et q non dégénérée et non définie ($r = s = 1$). Elle s'écrit $q(xe_1 + x_2e_2) = x^2 - y^2$: son cône isotrope est la réunion de deux droites distinctes et Σ_q est donc un ensemble de 4 points deux à deux symétriques par rapport à l'origine.

Toujours pour $n = 2$ supposons q de rang 1. Quitte à prendre $-q$ on peut supposer $q(xe_1 + ye_2) = x^2$. Alors $C(q)$ est une droite et Σ_q est un ensemble de 2 points symétriques par rapport à l'origine.

• Supposons enfin $n = 3$. On discute encore selon le rang de q en excluant les cas triviaux mentionnés au début.

⋆ Si q est de rang 1 et si $r = 1$, $s = 0$, $q(xe_1 + ye_2 + ze_3) = x^2$ et $C(q)$ est un plan. Par suite Σ_q est un (grand) cercle de la sphère unité et est donc connexe.

⋆ Si q est de rang 2 et si $r = 2$, $s = 0$, $q(xe_1 + ye_2 + ze_3) = x^2 + y^2$ et $C(q)$ est une droite : dans ce cas Σ_q est un ensemble de 2 points symétriques par rapport à l'origine.

⋆ Si q est de rang 2 et si $r = s = 1$, alors $q(xe_1 + ye_2 + ze_3) = x^2 - y^2$ et $C(q)$ est la réunion de deux plans. Par suite Σ_q est la réunion de deux grands cercles de la sphère unité. C'est un ensemble connexe par arcs.

⋆ Si enfin q est de rang 3 et $r = 2$, $s = 1$, on a alors $q(xe_1 + ye_2 + ze_3) = x^2 + y^2 - z^2$ et $C(q)$ est un vrai cône de $\mathbb{R}^3$. Géométriquement, Σ_q est la réunion de deux courbes fermées symétriques par rapport à l'origine et possède donc deux composantes connexes par arcs. Pour prouver cela analytiquement on peut, en se plaçant dans une base orthonormale qui diagonalise A, supposer que $q(xe_1 + ye_2 + ze_3) = \frac{x^2}{a^2} + \frac{y^2}{b^2} - \frac{z^2}{c^2}$ l'équation de la sphère étant toujours $x^2 + y^2 + z^2 = 1$. En éliminant z, on voit que (x, y, z) est dans Σ_q si, et seulement si, il existe un réel $t \in [0, 2\pi]$ et $\varepsilon \in \{\pm 1\}$ tels que

$$x = \frac{a}{\sqrt{a^2+c^2}}\cos t,\quad y = \frac{b}{\sqrt{b^2+c^2}}\sin t \quad \text{et}$$

$$z = \varepsilon\sqrt{\frac{c^2}{a^2+c^2}\cos^2 t + \frac{c^2}{b^2+c^2}\sin^2 t}.$$

On obtient donc bien la réunion de deux courbes fermées comme annoncé.

2. Soient x, y dans Σ_q. On peut évidemment supposer y distinct de x et de $-x$. Dans ce cas la famille (x, y) est libre. Choisissons un vecteur z quelconque en dehors du plan P, ce qui est possible car $n \geqslant 3$. La question précédente montre que $\mathrm{Vect}(x, y, z) \cap \Sigma_q$ est connexe par arcs ou possède deux composantes connexes par arcs symétriques par rapport à l'origine. On en déduit que y est soit dans la composante connexe par arcs de x soit dans celle de $-x$. Il en est de même pour les composantes connexes par arcs de Σ_q. ◁

La solution s'écrit plus simplement avec la notion de signature qui n'est plus au programme des classes préparatoires (cf exercices 3.22 et suivants).

Il est classique de prouver qu'un polynôme positif de $\mathbb{R}[X]$ peut s'écrire comme une somme de deux carrés (voir l'exercice 5.16 – ou 5.20 dans la seconde édition – du tome Algèbre 1). Le résultat est faux pour des polynômes à plusieurs variables : le contre-exemple de Motzkin $P(X,Y) = X^2Y^2(X^2+Y^2-3)+1$ a d'ailleurs été proposé récemment à l'oral de l'École normale supérieure. Un résultat difficile affirme toutefois que tout polynôme positif à plusieurs variable peut s'écrire comme une somme de carrés de fractions rationnelles (c'est le dix-septième problème de Hilbert). L'exercice suivant concerne, dans le cas d'une variable, le passage d'une décomposition en fractions à une décomposition en polynômes.

3.6. Théorème de Pfister (1965)

Soit K un corps commutatif de caractéristique différente de 2, $\alpha_1, \ldots, \alpha_n$ dans K^*. On suppose qu'il existe des fractions rationnelles non toutes nulles $r_1, \ldots, r_n$ de $K(X)$ telles que $\alpha_1 r_1^2 + \cdots + \alpha_n r_n^2 = 0$. Montrer que pour tout polynôme $f \in K[X]$ il existe des polynômes $f_1, \ldots, f_n$ de $K[X]$ tels que

$$f = \alpha_1 f_1^2 + \cdots + \alpha_n f_n^2.$$

(École normale supérieure)

▷ **Solution.**

Posons $k = \mathrm{K}(\mathrm{X})$, et notons Q la forme quadratique définie sur k^n par $\mathrm{Q}(x_1, x_2, \ldots, x_n) = \sum_{i=1}^{n} \alpha_i x_i^2$ et φ sa forme polaire. En multipliant éventuellement les fractions r_i par un dénominateur commun et en divisant les polynômes obtenus par leur pgcd, on peut se ramener au cas où $r_1, \ldots, r_n$ sont des polynômes premiers entre eux. Par hypothèse le vecteur $r = (r_1, r_2, \ldots, r_n)$ est isotrope, *i.e.* vérifie $\mathrm{Q}(r) = 0$. Il s'agit de démontrer que $\mathrm{Q}(\mathrm{K}[\mathrm{X}]^n) = \mathrm{K}[\mathrm{X}]$.

On a $\mathrm{pgcd}(\alpha_1 r_1, \ldots, \alpha_n r_n) = \mathrm{pgcd}(r_1, \ldots, r_n) = 1$ et par l'identité de Bezout, il existe des polynômes $s_1, \ldots, s_n$ tels que $\sum_{i=1}^{n} \alpha_i r_i s_i = 1$, c'est-à-dire $\varphi(r, s) = 1$, où $s = (s_1, \ldots, s_n)$. Posons $t = s - \frac{1}{2}\mathrm{Q}(s)r$. Comme $\mathrm{Q}(s) \in \mathrm{K}[\mathrm{X}]$, $t \in \mathrm{K}[\mathrm{X}]^n$ et de plus on a

$$\mathrm{Q}(t) = \mathrm{Q}(s) + \frac{1}{4}\mathrm{Q}(s)\mathrm{Q}(r) - \mathrm{Q}(s)\varphi(r, s) = 0 \quad \text{et} \quad \varphi(r, t) = \varphi(r, s) = 1.$$

Soit $f \in \mathrm{K}[\mathrm{X}]$. On a

$$\mathrm{Q}\left(r + \frac{f}{2}t\right) = \mathrm{Q}(r) + \frac{f^2}{4}\mathrm{Q}(t) + f\varphi(r, t) = f$$

et $r + \frac{f}{2}t \in \left(\mathrm{K}[\mathrm{X}]\right)^n$. On a donc bien $\mathrm{Q}(\mathrm{K}[\mathrm{X}]^n) = \mathrm{K}[\mathrm{X}]$. ◁

Il ne s'agit en réalité que du cas simple du théorème de Pfister : celui-ci affirme que si $f \in \mathrm{K}[\mathrm{X}]$ peut s'écrire $f = \alpha_1 r_1^2 + \cdots + \alpha_n r_n^2$ avec $r_1, \ldots, r_n$ dans $\mathrm{K}(\mathrm{X})$, alors f peut aussi s'écrire $f = \alpha_1 f_1^2 + \cdots + \alpha_n f_n^2$ avec $f_1, \ldots, f_n$ dans $\mathrm{K}[\mathrm{X}]$. Le cas où la forme quadratique est anisotrope est un peu plus difficile.

Notons que la solution de l'exercice montre que $\mathrm{Q}(k^n) = k$. C'est un résultat général : si q est une forme quadratique non dégénérée sur un k-espace vectoriel V, où le corps k est de caractéristique différente de 2, et si q possède un vecteur isotrope non nul, alors $q(\mathrm{V}) = k$. Si $q(x) = 0$, comme q est non dégénérée, on peut trouver $y \in \mathrm{V}$ tel que $\varphi(x, y) = 1$, puis l'on construit z tel que $q(z) = 0$ et $\varphi(x, z) = 1$. On conclut comme dans l'exercice.

L'énoncé suivant utilise les principales propriétés de l'orthogonalité au sens d'une forme bilinéaire symétrique b. Rappelons que si A *est une partie de* E *son orthogonal est* $\mathrm{A}^\perp = \{x \in \mathrm{E},\ \forall a \in \mathrm{A},\ b(x, a) = 0\}$. *C'est un sous-espace vectoriel de* E. *Rappelons aussi que b est dite non dégénérée lorsque l'application de* E *dans son dual* E^* *qui à tout vecteur x associe la forme linéaire $b_x : y \longmapsto b(x, y)$ est un isomorphisme. Si*

A est la matrice de b dans une base quelconque de E *cela se caractérise simplement par l'inversibilité de* A.

3.7. Indice d'une forme quadratique

Soit q une forme quadratique non dégénérée sur un $\mathbb{K}$-espace vectoriel de dimension finie E ($\mathbb{K}$ étant un corps quelconque de caractéristique *différente de* 2).

1. Soit F un sous-espace vectoriel de E. Montrer que

$$\dim \mathrm{F}^{\perp} = \dim \mathrm{E} - \dim \mathrm{F}.$$

2. Un sous-espace F est dit totalement isotrope lorsque la restriction de q à F est nulle, autrement dit lorsque $\mathrm{F} \subset \mathrm{F}^{\perp}$. On qualifiera de SETIM un sous-espace totalement isotrope qui est maximal au sens de l'inclusion. Montrer que tout sous-espace totalement isotrope est inclus dans un SETIM.

3. Montrer que tous les SETIM de E ont la même dimension $\nu(q)$ (qui est appelée *l'indice de* q). Vérifier que $2\nu(q) \leqslant \dim \mathrm{E}$ et donner un exemple où il y a égalité.

(École normale supérieure)

▷ **Solution.**

1. Notons b la forme polaire de q, n la dimension de E et soit F un sous-espace de dimension p. On peut supposer $p \geqslant 1$, car le résultat est clair si F est nul. Soit $(f_1, \ldots, f_p)$ une base de F. Comme q est non dégénérée, les formes linéaires $\varphi_i : x \longmapsto b(f_i, x)$ pour $1 \leqslant i \leqslant p$ forment une famille libre de E^* que nous pouvons donc compléter en une base $(\varphi_1, \ldots, \varphi_p, \varphi_{p+1}, \ldots, \varphi_n)$ de E^*. Soit $(e_1, \ldots, e_n)$ la base antéduale de cette base. On a

$$\begin{aligned} x \in \mathrm{F}^{\perp} &\iff \forall i \in [\![1, p]\!],\ x \perp f_i \\ &\iff \forall i \in [\![1, p]\!],\ \varphi_i(x) = 0 \\ &\iff x \in \mathrm{Vect}(e_{p+1}, \ldots, e_n). \end{aligned}$$

Le résultat en découle.

On prendra garde que, contrairement au cas euclidien, $\mathrm{F}^{\perp}$ *n'est pas forcément en somme directe avec* F. *En effet, la restriction de* q *à* F *peut être dégénérée.*

2. Soit F un sous-espace isotrope. Parmi tous les sous-espaces isotropes qui contiennent F (et il en existe puisqu'il y a au moins F lui-même), il suffit d'en prendre un de dimension maximale. Il s'agit forcément d'un *SETIM*.

3. Soient F et G deux SETIM de E. Considérons F_1 (resp. G_1) un supplémentaire de $F \cap G$ dans F (resp. dans G). On a donc

$$F = (F \cap G) \dotplus F_1 \quad \text{et} \quad G = (F \cap G) \dotplus G_1$$

et il nous suffit de prouver que F_1 et G_1 ont la même dimension. Pour cela il nous suffit de prouver que $G_1 \cap F_1^\perp = \{0\}$. En effet cela implique $\dim G_1 + \dim F_1^\perp \leqslant \dim E$ et donc $\dim G_1 \leqslant \dim F_1$ et la symétrie du problème montre qu'il y a alors forcément égalité.

Supposons donc par l'absurde qu'il existe un vecteur x non nul dans $G_1 \cap F_1^\perp$. Il s'agit évidemment d'obtenir une contradiction avec la maximalité de F ou de G. Comme x est dans G on va plutôt prendre F. Effectivement $x \notin F$, car sinon x serait dans $F \cap G$ et dans G_1 donc serait nul. Mais $x \in F^\perp$. En effet, si $y \in F$ on peut le décomposer sous la forme $y = y_1 + y_2$ avec $y_1 \in F \cap G$ et $y_2 \in F_1$. Or $b(x, y_1) = 0$, car x et y_1 sont tous les deux dans G et $b(x, y_2) = 0$, car $x \in F_1^\perp$ et $y_2 \in F_1$. Donc $b(x, y) = b(x, y_1) + b(x, y_2) = 0$. Il est alors facile de voir que le sous-espace $F' = F \oplus \mathbb{K}x$ est encore totalement isotrope ce qui contredit la maximalité de F.

Notons $\nu(q)$ la dimension des SETIM et soit F l'un d'entre eux. Comme $F \subset F^\perp$ on a $\dim F \leqslant \dim F^\perp = \dim E - \dim F$ et donc $2\nu(q) \leqslant \dim E$. Pour avoir égalité E doit être de dimension paire. Prenons par exemple $\mathbb{R}^{2p}$ et q la forme quadratique définie par

$$q(x_1, \ldots, x_{2p}) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_{2p}^2.$$

En notant $(e_1, \ldots, e_{2p})$ la base canonique de $\mathbb{R}^{2p}$ il est clair que $e_1 + e_{p+1}$, $e_2 + e_{p+2}, \ldots, e_p + e_{2p}$ sont isotropes et deux à deux orthogonaux. Ils engendrent donc un sous-espace totalement isotrope de dimension p qui est, d'après ce qui précède, forcément maximal. ◁

De manière plus générale, le lecteur pourra montrer que sur $\mathbb{R}^n$ l'indice d'une forme quadratique non dégénérée de signature (r, s) est égal à $\min(r, s)$.

Il est facile de voir par récurrence sur la dimension que toute forme quadratique q admet une base orthogonale : si q est nulle c'est évident et sinon on choisit e_1 tel que $q(e_1) \neq 0$. Alors $H = e_1^\perp$ est un hyperplan qui ne contient pas e_1 et il suffit d'appliquer l'hypothèse de récurrence à la restriction de q à H pour compléter e_1 en une base orthogonale. Cette existence de bases orthogonales est essentielle (mais insuffisante) pour le problème de la classification des formes quadratiques (voir page 212). Matriciellement cela montre que si $A \in \mathcal{S}_n(\mathbb{K})$ est une matrice symétrique alors A est congruente à une matrice diagonale : il existe $P \in \mathrm{GL}_n(\mathbb{K})$ et D diagonale telles que ${}^tPAP = D$. Lorsque $\mathbb{K} = \mathbb{R}$ nous savons par le théorème spectral que P peut être choisie orthogonale.

3.8. Famille de formes quadratiques

Soit $a \in \mathbb{R}$. Pour $(x, y, z) \in \mathbb{R}^3$, on pose

$$q_a(x, y, z) = a(x^2 + y^2 + z^2) - (2xy + 2xz + 2yz).$$

1. Pour quelles valeurs de a, la forme quadratique q_a est-elle non dégénérée ?

2. Trouver une base de $\mathbb{R}^3$ orthogonale pour toutes les q_a.

(**École polytechnique**)

▷ **Solution.**

1. La matrice de q_a dans la base canonique de $\mathbb{R}^3$ est

$$\mathrm{M}_a = \begin{pmatrix} a & -1 & -1 \\ -1 & a & -1 \\ -1 & -1 & a \end{pmatrix}.$$

La forme quadratique q_a est non dégénérée si et seulement si $\det \mathrm{M}_a \neq 0$. Or $\det \mathrm{M}_a = \det(a\mathrm{I}_3 - \mathrm{A}) = -\chi_{\mathrm{A}}(a)$ où A est la matrice $\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$. Il est clair que -1 est valeur propre de A et que l'espace propre associé est de dimension 2. Comme $\mathrm{Tr}\,\mathrm{A} = 0$ on en déduit que 2 est aussi valeur propre de A. Ainsi $\det \mathrm{M}_a = -\chi_{\mathrm{A}}(a) = (a+1)^2(a-2)$.

Conclusion. La forme quadratique q_a est non dégénérée si, et seulement si, $a \notin \{2, -1\}$.

2. Une base orthonormée (pour la structure euclidienne canonique de $\mathbb{R}^3$) formée de vecteurs propres de M_a constitue une base orthogonale pour la forme quadratique q_a. Or une base orthonormée de vecteurs propres de A est une base de vecteurs propres de M_a pour tout réel a. La droite $\mathrm{Ker}(\mathrm{A} - 2\mathrm{I}_3)$ est dirigée par le vecteur unitaire $f_1 = \frac{1}{\sqrt{3}}(1, 1, 1)$. Le plan propre de A pour la valeur propre -1 a pour équation $x + y + z = 0$. Une base orthonormée de ce plan est donnée par exemple par $f_2 = \frac{1}{\sqrt{2}}(1, -1, 0)$ et $f_3 = \frac{1}{\sqrt{6}}(1, 1, -2)$. La base (f_1, f_2, f_3) de $\mathbb{R}^3$ répond donc à la question. ◁

Rappelons que le rang d'une forme quadratique q sur E *est par définition le rang de l'application de* E *dans son dual* E* *qui à tout vecteur x associe la forme linéaire $b_x : y \longmapsto b(x, y)$. C'est aussi le rang de toute matrice* A *représentant q dans une base quelconque de* E. *Le résultat de l'exercice suivant est utilisé en Probabilités.*

3.9. Théorème de Fischer-Cochran

Soit E un espace euclidien de dimension n et m formes quadratiques $q_1, \dots, q_m$ sur E vérifiant $q_1 + \cdots + q_m = q$ où $q(x) = (x, x)$ est la forme euclidienne et $r_1 + \cdots + r_m = n$ où r_k désigne le rang de q_k. Montrer que E est somme directe orthogonale de sous-espaces $\mathrm{E}_1, \dots, \mathrm{E}_m$ tels que si p_k désigne le projecteur orthogonal sur E_k alors

$$\forall x \in \mathrm{E}. \quad q_k(x) = (p_k(x), x) = (p_k(x), p_k(x)).$$

(École polytechnique)

▷ **Solution.**

Voici une autre manière de comprendre le résultat demandé : on cherche à montrer l'existence d'une base orthonormée $(e_1, \dots, e_n)$ de E et d'un partage $\mathrm{I}_1, \dots, \mathrm{I}_m$ de l'ensemble $[\![1, n]\!]$ tel que pour tout $k \in [\![1, m]\!]$ et tout $x = \sum_{i=1}^{n} x_i e_i$ on ait $q_k(x) = \sum_{i \in \mathrm{I}_k} x_i^2$.

Nous savons que chaque forme q_k peut se représenter à l'aide d'un endomorphisme auto-adjoint u_k de E : $q_k(x) = (u_k(x), x)$ pour tout $x \in \mathrm{E}$. Les deux hypothèses se traduisent alors par

$$(1) \quad u_1 + \cdots + u_m = \mathrm{Id}_\mathrm{E} \quad \text{et} \quad (2) \quad \operatorname{rg} u_1 + \cdots + \operatorname{rg} u_m = n = \dim \mathrm{E}.$$

Posons $\mathrm{E}_k = \operatorname{Im} u_k$ pour tout $k \in [\![1, m]\!]$. L'hypothèse (1) implique que $\mathrm{E} = \mathrm{E}_1 + \mathrm{E}_2 + \cdots + \mathrm{E}_m$. Mais comme la somme des dimensions des E_k vaut exactement n par (2) on peut affirmer que la somme est directe : $\mathrm{E} = \mathrm{E}_1 \oplus \mathrm{E}_2 \oplus \cdots \oplus \mathrm{E}_m$. De plus, l'égalité (1) montre que u_k est exactement le projecteur sur E_k parallèlement à $\bigoplus_{i \neq k} \mathrm{E}_i$. Pour conclure il ne reste plus qu'à prouver que les E_k sont orthogonaux deux à deux et cela découle de ce que u_k est auto-adjoint (un projecteur d'un espace euclidien est orthogonal si, et seulement si, il est auto-adjoint). D'où le résultat. ◁

La plupart des exercices de ce chapitre sont consacrées aux formes quadratiques réelles. Avant d'aborder la classification de ces formes à l'aide de la notion de signature (récemment éliminée des programmes des classes préparatoires) nous proposons plusieurs exercices sur la notion de positivité. Rappelons qu'une forme quadratique q sur un espace vectoriel réel est dite positive (resp. définie positive) lorsque $q(x) \geqslant 0$ pour tout $x \in \mathrm{E}$ (resp. $q(x) > 0$ pour tout x non nul de E). Une matrice symétrique réelle A sera dite positive (définie positive) lorsque la forme quadratique q_A de $\mathbb{R}^n$ qui lui est canoniquement associée est positive (resp. définie

positive). Ces propriétés se lisent sur le spectre de A : A est positive (resp. définie positive) si, et seulement si, ses valeurs propres sont toutes positives (resp. strictement positives).

3.10. Une forme quadratique définie positive

Soit $(a_1, \ldots, a_n, b_1, \ldots, b_n)$ une famille de $2n$ réels strictement positifs. Pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on pose

$$q(x) = \sum_{1 \leqslant i,j \leqslant n} \frac{x_i x_j}{a_i b_j + a_j b_i}.$$

À quelle condition la forme quadratique q est-elle définie positive ?

(École polytechnique)

▷ **Solution.**

Un changement de base très simple permet de simplifier q. Posons $y_i = \dfrac{x_i}{a_i}$ pour tout i. On a alors

$$q(x) = \sum_{1 \leqslant i,j \leqslant n} \frac{a_i a_j y_i y_j}{a_i b_j + a_j b_i} = \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{\dfrac{b_i}{a_i} + \dfrac{b_j}{a_j}} = \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{c_i + c_j},$$

où $c_i = \dfrac{b_i}{a_i}$.

On remarque alors, et c'est la clé de l'exercice, que

$$\begin{aligned} \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{c_i + c_j} &= \sum_{1 \leqslant i,j \leqslant n} y_i y_j \int_0^1 t^{c_i + c_j - 1} \mathrm{d}t \\ &= \int_0^1 \left(\sum_{1 \leqslant i,j \leqslant n} y_i y_j t^{c_i - \frac{1}{2}} t^{c_j - \frac{1}{2}} \right) \mathrm{d}t \\ &= \int_0^1 \left(\sum_{i=1}^n y_i t^{c_i - \frac{1}{2}} \right)^2 \mathrm{d}t \geqslant 0. \end{aligned}$$

Comme la fonction $t \longmapsto \sum_{i=1}^n y_i t^{c_i - \frac{1}{2}}$ est continue sur $]0, 1]$, il y a égalité si, et seulement si, $\sum_{i=1}^n y_i t^{c_i - \frac{1}{2}} = 0$ pour tout $t \in]0, 1]$. La forme quadratique q sera définie positive si cela implique $y_1 = \cdots = y_n = 0$. Deux cas se présentent :

• Si les c_i ne sont pas distincts, alors q n'est pas définie positive. Si par exemple, on a $c_1 = c_2$, on annule q en prenant $y_1 = 1$, $y_2 = -1$ et $y_3 = \cdots = y_n = 0$.

positive). Ces propriétés se lisent sur le spectre de A : A est positive (resp. définie positive) si, et seulement si, ses valeurs propres sont toutes positives (resp. strictement positives).

3.10. Une forme quadratique définie positive

Soit $(a_1, \ldots, a_n, b_1, \ldots, b_n)$ une famille de $2n$ réels strictement positifs. Pour $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, on pose

$$q(x) = \sum_{1 \leqslant i,j \leqslant n} \frac{x_i x_j}{a_i b_j + a_j b_i}.$$

À quelle condition la forme quadratique q est-elle définie positive ?

(École polytechnique)

▷ **Solution.**

Un changement de base très simple permet de simplifier q. Posons $y_i = \dfrac{x_i}{a_i}$ pour tout i. On a alors

$$q(x) = \sum_{1 \leqslant i,j \leqslant n} \frac{a_i a_j y_i y_j}{a_i b_j + a_j b_i} = \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{\dfrac{b_i}{a_i} + \dfrac{b_j}{a_j}} = \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{c_i + c_j},$$

où $c_i = \dfrac{b_i}{a_i}$.

On remarque alors, et c'est la clé de l'exercice, que

$$\begin{aligned} \sum_{1 \leqslant i,j \leqslant n} \frac{y_i y_j}{c_i + c_j} &= \sum_{1 \leqslant i,j \leqslant n} y_i y_j \int_0^1 t^{c_i + c_j - 1} \mathrm{d}t \\ &= \int_0^1 \left(\sum_{1 \leqslant i,j \leqslant n} y_i y_j t^{c_i - \frac{1}{2}} t^{c_j - \frac{1}{2}} \right) \mathrm{d}t \\ &= \int_0^1 \left(\sum_{i=1}^n y_i t^{c_i - \frac{1}{2}} \right)^2 \mathrm{d}t \geqslant 0. \end{aligned}$$

Comme la fonction $t \longmapsto \sum_{i=1}^n y_i t^{c_i - \frac{1}{2}}$ est continue sur $]0, 1]$, il y a égalité si, et seulement si, $\sum_{i=1}^n y_i t^{c_i - \frac{1}{2}} = 0$ pour tout $t \in]0, 1]$. La forme quadratique q sera définie positive si cela implique $y_1 = \cdots = y_n = 0$. Deux cas se présentent :

• Si les c_i ne sont pas distincts, alors q n'est pas définie positive. Si par exemple, on a $c_1 = c_2$, on annule q en prenant $y_1 = 1$, $y_2 = -1$ et $y_3 = \cdots = y_n = 0$.

• Si les c_i sont deux à deux distincts, on peut supposer sans perte de généralité que $c_1 < \cdots < c_n$. On note $f(t) = \sum_{i=1}^{n} y_i t^{c_i - \frac{1}{2}}$ et on suppose que $f(t) = 0$ pour tout $t \in]0,1]$. Si $y_1, \ldots, y_n$ ne sont pas tous nuls, on considère le plus petit indice k tel que $y_k \neq 0$. On a alors $f(t) \underset{t \to 0}{\sim} y_k t^{c_k - \frac{1}{2}}$ ce qui contredit le fait que $f(t) = 0$ pour tout $t \in]0,1]$. On a donc $y = 0$ et $x = 0$.

Conclusion. La forme q est définie positive si et seulement les réels c_i sont deux à deux distincts, autrement dit si, et seulement si, $\frac{a_1}{b_1}, \ldots, \frac{a_n}{b_n}$ sont deux à deux distincts distincts. ◁

3.11. Une forme quadratique positive

Soit $\mathrm{A} = \sum_{k=0}^{2n} a_k \mathrm{X}^k \in \mathbb{R}_{2n}[\mathrm{X}]$. On considère l'application δ_A de $\mathbb{R}_{2n}[\mathrm{X}]$ dans lui-même définie par $\delta_\mathrm{A}(\mathrm{P}) = \sum_{k=0}^{2n} \frac{a_k}{k!} \mathrm{P}^{(k)}$. Montrer que les propriétés suivantes sont équivalentes :

(i) $\forall \mathrm{P} \in \mathbb{R}_{2n}[\mathrm{X}],\ \mathrm{P}(\mathbb{R}) \subset \mathbb{R}_+ \Longrightarrow \delta_\mathrm{A}(\mathrm{P})(\mathbb{R}_+) \subset \mathbb{R}_+$.

(ii) $\forall \mathrm{P} \in \mathbb{R}_{2n}[\mathrm{X}],\ \mathrm{P}(\mathbb{R}) \subset \mathbb{R}_+ \Longrightarrow \delta_\mathrm{A}(\mathrm{P})(0) \in \mathbb{R}_+$.

(iii) La forme quadratique définie par la matrice $(a_{i+j})_{0 \leqslant i,j \leqslant n}$ est positive.

(École polytechnique)

▷ **Solution.**

• L'implication $(i) \Longrightarrow (ii)$ est triviale.

• Montrons que $(ii) \Longrightarrow (i)$. Soit $\mathrm{P} \in \mathbb{R}_{2n}[\mathrm{X}]$ un polynôme à valeurs positives et $x \in \mathbb{R}$. Le polynôme $\mathrm{Q}(\mathrm{X}) = \mathrm{P}(\mathrm{X} + x)$ est alors également positif. Donc par hypothèse $\delta_\mathrm{A}(\mathrm{Q})(0) \geqslant 0$. Or, $\delta_\mathrm{A}(\mathrm{Q})(0) = \delta_\mathrm{A}(\mathrm{P})(x)$. Cela vaut pour tout réel x et (i) est donc vérifiée.

• Faisons maintenant le lien avec la forme quadratique q définie sur $\mathbb{R}^{n+1}$ par la matrice $(a_{i+j})_{0 \leqslant i,j \leqslant n}$. Soit $x = (x_0, x_1 \ldots, x_n) \in \mathbb{R}^{n+1}$. Notons P le polynôme $\mathrm{P}(\mathrm{X}) = (x_0 + x_1 \mathrm{X} + \cdots + x_n \mathrm{X}^n)^2 \in \mathbb{R}_{2n}[\mathrm{X}]$. On a

$$q(x) = \sum_{0 \leqslant i,j \leqslant n} a_{i+j} x_i x_j = \sum_{k=0}^{2n} a_k \sum_{i+j=k} x_i x_j = \sum_{k=0}^{2n} a_k \frac{\mathrm{P}^{(k)}(0)}{k!} = \delta_\mathrm{A}(\mathrm{P})(0),$$

car $\sum_{i+j=k} x_i x_j$ est le coefficient de X^k dans P. L'implication $(ii) \Longrightarrow (iii)$ découle déjà de ces observations.

• Pour avoir $(iii) \Longrightarrow (ii)$ il suffit d'observer que $\delta_\mathrm{A}(\mathrm{P}^2)(0) \geqslant 0$ pour tout polynôme de degré inférieur à n et de conclure en se rappelant que

tout polynôme positif de $\mathbb{R}_{2n}[X]$ peut s'écrire comme somme de deux carrés : c'est un résultat que le lecteur trouvera dans l'exercice 5.16 du tome 1 d'algèbre (ou 5.20 dans la seconde édition). ◁

3.12. Une forme quadratique définie négative

Soit A une matrice de $\mathcal{M}_n(\mathbb{R})$ symétrique définie positive et q la forme quadratique définie par $q(x) = \det\begin{pmatrix} 0 & {}^t x \\ x & A \end{pmatrix}$ pour tout vecteur colonne $x \in \mathbb{R}^n$. Montrer que q est définie négative.

(École polytechnique)

Solution.

Calculons $q(x)$ en développant le déterminant selon sa première colonne. En notant $x_1, \ldots, x_n$ les coordonnées du vecteur colonne x, il vient

$$q(x) = \sum_{i=1}^{n} x_i (-1)^{i+2} \det \mathrm{M}_i(x)$$

où $\mathrm{M}_i(x)$ est la matrice dont la première ligne est ${}^t x$ et dont les lignes suivantes sont celles de A exceptée la ligne d'indice i. On développe alors $\det \mathrm{M}_i(x)$ selon sa première ligne pour obtenir

$$q(x) = \sum_{i=1}^{n} x_i (-1)^i \sum_{j=1}^{n} (-1)^{j+1} x_j \det \mathrm{A}_{ij} = -\sum_{1 \leqslant i,j \leqslant n} (-1)^{i+j} x_i x_j \det \mathrm{A}_{ij}$$

où A_{ij} est la matrice extraite de A en otant la i-ième ligne et la j-ième colonne. Autrement dit, q est l'opposée de la forme quadratique associée à la comatrice de A. Or, la formule fondamentale sur la comatrice donne $\mathrm{Com}\,\mathrm{A} = (\det \mathrm{A})\mathrm{A}^{-1}$ (car A est symétrique) et cela montre clairement que la comatrice de A est aussi définie positive. Donc q est bien définie négative. ◁

3.13. Matrices symétriques à diagonale positive

Soit $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbb{R})$ une matrice symétrique. Montrer l'équivalence des deux propositions suivantes :

(*i*) $\forall i \in [\![1, n]\!]$, $a_{ii} > 0$;

(*ii*) il existe P symétrique définie positive et D diagonale à éléments diagonaux strictement positifs tels que $\mathrm{A} = \mathrm{DP} + \mathrm{PD}$.

(École polytechnique)

▷ **Solution.**

• Si la condition (ii) est réalisée, avec $D = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$ et $P = (p_{ij})$. on a pour tout $(i,j) \in [\![1,n]\!]^2$, $a_{ij} = \lambda_i p_{ij} + p_{ij}\lambda_j = (\lambda_i + \lambda_j)p_{ij}$ et en particulier $a_{ii} = 2\lambda_i p_{ii}$. Les λ_i sont strictement positifs par hypothèse et si l'on appelle Φ la forme quadratique associée à P dans la base canonique $(e_1, \dots, e_n)$ de $\mathbb{R}^n$, $p_{ii} = \Phi(e_i) > 0$, car P est définie positive. Ceci implique $a_{ii} > 0$ pour tout i.

• Réciproquement supposons que A a des termes diagonaux strictement positifs. Si l'on considère $D = \mathrm{Diag}(\lambda_1, \dots, \lambda_n)$, où les λ_i sont strictement positifs et $P = (p_{ij}) \in \mathcal{M}_n(\mathbb{R})$, on a, d'après la démonstration directe, $A = DP + PD$ si, et seulement si, pour tout $(i,j) \in [\![1,n]\!]^2$, $p_{ij} = \dfrac{a_{ij}}{\lambda_i + \lambda_j}$. La matrice P obtenue est symétrique et on cherche les λ_i pour qu'elle soit définie positive. Pour $X \in \mathbb{R}^n$, on a

$$\begin{aligned} {}^t\mathrm{XPX} &= \sum_{1 \leqslant i,j \leqslant n} \frac{a_{ij}}{\lambda_i + \lambda_j} x_i x_j = \sum_{i=1}^{n} \frac{a_{ii}}{2\lambda_i} x_i^2 + \sum_{i \neq j} \frac{a_{ij}}{\lambda_i + \lambda_j} x_i x_j \\ &= \frac{1}{2} \sum_{i=1}^{n} a_{ii} y_i^2 + \sum_{i \neq j} a_{ij} \frac{\sqrt{\lambda_i}\sqrt{\lambda_j}}{\lambda_i + \lambda_j} y_i y_j, \end{aligned}$$

où l'on a posé $y_i = \dfrac{x_i}{\sqrt{\lambda_i}}$. On note désormais $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$.

On a alors $\sum_{i=1}^{n} a_{ii} y_i^2 \geqslant \alpha \|Y\|^2$, où $\alpha = \min_{1 \leqslant i \leqslant n} a_{ii} > 0$. Il faut choisir les λ_i pour que la deuxième somme soit petite. Pour t et u strictement positifs, l'expression $\dfrac{\sqrt{t}\sqrt{u}}{t+u}$ est maximale et vaut $\dfrac{1}{2}$ pour $t = u$; si l'on fixe t, elle décroît avec u pour $u > t$ et tend vers 0 en $+\infty$. On va donc choisir des λ_i distincts et grands. Soit $\lambda > 1$ et, pour tout i, $\lambda_i = \lambda^{2i}$. On a alors, pour $i > j$,

$$\frac{\sqrt{\lambda_i}\sqrt{\lambda_j}}{\lambda_i + \lambda_j} = \frac{\lambda^{i+j}}{\lambda^{2i} + \lambda^{2j}} = \frac{\lambda^{i-j}}{1 + \lambda^{2(i-j)}} \leqslant \frac{\lambda}{1 + \lambda^2} \leqslant \frac{1}{\lambda}.$$

En notant $K = \sup_{i \neq j} |a_{ij}|$, on obtient

$$\left| \sum_{i \neq j} a_{ij} \frac{\sqrt{\lambda_i}\sqrt{\lambda_j}}{\lambda_i + \lambda_j} y_i y_j \right| \leqslant \frac{K(n^2 - n)}{\lambda} \|Y\|_\infty^2 \leqslant \frac{K(n^2 - n)}{\lambda} \|Y\|^2$$

et donc

$${}^t\mathrm{XPX} \geqslant \left(\frac{\alpha}{2} - \frac{K(n^2 - n)}{\lambda} \right) \|Y\|^2.$$

Si λ est assez grand pour que $\frac{\alpha}{2} - \frac{K(n^2-n)}{\lambda}$ soit strictement positif, on aura pour $X \neq 0$, $Y \neq 0$ et ${}^tXPX > 0$. Pour un tel choix, P est symétrique définie positive. ◁

3.14. Restriction à un plan définie positive

Soit q la forme quadratique de $\mathbb{R}^3$ définie par

$$q(x,y,z) = x^2 + z^2 + 4xy + 2xz + 2yz.$$

Trouver les plans P de $\mathbb{R}^3$ tels que $q_{|P}$ est définie positive.

(École polytechnique)

▷ **Solution.**

Commençons par réduire la forme quadratique q en appliquant l'algorithme de Gauss. On a

$$\begin{aligned} q(x,y,z) &= (x+2y+z)^2 - 4y^2 - 2yz \\ &= (x+2y+z)^2 - 4(y+\frac{1}{4}z)^2 + \frac{z^2}{4} \\ &= X^2 - Y^2 + Z^2 \end{aligned}$$

en posant $X = x + 2y + z$. $Y = 2y + \frac{z}{2}$ et $Z = \frac{z}{2}$. Autrement dit, on vient de donner l'expression de q dans la base de $\mathbb{R}^3$ définie par les colonnes de la matrice P avec $P^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ 0 & 2 & 1/2 \\ 0 & 0 & 1/2 \end{pmatrix}$. Si on préfère voir les choses matriciellement on a donc ${}^tPAP = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ où $A = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}$ désigne la matrice de q dans la base canonique de $\mathbb{R}^3$. Nous allons bien entendu travailler dans la nouvelle base. Nous voyons que la forme q n'est pas positive : elle est de signature[2] $(2,1)$. Son cône isotrope est un vrai cône de $\mathbb{R}^3$ et cela permet déjà de voir géométriquement la solution à la question posée : les plans de $\mathbb{R}^3$ sur lesquels q est définie positive sont les plans qui ne coupent ce cône qu'en 0. La restriction de q à un plan coupant le cône selon une génératrice (un plan tangent au cône)

2. Voir l'exercice 3.22 pour cette notion qui a disparu des programmes des classes préparatoires.

est positive mais dégénérée et la restriction de q à un plan coupant le cône selon deux droites a une signature $(1,1)$. L'énoncé semble attendre une description analytique des plans solutions. Soit P un plan d'équation $aX+bY+cZ=0$. On peut supposer b non nul, car sinon P ne convient pas (q est strictement négative sur le vecteur $(0,1,0)$). Écrivons l'équation de P sous la forme $Y=\alpha X+\beta Z$. Pour $(X,Y,Z)\in P$ on a

$$X^2-Y^2+Z^2=X^2-(\alpha X+\beta Z)^2+Z^2=(1-\alpha^2)X^2+(1-\beta^2)Z^2-2\alpha\beta XZ.$$

Il en découle que le plan P est solution si, et seulement si, la forme quadratique $(X,Z)\longmapsto(1-\alpha^2)X^2+(1-\beta^2)Z^2-2\alpha\beta XZ$, de matrice $M=\begin{pmatrix}1-\alpha^2 & -\alpha\beta\\ -\alpha\beta & 1-\beta^2\end{pmatrix}$ est définie positive. C'est le cas si, et seulement si, $1-\alpha^2>0$ et $\det M>0$ (voir exercices 2.18 et 3.15). Cela équivaut facilement à $\alpha^2+\beta^2<1$. Pour finir on peut revenir à la base canonique : les plans de $\mathbb{R}^3$ sur lesquels q est définie positive sont les plans d'équation

$$4y+z=2\alpha(x+2y+z)+\beta z$$

avec $\alpha^2+\beta^2<1$. ◁

L'exercice suivant établit un résultat important qui a déjà été rencontré en 2.18 avec une autre solution qui est plus dans l'esprit actuel du programme des classes préparatoires.

3.15. Caractérisation de Sylvester des matrices définies positives

Soit $M=(m_{ij})_{1\leqslant i,j\leqslant n}$ une matrice symétrique réelle. Montrer que M est définie positive si, et seulement si, tous les mineurs principaux $D_k=\det(m_{ij})_{1\leqslant i,j\leqslant k}$ (pour $1\leqslant k\leqslant n$) sont strictement positifs.

(École polytechnique)

▷ **Solution.**

On note $(e_1,\dots,e_n)$ la base canonique de $\mathbb{R}^n$ et q la forme quadratique canoniquement associée à M : $q_M(X)={}^tXMX$ pour $X\in\mathbb{R}^n$.

• Le sens direct est facile. En effet, si q est définie positive il en est de même de sa restriction au sous-espace $\mathrm{Vect}(e_1,\dots,e_k)$ pour tout $k\in[\![1,n]\!]$. Or $(m_{ij})_{1\leqslant i,j\leqslant k}$ n'est autre que la matrice de cette restriction. Elle a donc un déterminant D_k strictement positif.

• Pour la réciproque, nous procédons par récurrence sur n. Le résultat est clair pour $n=1$. Supposons que $n\geqslant 2$ et que pour tout $k\in[\![1,n]\!]$ le mineur D_k soit strictement positif. L'hypothèse de récurrence ap-

pliquée à la matrice $(m_{ij})_{1\leqslant i,j\leqslant n-1}$ montre que la restriction de q à $\mathrm{F} = \mathrm{Vect}(e_1, \ldots, e_{n-1})$ est définie positive. Considérons l'orthogonal G de F. Comme q est non dégénérée (puisque $\det \mathrm{M} = \mathrm{D}_n > 0$), sa dimension est 1. Si $x \in \mathrm{F} \cap \mathrm{G}$, x est isotrope et donc nul puisque q est définie positive sur F. Ainsi, on a $\mathbb{R}^n = \mathrm{F} \overset{\perp}{\oplus} \mathrm{G}$. Prenons $(\varepsilon_1, \ldots, \varepsilon_{n-1})$ une base orthonormale de F pour q et ε_n un vecteur non nul de G. Si on note P la matrice de passage des e_i aux ε_i, on obtient pour la matrice de q_{M} dans la base des ε_i

$$ {}^t\mathrm{PMP} = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & 1 & 0 \\ 0 & \ldots & \ldots & 0 & \alpha \end{pmatrix}. $$

En prenant le déterminant, il vient $\alpha = (\det \mathrm{P})^2 \det \mathrm{M} = (\det \mathrm{P})^2 \mathrm{D}_n > 0$ et finalement q_{M} est définie positive. La matrice M est donc définie positive. ◁

Il est facile de déduire de cette caractérisation que l'ensemble des matrices réelles symétriques définies positives est un ouvert de $\mathcal{S}_n(\mathbb{R})$.

Si q est une forme quadratique positive de forme polaire b on dispose de l'importante inégalité de Cauchy-Schwarz : $b(x,y)^2 \leqslant q(x)q(y)$ pour tout couple $(x,y) \in \mathrm{E}^2$. Cette inégalité sera au cœur des énoncés suivants. Un premier corollaire immédiat de cette inégalité est que le cône isotrope d'une forme positive q coïncide avec son noyau (ou radical) :

$$ \mathrm{C}(q) = \{x \in \mathrm{E},\ q(x) = 0\} = \{x \in \mathrm{E},\ \forall y \in \mathrm{E},\ b(x,y) = 0\} = \mathrm{rad}(q). $$

Cette remarque est la clé du petit exercice suivant.

3.16. Comparaison de noyaux

Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ vérifiant ${}^t\mathrm{XAX} \geqslant 0$ pour tout $\mathrm{X} \in \mathbb{R}^n$.
Comparer $\mathrm{Ker}\,\mathrm{A}$ et $\mathrm{Ker}\,{}^t\mathrm{A}$.

(École polytechnique)

▷ **Solution.**

L'application $q : \mathrm{X} \longmapsto {}^t\mathrm{XAX}$ est une forme quadratique sur $\mathbb{R}^n$. Il faut simplement prendre garde au fait que la matrice A n'étant pas supposée symétrique elle ne représente pas la matrice de q dans la base canonique de $\mathbb{R}^n$: cette matrice est la partie symétrique de A à savoir $\mathrm{B} = \dfrac{1}{2}({}^t\mathrm{A} + \mathrm{A})$. Par hypothèse q est une forme quadratique positive.

Prenons X dans Ker A. On a alors $q(X) = 0$: X est un vecteur isotrope pour q. Comme q est positive, X est dans le radical de q et on a ${}^tYBX = 0$ pour tout vecteur $Y \in \mathbb{R}^n$. Cela donne ${}^tY\,{}^tAX = 0$ pour tout $Y \in \mathbb{R}^n$. Il en résulte que ${}^tAX = 0$. On a donc $\operatorname{Ker} A \subset \operatorname{Ker} {}^tA$. Les matrices A et ${}^tA$ ayant même rang, leur noyau est le même. ◁

3.17. Une forme quadratique positive

Soit q une forme quadratique positive sur $\mathbb{R}^n$ de matrice dans la base canonique $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$. Pour $x = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$, on pose

$$q'(x) = a_{nn} \sum_{1 \leqslant i,j \leqslant n-1} a_{ij} x_i x_j - \sum_{1 \leqslant i,j \leqslant n-1} a_{ni} a_{nj} x_i x_j.$$

Montrer que q' est une forme quadratique positive.

(École polytechnique)

▷ **Solution.**

Il suffit d'interpréter convenablement q'. Notons $(e_1, \ldots, e_n)$ la base canonique de $\mathbb{R}^n$ et b la forme polaire de q. Soit $x = (x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ et $X = (x_1, \ldots, x_{n-1}, 0) \in \mathbb{R}^n$.

On a $a_{nn} = q(e_n)$, $\displaystyle\sum_{1 \leqslant i,j \leqslant n-1} a_{ij} x_i x_j = q(X)$ et $\displaystyle\sum_{i=1}^{n-1} a_{ni} x_i = b(e_n, X)$.

Ainsi, la positivité de q' résulte simplement de l'inégalité de Cauchy-Schwarz pour la forme positive q puisque

$$q'(x) = q(e_n) q(X) - b(e_n, X)^2 \geqslant 0.$$

D'où le résultat. ◁

3.18. Condition pour qu'une matrice soit positive

Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$ une matrice symétrique réelle positive. On suppose que les coefficients de A sont tous non nuls. Montrer que la matrice $\left(\dfrac{1}{a_{ij}}\right)_{1 \leqslant i,j \leqslant n}$ est positive si, et seulement si, $\operatorname{rg} A = 1$.

(École polytechnique)

▷ **Solution.**

• Supposons d'abord que A soit de rang 1. Nous savons que A est diagonalisable en base orthonormée. Comme 0 est valeur propre avec un espace propre associé de dimension $n-1$, la matrice A admet une

unique valeur propre λ strictement positive (car A est positive) l'espace $\mathrm{Ker}(\mathrm{A}-\lambda \mathrm{I}_n)$ étant de dimension 1. Il existe donc $\mathrm{P} \in \mathrm{O}_n(\mathbb{R})$ telle que $\mathrm{A} = {}^t\mathrm{P}\,\mathrm{Diag}(\lambda, 0, \ldots, 0)\mathrm{P}$. En posant $\mathrm{M} = \mathrm{Diag}(\sqrt{\lambda}, 0, \ldots, 0)\mathrm{P}$ on a donc $\mathrm{A} = {}^t\mathrm{M}\mathrm{M}$. Notons $m_1, \ldots, m_n$ les coefficients de la première ligne de M, les autres coefficients de M étant nuls. On a alors $a_{ij} = m_i m_j$ pour tout couple (i, j). Autrement dit, si q désigne la forme quadratique associée à A on a pour tout X de coordonnées $x_1, \ldots, x_n$,

$$q(\mathrm{X}) = {}^t\mathrm{X}\mathrm{A}\mathrm{X} = \sum_{1 \leqslant i,j \leqslant n} m_i m_j x_i x_j = \left(\sum_{k=1}^{n} m_k x_k \right)^2 .$$

Il est alors clair que la matrice $\mathrm{A}' = \left(\frac{1}{a_{ij}} \right)_{1 \leqslant i,j \leqslant n}$ est positive. En effet, on a pour tout X,

$${}^t\mathrm{X}\mathrm{A}'\mathrm{X} = \sum_{1 \leqslant i,j \leqslant n} \frac{x_i x_j}{m_i m_j} = \left(\sum_{k=1}^{n} \frac{x_k}{m_k} \right)^2 \geqslant 0.$$

• Traitons maintenant la réciproque. Comme A est positive, l'inégalité de Cauchy-Schwarz montre que $a_{ij}^2 \leqslant a_{ii} a_{jj}$ pour tout couple (i, j). Mais par hypothèse la matrice $\left(\frac{1}{a_{ij}} \right)_{1 \leqslant i,j \leqslant n}$ est aussi positive de sorte qu'on a aussi $\frac{1}{a_{ij}^2} \leqslant \frac{1}{a_{ii}} \frac{1}{a_{jj}}$. Finalement, on a $a_{ij}^2 = a_{ii} a_{jj}$ pour tout couple (i, j). Sommons ces égalités sur i et j. Il vient $\mathrm{Tr}({}^t\mathrm{A}\mathrm{A}) = (\mathrm{Tr}\,\mathrm{A})^2$ c'est-à-dire $\mathrm{Tr}\,\mathrm{A}^2 = (\mathrm{Tr}\,\mathrm{A})^2$ ou encore $\sum_{i=1}^{n} \lambda_i^2 = \left(\sum_{i=1}^{n} \lambda_i \right)^2$ où $\lambda_1, \ldots, \lambda_n$ sont les valeurs propres de A prises avec multiplicité. En développant on a donc $\sigma_2(\lambda_1, \ldots, \lambda_n) = 0$ et comme les λ_i sont tous positifs, il y en a au plus un qui est non nul. Comme A est non nulle, elle admet une valeur propre simple strictement positive et 0 pour valeur propre de multiplicité $n - 1$. Elle est donc bien de rang 1. ◁

3.19. Inégalité de Bergström

Soit $\mathrm{A} = (a_{ij})_{1 \leqslant i,j \leqslant n}$ et $\mathrm{B} = (b_{ij})_{1 \leqslant i,j \leqslant n}$ deux matrices symétriques définies positives. On considère $\mathrm{A}_1 = (a_{ij})_{2 \leqslant i,j \leqslant n}$ et $\mathrm{B}_1 = (b_{ij})_{2 \leqslant i,j \leqslant n}$. Montrer que

$$\frac{\det \mathrm{A}}{\det \mathrm{A}_1} + \frac{\det \mathrm{B}}{\det \mathrm{B}_1} \leqslant \frac{\det(\mathrm{A}+\mathrm{B})}{\det(\mathrm{A}_1 + \mathrm{B}_1)}.$$

(École normale supérieure)

▷ **Solution.**

Le terme $\det A_1$ est le coefficient $(1,1)$ de la comatrice de A. La formule classique $A^{-1} = \dfrac{1}{\det A}\,{}^t\mathrm{Com}\,A$ montre que $\dfrac{\det A_1}{\det A}$ est le coefficient d'indice $(1,1)$ de A^{-1}. Notons alors u et v les endomorphismes de $\mathbb{R}^n$ (auto-adjoints définis positifs) associés à A et B dans la base canonique $(e_1,\dots,e_n)$. On vient donc de voir que

$$\frac{\det A_1}{\det A} = \langle u^{-1}(e_1), e_1\rangle.$$

On peut évidemment tenir le même raisonnement pour B et A + B. L'inégalité à démontrer s'écrit donc

$$\frac{1}{\langle u^{-1}(e_1), e_1\rangle} + \frac{1}{\langle v^{-1}(e_1), e_1\rangle} \leqslant \frac{1}{\langle (u+v)^{-1}(e_1), e_1\rangle}.$$

Montrons plus généralement, que pour tout $x \in E \setminus \{0\}$, on a

$$\frac{1}{\langle u^{-1}(x), x\rangle} + \frac{1}{\langle v^{-1}(x), x\rangle} \leqslant \frac{1}{\langle (u+v)^{-1}(x), x\rangle}.$$

Pour cela, on démontre que, pour tout $(x,y) \in E^2$, on a

$$\langle x, y\rangle^2 \leqslant \langle u^{-1}(x), x\rangle \langle u(y), y\rangle.$$

L'application $\varphi : (x,y) \longmapsto \langle u(x), y\rangle$ est un produit scalaire, puisque u est symétrique défini positif. Il suffit de lui appliquer l'inégalité de Cauchy-Schwarz :

$$\begin{aligned}\langle x, y\rangle^2 &= \langle u(u^{-1}(x)), y)\rangle^2 = \varphi(u^{-1}(x), y)^2 \\ &\leqslant \varphi(u^{-1}(x), u^{-1}(x))\varphi(y,y) \leqslant \langle x, u^{-1}(x)\rangle\langle u(y), y\rangle.\end{aligned}$$

On en déduit que si y n'est pas orthogonal à x, on a

$$\frac{1}{\langle u^{-1}(x), x\rangle} \leqslant \frac{\langle u(y), y\rangle}{\langle x, y\rangle^2}.$$

On peut écrire une inégalité du même type pour v, et l'on obtient, si $\langle x, y\rangle \neq 0$,

$$\frac{1}{\langle u^{-1}(x), x\rangle} + \frac{1}{\langle v^{-1}(x), x\rangle} \leqslant \frac{\langle u(y), y\rangle}{\langle x, y\rangle^2} + \frac{\langle v(y), y\rangle}{\langle x, y\rangle^2} \leqslant \frac{\langle (u+v)(y), y\rangle}{\langle x, y\rangle^2}.$$

En prenant $y = (u+v)^{-1}(x)$, on a l'inégalité voulue. ◁

Signalons que l'exercice est souvent proposé à l'oral avec une question préliminaire, qui le rend plus facile, consistant à prouver l'inégalité $\langle x, y\rangle^2 \leqslant \langle u^{-1}(x), x\rangle\langle u(y), y\rangle$.

La notion de positivité permet de munir l'espace des matrices symétriques réelles d'une relation d'ordre. L'exercice classique suivant généralise un résultat bien connu sur les suites réelles (cas $n = 1$*).*

3.20. Convergence d'une suite croissante et majorée

Soit $(A_k)_{k\geqslant 0}$ une suite de matrices symétriques de taille n. Soit A une matrice symétrique positive de taille n. On suppose que les matrices $A - A_k$ et $A_{k+1} - A_k$ sont toutes positives. Montrer que la suite $(A_k)_{k\geqslant 0}$ est convergente.

(École polytechnique)

▷ **Solution.** On montre facilement que la relation $\geqslant$ définie sur l'ensemble des matrices symétriques par

$$A \geqslant B \Longleftrightarrow A - B \text{ est positive}$$

est une relation d'ordre appelé ordre de Loewner. Il s'agit de démontrer qu'une suite de matrices symétriques $(A_k)_{k\geqslant 0}$ croissante et majorée au sens de cet ordre est convergente. On note, pour tout $k \in \mathbb{N}$, $A_k = (a_{ij}^k)$. Soit $X \in \mathbb{R}^n$. L'hypothèse implique que

$$\begin{aligned} {}^tXAX - {}^tXA_kX &= {}^tX(A - A_k)X \geqslant 0 \\ {}^tXA_{k+1}X - {}^tXA_kX &= {}^tX(A_{k+1} - A_k)X \geqslant 0, \end{aligned}$$

pour tout entier naturel k. La suite réelle $({}^tXA_kX)_{k\geqslant 0}$ est croissante et majorée donc est convergente. On note $\Phi(X)$ sa limite et on pose pour $(X, Y) \in (\mathbb{R}^n)^2$, $\varphi(X, Y) = \frac{1}{2}(\Phi(X + Y) - \Phi(X) - \Phi(Y))$, ce qui donne

$$\begin{aligned} \varphi(X, Y) &= \lim_{k\to+\infty} \frac{1}{2}({}^t(X + Y)A_k(X + Y) - {}^tXA_kX - {}^tYA_kY) \\ &= \lim_{k\to+\infty} {}^tXA_kY. \end{aligned}$$

L'application φ est une forme bilinéaire symétrique sur $\mathbb{R}^n$, car limite simple d'une suite de formes bilinéaires symétriques. Soit $B = (b_{ij})$ la matrice de φ dans la base canonique $(e_1, \ldots, e_n)$ de $\mathbb{R}^n$. On a, pour tout $(i, j) \in [\![1, n]\!]^2$,

$$b_{ij} = \varphi(e_i, e_j) = \lim_{k\to+\infty} {}^te_iA_ke_j = \lim_{k\to+\infty} a_{ij}^k.$$

La matrice B est la limite de la suite $(A_k)_{k\geqslant 0}$. ◁

3.21. Décroissance de la fonction inverse

1. Soit $A \in \mathcal{M}_n(\mathbb{R})$ symétrique définie positive, $Y \in \mathbb{R}^n$. Trouver le minimum de $J : \mathbb{R}^n \longrightarrow \mathbb{R}$ définie par

$$J(X) = {}^tXAX - 2{}^tYX.$$

2. Soient $A, B \in \mathcal{M}_n(\mathbb{R})$ deux matrices symétriques définies positives. Montrer que si $A - B$ est définie positive alors $B^{-1} - A^{-1}$ l'est aussi.

(École normale supérieure)

▷ **Solution.**

1. La fonction J est clairement de classe $\mathcal{C}^1$ sur $\mathbb{R}^n$ donc, si elle admet un minimum en un point X_0, sa différentielle en X_0 est nulle. Or, celle-ci est définie par

$$dJ_{X_0}(H) = 2\,{}^tX_0AH - 2\,{}^tYH$$

pour tout $H \in \mathbb{R}^n$ (autrement dit, si $\mathbb{R}^n$ est muni de sa structure euclidienne canonique, le gradient de J en X_0 est le vecteur $2(AX_0 - Y)$). Le seul point où J peut admettre son minimum est donc $X_0 = A^{-1}Y$. Effectivement, on a pour tout $H \in \mathbb{R}^n$,

$$J(X_0 + H) - J(X_0) = {}^tHAH + 2({}^tX_0AH - {}^tYH) = {}^tHAH \geqslant 0,$$

car la matrice A est définie positive.

Conclusion. Le minimum de J est atteint en $X_0 = A^{-1}Y$ et seulement en ce point. Il vaut $-{}^tYA^{-1}Y$.

2. Supposons que $A - B$ soit définie positive. Pour $X \in \mathbb{R}^n$ on a donc ${}^tXBX \leqslant {}^tXAX$ l'inégalité étant stricte si X est non nul. Soit $Y \in \mathbb{R}^n$ un vecteur non nul fixé. On a pour tout X non nul,

$$J_B(X) = {}^tXBX - 2\,{}^tYX < {}^tXAX - 2\,{}^tYX = J_A(X).$$

En particulier on peut prendre le point $X = A^{-1}Y$ où le membre de droite atteint son minimum. On en déduit que

$$-{}^tYB^{-1}Y = \min J_B \leqslant J_B(A^{-1}Y) < J_A(A^{-1}Y) = -{}^tYA^{-1}Y.$$

Bref, comme ${}^tY(B^{-1} - A^{-1})Y > 0$ pour tout Y non nul la matrice $B^{-1} - A^{-1}$ est bien définie positive. ◁

Nous abordons maintenant le problème de la classification des formes quadratiques réelles et la notion de signature. Cette section n'est plus actuellement dans les programmes des classes préparatoires mais nous avons préféré inclure ces exercices pour nos nombreux lecteurs agrégatifs. Les taupins pourront passer directement au thème suivant qui concerne la réduction simultanée de deux formes quadratiques. Voici toutefois un rapide rappel du cours. Deux formes quadratiques q et q' sur un même $\mathbb{K}$-espace vectoriel E de dimension n sont dites équivalentes s'il existe un automorphisme u de E tel que $q' = q \circ u$. Si A et A′ désignent les matrices

de q et q' dans une base de E *cela revient à l'existence d'une matrice inversible* P *(à savoir la matrice de u dans la base concernée) telle que* $A' = {}^tPAP$. *On dit alors que les matrices* A *et* A′ *sont congruentes. On définit ainsi des relations d'équivalence sur respectivement* Q(E) *(l'espace des formes quadratiques de* E*) et $\mathcal{S}_n(\mathbb{K})$ l'espace des matrices symétriques de taille n. Le problème de la classification des formes quadratiques consiste à décrire les classes d'équivalence. Il est très difficile dans le cas général et fait intervenir les propriétés arithmétiques du corps $\mathbb{K}$. Traitons ici les cas $\mathbb{K} = \mathbb{C}$ et $\mathbb{K} = \mathbb{R}$. L'existence d'une base q-orthogonale (voir page 198) montre que toute forme quadratique q de $\mathbb{K}^n$ est équivalente à une forme du type $(x_1, \dots, x_n) \longmapsto a_1x_1^2 + \cdots + a_tx_t^2$ où les a_i sont des scalaires non nuls. L'entier t est simplement le rang de q. Si $\mathbb{K} = \mathbb{C}$ les a_i sont tous des carrés et il est facile de voir que q est équivalente à $(x_1, \dots, x_n) \longmapsto x_1^2 + \cdots + x_t^2$. On en déduit que deux formes quadratiques de* E *sont équivalentes si et seulement si elles ont le même rang. Si $\mathbb{K} = \mathbb{R}$ on peut, en séparant les a_i positifs et les a_i négatifs, montrer que q est équivalente à une forme du type $(x_1, \dots, x_n) \longmapsto x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots x_{r+s}^2$. L'entier $r + s$ est le rang de q. Il est remarquable, et cela constitue la loi d'inertie de Sylvester, que quelle que soit la base q-orthogonale utilisée, le nombre de coefficients positifs et le nombre de coefficients négatifs sont toujours les mêmes. Autrement dit le couple (r, s) est intrinsèquement associé à q : on l'appelle la signature de q. On a ainsi le théorème suivant : deux formes quadratiques sur* E *sont équivalentes si et seulement si elles ont la même signature. L'exercice suivant redémontre ce résultat et donne une vision géométrique de la signature. L'exercice 3.27 s'occupera de la classification des formes quadratiques lorsque $\mathbb{K}$ est un corps fini.*

3.22. Signature d'une forme quadratique réelle

Soit E un $\mathbb{R}$-espace vectoriel de dimension finie n, q une forme quadratique sur E. Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base q-orthogonale de E. On pose $q(x) = \sum_{i=1}^{n} a_i x_i^2$ où $x = \sum_{i=1}^{n} x_i e_i$ et on note r (resp. s) le nombre de réels a_i strictement positifs (resp. strictement négatifs).

1. Montrer que r (resp. s) est la dimension maximale d'un sous-espace vectoriel de E sur lequel q est définie positive (resp. définie négative). En déduire que le couple (r, s) ne dépend pas de la base orthogonale choisie. *Par définition ce couple est la signature de q.*

2. Montrer que deux formes quadratiques de E sont équivalentes si, et seulement si, elles ont la même signature.

(École polytechnique)

▷ **Solution.**

1. Quitte à permuter les vecteurs de la base B, on peut supposer que $a_1, \ldots, a_r$ sont strictement positifs, que $a_{r+1}, \ldots, a_{r+s}$ sont strictement négatifs et que les autres a_i sont nuls. Il est clair que la restriction de q au sous-espace $\mathrm{Vect}(e_1, \ldots, e_r)$ est définie positive. La dimension maximale d'un tel sous-espace est donc au moins égale à r. Inversement, soit F un sous-espace de E sur lequel q est définie positive. Comme q est négative sur le sous-espace $\mathrm{G} = \mathrm{Vect}(e_{r+1}, \ldots, e_n)$ on est certain que F et G sont en somme directe. Par suite $\dim \mathrm{F} \leqslant n - \dim \mathrm{G} = r$. D'où le résultat. On raisonne bien entendu de même pour prouver que s est la dimension maximale d'un sous-espace sur lequel q est définie négative. Ces dimensions sont intrinsèquement associées à q et ne dépendent pas de la base orthogonale B choisie.

2. Si q est de signature (r, s) elle s'écrit $x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots x_{r+s}^2$ dans une bonne base de E. Il en résulte que deux formes quadratiques de même signature sont équivalentes. La réciproque est tout aussi rapide puisque si $q' = q \circ u$ avec $u \in \mathrm{GL(E)}$ alors les sous-espaces sur lesquels q' est définie positive (resp. négative) sont les images par u^{-1} des sous-espaces sur lesquels q' est définie positive (resp. négative). Comme u^{-1} préserve les dimensions, la question précédente montre que q et q' ont la même signature. ◁

Voici une première application de cette notion.

3.23. Composantes connexes

Soit $(\mathrm{E}, \|\ \|)$ un $\mathbb{R}$-espace vectoriel normé de dimension $n \geqslant 1$. On note Q(E) l'ensemble des formes quadratiques sur E et Ω(E) le sous-ensemble des formes quadratiques non dégénérées.

1. Montrer que Ω(E) est un ouvert de Q(E).

2. Soit $q \in \Omega(\mathrm{E})$. Établir l'existence de deux sous-espaces vectoriels supplémentaires F et G de E et d'un réel $k > 0$ tels que

$$\forall x \in \mathrm{F},\quad q(x) \geqslant k\|x\|^2 \quad \text{et} \quad \forall x \in \mathrm{G},\quad q(x) \leqslant -k\|x\|^2.$$

3. Montrer qu'on peut trouver une norme N sur Q(E) telle que si $q' \in \mathrm{Q(E)}$ vérifie $\mathrm{N}(q - q') < k$ alors q' a la même signature que q.

4. En déduire les composantes connexes de Ω(E).

(École polytechnique)

▷ **Solution.**

1. En choisissant une base de E, Q(E) s'identifie à $\mathcal{S}_n(\mathbb{R})$ l'espace des matrices symétriques réelles et Ω(E) s'identifie au sous-ensemble formé

des matrices symétriques inversibles. Comme la fonction déterminant est continue, $\mathrm{GL}_n(\mathbb{R}) = \det^{-1}(\mathbb{R}^*)$ est ouvert et $\mathcal{S}_n(\mathbb{R}) \cap \mathrm{GL}_n(\mathbb{R})$ est donc un ouvert de $\mathcal{S}_n(\mathbb{R})$.

2. Notons (r, s) la signature de q. Comme q est non dégénérée on a $r + s = n$. Comme on le voit en se plaçant dans une base q-orthogonale de E, il existe deux sous-espaces supplémentaires F et G de dimensions respectives r et s tels que q soit définie positive sur F et définie négative sur G. Montrons que ces sous-espaces conviennent.

La restriction de q à F étant définie positive, la fonction $\sqrt{q}$ est une norme (euclidienne) sur F. Comme F est de dimension finie cette norme est équivalente à la restriction de la norme de E. En particulier il existe $k_1 > 0$ tel que pour tout $x \in \mathrm{F}$, $\sqrt{q(x)} \geqslant k_1\|x\|$ et donc $q(x) \geqslant k_1^2\|x\|^2$. De même $-q$ définit une structure euclidienne sur G et il existe $k_2 > 0$ tel que pour tout $x \in \mathrm{G}$, $\sqrt{-q(x)} \geqslant k_2\|x\|$ soit encore $q(x) \leqslant -k_2^2\|x\|^2$. Il suffit de prendre $k = \min(k_1^2, k_2^2)$ pour conclure.

3. On va maintenant montrer que si q' est une forme quadratique assez proche de q elle va rester définie positive sur F et définie négative sur G et par suite elle va avoir la même signature que q. Pour $q' \in \mathrm{Q}(\mathrm{E})$ posons $\mathrm{N}(q') = \sup_{\|x\|=1} |q'(x)|$. Il s'agit clairement d'une norme sur $\mathrm{Q}(\mathrm{E})$. Par homogénéité on a $|q'(x)| \leqslant \mathrm{N}(q')\|x\|^2$ pour tout vecteur x de E. Soit alors $q' \in \mathrm{Q}(\mathrm{E})$ telle que $\mathrm{N}(q - q') < k$. On a donc $|q(x) - q'(x)| < k\|x\|^2$ pour tout vecteur non nul x. Ainsi, si $x \in \mathrm{F} \setminus \{0\}$, on a par inégalité triangulaire, $q'(x) > q(x) - k\|x\|^2 \geqslant 0$ et de même, si $x \in \mathrm{G} \setminus \{0\}$ on a $q'(x) < q(x) + k\|x\|^2 \leqslant 0$. Donc q' est définie positive sur F et définie négative sur G. Sa signature est la même que celle de q.

4. Pour $k \in [\![0, n]\!]$, notons $\Omega_k(\mathrm{E})$ l'ensemble des formes quadratiques (non dégénérées) de signature $(k, n-k)$. On vient de montrer que les $\Omega_k(\mathrm{E})$ sont tous ouverts. Ils forment par ailleurs une partition de $\Omega(\mathrm{E})$. On va montrer que les $\Omega_k(\mathrm{E})$ sont tous connexes : ce seront donc les composantes connexes de $\Omega(\mathrm{E})$. Travaillons matriciellement. Soit A, B deux matrices symétriques de signature $(k, n-k)$. Il existe donc P et Q inversibles telles que $\mathrm{A} = {}^t\mathrm{P}\mathrm{D}_k\mathrm{P}$ et $\mathrm{B} = {}^t\mathrm{Q}\mathrm{D}_k\mathrm{Q}$ où D_k est la matrice diagonale contenant k coefficients 1 et $n-k$ coefficients -1. On peut très bien supposer P et Q de déterminant strictement positif (quitte à changer en son opposée la première ligne de P ou de Q). Or l'ensemble des matrices réelles inversibles de déterminant > 0 est connexe par arcs. On peut donc y trouver un chemin continu $t \in [0, 1] \longmapsto \mathrm{P}(t)$ tel que $\mathrm{P}(0) = \mathrm{P}$ et $\mathrm{P}(1) = \mathrm{Q}$. Alors $t \longmapsto {}^t\mathrm{P}(t)\mathrm{D}_k\mathrm{P}(t)$ est un chemin continu formé de matrices symétriques de signature $(k, n-k)$ et joignant A à B. ◁

Les exercices suivants sont consacrés à des calculs de signature. Plutôt que de multiplier les exemples nous avons choisi d'illustrer les différentes techniques.

3.24. Un calcul de signature (1)

Montrer que $\Phi : (M, N) \longmapsto \mathrm{Tr}(MN)$ est une forme bilinéaire symétrique sur $\mathcal{M}_n(\mathbb{R})$ et déterminer la signature de la forme quadratique associée.

(École polytechnique)

▷ **Solution.**

La bilinéarité résulte directement de la bilinéarité du produit et de la linéarité de la trace. La forme quadratique associée est définie par $q(M) = \mathrm{Tr}(M^2)$. Si $M = (m_{ij})$ est symétrique, on a

$$q(M) = \mathrm{Tr}(M^2) = \mathrm{Tr}({}^tMM) = \sum_{1 \leqslant i,j \leqslant n} m_{ij}^2.$$

Donc q est définie positive sur le sous-espace vectoriel des matrices symétriques. Si M est antisymétrique, le calcul précédent montre que $q(M) = -\mathrm{Tr}({}^tMM) = -\sum_{1 \leqslant i,j \leqslant n} m_{ij}^2$, de sorte que q est définie négative sur le sous-espace des matrices antisymétriques. On peut aussi observer que ces deux sous-espaces de $\mathcal{M}_n(\mathbb{R})$ sont supplémentaires orthogonaux bien que cela ne soit pas indispensable pour conclure (cf. exercice 3.22). En effet, si M est symétrique et N antisymétrique,

$$\mathrm{Tr}(MN) = \mathrm{Tr}({}^t(MN)) = \mathrm{Tr}({}^tN{}^tM) = -\mathrm{Tr}(NM) = -\mathrm{Tr}(MN).$$

Bref, q est de rang n^2 et de signature $\left(\frac{n(n+1)}{2}, \frac{n(n-1)}{2}\right)$. ◁

3.25. Un calcul de signature (2)

Soit q une forme quadratique de signature $(n-1, 1)$ sur un espace vectoriel réel E de dimension n et F un sous-espace de E dans lequel il existe x tel que $q(x) < 0$. Quelle est la signature de q restreinte à F ?

(École polytechnique)

▷ **Solution.**

Notons q' la restriction de q à F. La forme quadratique q' est définie

négative sur la droite D engendrée par x. Par ailleurs il ne peut pas exister de sous-espace de F de dimension > 1 sur lequel q' est définie négative puisque la signature de q est $(n-1,1)$. Pour avoir la signature de q' il suffit alors de connaître son rang. Montrons en fait que q' est non dégénérée, ce qui permettra de dire que sa signature est $(\dim \mathrm{F}-1,1)$. Soit $y \in \operatorname{Ker} q'$. En particulier y est orthogonal à x donc appartient à l'hyperplan $\mathrm{H} = \mathrm{D}^{\perp}$. Or E est somme directe orthogonale de D et de H, car x n'est pas isotrope. Il est clair que la restriction de q à H est définie positive. Comme $y \in \mathrm{H}$ et comme $q(y) = q'(y) = 0$ on a forcément $y = 0$.

Conclusion. La signature de q' est $(\dim \mathrm{F}-1,1)$. ◁

Si A *est la matrice d'une forme* q *dans une base, la signature de* q *est* (r,s)*, où* r *(resp.* s*) est le nombre de valeurs propres strictement positives (resp. strictement négatives) de* A*. En effet, il existe* D *diagonale et* P *orthogonale telle que* $\mathrm{A} = \mathrm{P}\mathrm{D}\,{}^t\mathrm{P}$*. La matrice* A *est congruente à* D *donc a même signature que* D*.*

3.26. Un calcul de signature (3)

Déterminer la signature de la forme quadratique q de $\mathbb{R}^n$ définie par $q(x_1,\dots,x_n) = \sum_{i \neq j} x_i x_j$.

(**École polytechnique**)

Solution.

La matrice de q dans la base canonique est $\mathrm{A} = \mathrm{J} - \mathrm{I}_n$ où J est la matrice dont tous les coefficients valent 1. Il est clair que J est de rang 1 ce qui montre que -1 est valeur propre de A avec un espace propre associé de dimension $n-1$. Comme la trace de A est nulle, la dernière valeur propre de A est $n-1$. On en déduit que la signature de q est $(1,n-1)$. ◁

3.27. Formes quadratiques sur un corps fini

Soit $\mathbb{K}$ un corps fini de cardinal q et de caractéristique $p > 2$.

1. Montrer qu'il y a $\dfrac{q-1}{2}$ carrés dans $\mathbb{K}^*$.

2. Soit a et b dans $\mathbb{K}^*$. Montrer que l'équation $ax^2 + by^2 = 1$ admet toujours au moins une solution.

3. En déduire qu'il y a exactement deux classes d'équivalence de formes quadratiques non dégénérées sur $\mathbb{K}^n$.

(**École normale supérieure**)

▷ **Solution.**

1. L'application $\psi : x \longmapsto x^2$ est un morphisme de groupes de $(\mathbb{K}^*, \times)$ dans lui-même dont l'image est exactement l'ensemble C des carrés non nuls de $\mathbb{K}$. La caractéristique de $\mathbb{K}$ n'étant pas égale à 2 le noyau de ψ est égal à $\{\pm 1\}$ et donc de cardinal 2. Il en découle que $|\mathrm{C}| = \dfrac{q-1}{2}$. Notons, car cela servira dans la suite, que si δ n'est pas un carré de $\mathbb{K}^*$, alors les autres éléments non carrés de $\mathbb{K}^*$ sont les éléments de la forme δc lorsque c parcourt C.

2. On utilise un argument combinatoire. En comptant 0, on vient de voir qu'il y a $\dfrac{q+1}{2}$ carrés dans $\mathbb{K}$. On en déduit que les ensembles $\{1 - ax^2,\ x \in \mathbb{K}\}$ et $\{br^2,\ x \in \mathbb{K}\}$ sont tous deux de cardinal $\dfrac{q+1}{2}$ (la multiplication par un élément non nul et les translations sont des opérations bijectives de $\mathbb{K}$) et ne peuvent donc pas être disjoints. D'où l'existence d'au moins un couple (x, y) tel que $1 - ax^2 = by^2$.

3. On va procéder par récurrence sur n.

Pour $n = 1$ une forme quadratique non dégénérée s'écrit $x \longmapsto ax^2$ avec $a \neq 0$. Si a est un carré elle est équivalente à la forme $q_1 : x \longmapsto x^2$ et si a n'est pas un carré la remarque faite en 1 montre qu'elle est équivalente à $q_2 : x \longmapsto \delta x^2$ où δ est un élément non carré quelconque fixé. Il y a donc au plus 2 classes d'équivalence de formes quadratiques non dégénérées. Un changement de base revient à poser $x = ux'$ avec u non nul et $q_1 : x \longmapsto x^2$ se transforme en $x' \longmapsto u^2x'^2$. Cela montre que q_1 et q_2 ne sont pas équivalentes.

Traitons aussi le cas $n = 2$ qui est fondamental dans le passage du rang n au rang $n+1$. Soit q une forme quadratique non dégénérée sur $\mathbb{K}^2$. Quitte à se placer dans une base q-orthogonale on peut supposer que $q(x, y) = ax^2 + by^2$ avec a, b non nuls. D'après la question 2, on peut trouver un vecteur $f_1 = (x_1, y_1)$ tel que $q(x_1, y_1) = 1$. Soit f_2 un vecteur dirigeant la droite $f_1^{\perp}$. Alors (f_1, f_2) est une base de $\mathbb{K}^2$ et on a $q(xf_1 + yf_2) = x^2 + \alpha y^2$ avec $\alpha \in \mathbb{K}^*$. Si α est un carré, alors q est équivalente à la forme $q_1 : (x, y) \longmapsto x^2 + y^2$ et sinon q est équivalente à la forme $q_2 : (x, y) \longmapsto x^2 + \delta y^2$ où δ désigne toujours un élément non carré de $\mathbb{K}$ fixé. Les formes q_1 et q_2 ne sont pas équivalentes puisque sinon il existerait une matrice $\mathrm{P} \in \mathrm{GL}_2(\mathbb{K})$ telle que $\begin{pmatrix} 1 & 0 \\ 0 & \delta \end{pmatrix} = {}^t\mathrm{P}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\mathrm{P}$, ce qui est impossible, car en passant au déterminant on aurait $\delta = (\det \mathrm{P})^2$.

Montrons maintenant par récurrence sur n que toute forme quadratique non dégénérée sur $\mathbb{K}^n$ est équivalente soit à $q_1 : (x_1, \ldots, x_n) \longmapsto x_1^2 + \cdots + x_n^2$ soit à $q_2 : (x_1, \ldots, x_n) \longmapsto x_1^2 + \cdots + x_{n-1}^2 + \delta x_n^2$ où δ est comme avant un élément non carré de $\mathbb{K}$ fixé. Bien entendu q_1 et q_2 ne sont pas équivalentes (on le voit comme en dimension 2 à l'aide du déterminant). Ce qui précède montre que le résultat est vrai aux

rangs $n = 1$ et $n = 2$. Supposons le résultat vrai au rang $n-1$ et soit q une forme quadratique non dégénérée sur $\mathbb{K}^n$. En prenant une base q-orthogonale on peut supposer que $q(x_1, \ldots, x_n) = a_1x_1^2 + \cdots + a_nx_n^2$ où les a_i sont tous non nuls. On applique alors le cas $n = 2$ à la restriction de q au plan correspondant aux deux premières coordonnées. On en déduit que q est équivalente à une forme du type $q' : (x_1, \ldots, x_n) \longmapsto x_1^2 + a'_2x_2^2 + \cdots + a_nx_n^2$ avec $a'_2 \neq 0$. Il suffit enfin d'appliquer l'hypothèse de récurrence à la restriction de q' à l'hyperplan correspondant aux $n-1$ dernières coordonnées pour conclure. ◁

Le théorème spectral affirme qu'un endomorphisme auto-adjoint d'un espace euclidien E *peut se diagonaliser dans une base orthonormée de* E. *Le lecteur trouvera déjà beaucoup d'exercices sur ce premier aspect purement linéaire dans le chapitre 2. Mais ce théorème admet une autre lecture souvent moins bien comprise des étudiants que nous allons exposer maintenant.*

Matriciellement, le théorème spectral dit que si $A \in \mathcal{S}_n(\mathbb{R})$ *est une matrice symétrique réelle il existe* $P \in O_n(\mathbb{R})$ *telle que* $D = P^{-1}AP$ *soit diagonale. Mais comme* $P^{-1} = {}^tP$, *cette égalité admet une nouvelle lecture bilinéaire : elle signifie aussi que* A *est congruente à la matrice* D, *où si l'on préfère que dans la nouvelle base de* $\mathbb{R}^n$ *définie par les colonnes de* P, *la matrice de la forme quadratique* $q_A : X \longmapsto {}^tXAX$ *canoniquement associée à* A *est diagonale. Cette nouvelle base est donc orthogonale pour* q_A *tout en étant orthonormée pour le produit scalaire canonique de* $\mathbb{R}^n$. *Une version abstraite de ce résultat est donc donnée par le théorème suivant :* si q, q' sont deux formes quadratiques sur un espace vectoriel réel E de dimension finie, et si q est définie positive, il existe une base de E qui est orthonormée pour q et orthogonale pour q'. *On notera qu'en général il n'est pas possible de trouver une base orthogonale simultanément pour deux formes quadratiques quelconques : prendre par exemple* $q(x, y) = x^2 - y^2$ *et* $q'(x, y) = 2xy$ *sur* $\mathbb{R}^2$ *(voir aussi l'exercice 3.34).*

Dans nos exercices, c'est toutefois essentiellement dans la version matricielle suivante que ce résultat sera utilisé : si A *et* B *sont deux matrices symétriques réelles de taille* n, *et si* A *est définie positive, il existe* $P \in GL_n(\mathbb{R})$ *et* D *matrice diagonale réelle telles que*

$$^tPAP = I_n \quad et \quad {}^tPBP = D.$$

Nous parlerons alors de pseudo-réduction simultanée car il ne s'agit pas ici d'une réduction au sens linéaire du terme. Attention au fait que la matrice P *est simplement inversible (et pas orthogonale !) : on applique ce qui précède avec le produit scalaire défini par* A *et pas avec le produit scalaire canonique de* $\mathbb{R}^n$; *les colonnes de* P *définissent une nouvelle base*

de $\mathbb{R}^n$ qui est orthonormée pour le produit scalaire $(X,Y) \longmapsto {}^tXAY$ et orthogonale pour la forme bilinéaire symétrique $(X,Y) \longmapsto {}^tXBY$.

Dans l'exercice suivant on montre, comme première application, que si A, B sont deux matrices symétriques réelles avec $0 \leqslant A \leqslant B$ (au sens de l'ordre rencontré dans l'exercice 3.20) alors $\det A \leqslant \det B$.

3.28. Matrices symétriques positives telles que $A \leqslant B$

Soit A et B deux matrices symétriques réelles positives, q_A et q_B les formes quadratiques de $\mathbb{R}^n$ associées. On suppose $q_B \leqslant q_A$. Montrer que $\det B \leqslant \det A$.

(École polytechnique)

▷ **Solution.**

• Supposons d'abord que $\det A = 0$ c'est-à-dire que q_A soit dégénérée. Comme $\det B \geqslant 0$ on doit montrer que $\det B = 0$. Soit $X \in \operatorname{Ker} A$. On a alors $q_A(X) = {}^tXAX = 0$ et comme $0 \leqslant q_B(X) \leqslant q_A(X) = 0$, X est également un vecteur isotrope de q_B. Comme q_B est positive, l'inégalité de Cauchy-Schwarz montre que ${}^tYBX = 0$ pour tout vecteur $Y \in \mathbb{R}^n$, autrement dit que $BX = 0$. Donc q_B est également dégénérée et on a $\det B = \det A = 0$.

• Supposons maintenant q_A définie positive. Le théorème de pseudo-réduction simultanée garantit l'existence de $P \in \mathrm{GL}_n(\mathbb{R})$ telle que ${}^tPAP = I_n$ et ${}^tPBP = D$ avec $D = \operatorname{Diag}(d_1, \ldots, d_n)$ matrice diagonale. Comme $q_B \leqslant q_A$, on a pour tout $i \in [\![1, n]\!]$, $d_i = q_B(Pe_i) \leqslant q_A(Pe_i) = 1$ où $(e_1, \ldots, e_n)$ est la base canonique de $\mathbb{R}^n$. On a donc

$$(\det P)^2 \det B = \det D = \prod_{i=1}^{n} d_i \leqslant 1 = \det I_n = (\det P)^2 \det A$$

ce qui conduit à l'inégalité souhaitée en simplifiant par $(\det P)^2$.

Dans ce second cas, il y a égalité si, et seulement si, les d_i sont tous égaux à 1 i.e. si, et seulement si, $A = B$. ◁

L'exercice suivant est un énoncé plus précis qui implique le précédent.

3.29. Inégalité sur le déterminant de matrices symétriques

Soit u et v deux endomorphismes symétriques positifs d'un espace euclidien. Montrer que $\det(u+v) \geqslant \det u + \det v$. Dans quel cas a-t-on égalité ?

(École polytechnique)

▷ **Solution.**

Si $\det u = \det v = 0$, l'inégalité est évidente, car $u + v$ est un endomorphisme symétrique positif.

Supposons par exemple $\det u > 0$. Notons A et B les matrices respectives de u et v dans une base orthonormale quelconque de E. Les matrices A et B sont symétriques positives. A étant définie positive, D'après le théorème de pseudo-réduction simultanée, il existe $P \in GL_n(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale telles que $A = {}^tPP$ et $B = {}^tPDP$. On a alors $\det u = \det A = (\det P)^2$, $\det v = \det B = \det D(\det P)^2$ et $\det(u+v) = \det(A+B) = \det(I_n + D)(\det P)^2$. On est donc ramené à démontrer que $\det(I_n + D) \geqslant 1 + \det D$, c'est-à-dire

$$\prod_{i=1}^{n}(1+\lambda_i) \geqslant 1 + \prod_{i=1}^{n}\lambda_i.$$

Cette inégalité est évidente, car les λ_i sont positifs : en développant le produit qui est dans le membre de gauche de l'inégalité, on obtient le membre de droite plus des termes positifs.

Dans le second cas on a égalité si, et seulement si, tous les λ_i sont nuls c'est-à-dire si et seulement si $v = 0$. Ainsi, il y a égalité si, et seulement si, u, v et $u + v$ ne sont pas définis positifs ou si $u = 0$ ou $v = 0$. ◁

Dans le prochain exercice intervient la relation suivante sur l'espace des matrices symétriques réelles : $B > A \iff B - A$ *est définie positive. Ce n'est pas une relation d'ordre et il ne faut pas non plus la confondre avec l'ordre strict associé à la relation d'ordre de Löewner (voir exercice 4.20) défini par* $B \geqslant A$ *et* $B \neq A$. *Par exemple si on prend* $B = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ *et* $A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ *on a* $B \geqslant A$ *et* $B \neq A$ *mais pas* $B > A$.

3.30. Comparaison d'endomorphismes symétriques définis positifs

Soit E un espace euclidien. On définit sur l'ensemble des endomorphismes auto-adjoints définis positifs la relation

$$a > b \iff \forall x \in E \setminus \{0\},\ \ \langle a(x), x\rangle > \langle b(x), x\rangle.$$

Montrer que $a > b \iff \mathrm{Sp}(a^{-1}b) \subset]0, 1[$.

(École polytechnique)

▷ **Solution.**

On note A et B les matrices de a et b dans une base orthonormale $\mathcal{B}$ de

E. Elles sont symétriques définies positives. Il existe donc $P \in GL_n(\mathbb{R})$ et $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$, diagonale à termes diagonaux strictement positifs telles que

$$A = {}^tPP \quad \text{et} \quad B = {}^tPDP.$$

On a alors $A^{-1}B = P^{-1}\,{}^tP^{-1}\,{}^tPDP = P^{-1}DP$ qui est semblable à D donc est diagonalisable, avec des valeurs propres strictement positives. Ceci montre déjà que $\mathrm{Sp}(a^{-1}b) \subset \mathbb{R}_+^*$.

• Supposons maintenant $a > b$. Soit $\lambda \in \mathrm{Sp}(a^{-1}b)$ et x un vecteur propre associé. On a $a^{-1}b(x) = \lambda x$ et donc $b(x) = \lambda a(x)$. L'hypothèse $\langle a(x), x\rangle > \langle b(x), x\rangle = \lambda\langle a(x), x\rangle$ implique $\lambda < 1$, car $\langle a(x), x\rangle > 0$ pour $x \neq 0$. On a donc $\mathrm{Sp}(a^{-1}b) \subset]0, 1[$.

• Supposons réciproquement que $\mathrm{Sp}(a^{-1}b) = \mathrm{Sp}\,D \subset]0, 1[$. Pour tout $x \in E$, dont la matrice des coordonnées dans la base $\mathcal{B}$ est X, on a en notant ${}^t(PX) = {}^tY = (y_1, \ldots, y_n)$,

$$\langle a(x), x\rangle = {}^tXAX = {}^tX\,{}^tPPX = {}^tYY = \sum_{i=1}^{n} y_i^2 \quad \text{et}$$

$$\langle b(x), x\rangle = {}^tXBX = {}^tX\,{}^tPDPX = {}^tYDY = \sum_{i=1}^{n} \lambda_i y_i^2 < \langle a(x), x\rangle,$$

car les λ_i appartiennent à $]0, 1[$ et Y n'est pas nul. ◁

Le prochain exercice démontre que le déterminant est une fonction logarithmiquement concave sur le cône des matrices symétriques définies positives.

3.31. Convexité logarithmique

Soit A et B deux matrices symétriques réelles définies positives, α et β deux réels positifs tels que $\alpha + \beta = 1$. Montrer que

$$\det(\alpha A + \beta B) \geqslant (\det A)^{\alpha}(\det B)^{\beta}.$$

(École polytechnique)

▷ **Solution.**

Le théorème de pseudo-réduction simultanée affirme l'existence de $P \in GL_n(\mathbb{R})$ et de $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale réelle telles que

$$A = {}^tPP \quad \text{et} \quad B = {}^tPDP.$$

Les λ_i sont strictement positifs car B est définie positive. On a donc

$$(\det A)^{\alpha}(\det B)^{\beta} = (\det P^2)^{\alpha}(\det P^2 \det D)^{\beta} = \det P^2 (\det D)^{\beta},$$

puisque $\alpha + \beta = 1$, et $\det(\alpha A + \beta B) = \det P^2 \det(\alpha I_n + \beta D)$.

On est donc ramené à montrer que $\det(\alpha I_n + \beta D) \geqslant (\det(D))^{\beta}$, c'est-à-dire que

$$\prod_{i=1}^{n}(\alpha + \beta\lambda_i) \geqslant \left(\prod_{i=1}^{n}\lambda_i\right)^{\beta}$$

ou encore, en prenant le logarithme, que

$$\sum_{i=1}^{n}\ln(\alpha + \beta\lambda_i) \geqslant \beta\sum_{i=1}^{n}\ln\lambda_i.$$

Pour tout i, on a $\ln(\alpha+\beta\lambda_i) \geqslant \alpha\ln 1 + \beta\ln\lambda_i = \beta\ln\lambda_i$ par concavité du logarithme. Le résultat s'obtient en sommant ces inégalités pour i allant de 1 à n. ◁

Si $\alpha \in]0,1[$ et si $A \neq B$, un des λ_i est différent de 1 donc, par stricte concavité du logarithme, l'une des inégalités est stricte et

$$\det(\alpha A + \beta B) > (\det A)^{\alpha}(\det B)^{\beta}.$$

3.32. Inégalité de Minkowski

Soit A et B deux matrices symétriques définies positives d'ordre n. Montrer que

$$(\det A)^{1/n} + (\det B)^{1/n} \leqslant (\det(A + B))^{1/n}.$$

(École polytechnique)

▷ **Solution.**

Le théorème de pseudo-réduction simultanée affirme l'existence d'une matrice $P \in GL_n(\mathbb{R})$ et de $D = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$ diagonale réelle telles que

$$A = {}^tPP \quad \text{et} \quad B = {}^tPDP.$$

On a donc $\det A = \det {}^tP \det P = \det P^2$, $\det B = \det P^2 \det D$ et enfin $\det(A + B) = \det P^2 \det(I_n + D)$. L'inégalité à démontrer équivaut à $1 + (\det D)^{1/n} \leqslant (\det(I_n + D))^{1/n}$, c'est-à-dire à

$$1 + \left(\prod_{i=1}^{n}\lambda_i\right)^{1/n} \leqslant \left(\prod_{i=1}^{n}(1 + \lambda_i)\right)^{1/n}.$$

La relation est évidente si l'un des λ_i est nul. S'ils sont tous strictement positifs, elle équivaut à

$$\ln\left(1+e^{\frac{1}{n}\sum_{i=1}^{n}\ln\lambda_i}\right)\leqslant\frac{1}{n}\sum_{i=1}^{n}\ln(1+\lambda_i).$$

Cela résulte de la convexité sur $\mathbb{R}$ de la fonction $f : x \longmapsto \ln(1+e^x)$ appliquée aux points $\ln\lambda_1,\dots,\ln\lambda_n$, avec des coefficients tous égaux à $\frac{1}{n}$. On vérifie en effet que, pour tout $x\in\mathbb{R}$, $f''(x)=\frac{e^x}{(1+e^x)^2}>0$. ◁

Le lecteur trouvera une autre preuve de cette inégalité dans la note qui suit l'exercice 2.32.

L'énoncé suivant est très voisin de l'exercice 3.28.

3.33. Inégalité sur le déterminant de matrices symétriques

Soit $A_1,\dots,A_k$ des matrices symétriques positives de $\mathcal{M}_n(\mathbb{R})$ et $(\lambda_1,\dots,\lambda_k)\in\mathbb{R}^k$. Montrer que

$$\left|\det\left(\sum_{i=1}^{k}\lambda_iA_i\right)\right|\leqslant\det\left(\sum_{i=1}^{k}|\lambda_i|A_i\right).$$

(École normale supérieure)

▷ **Solution.**

Posons $H=\sum_{i=1}^{k}\lambda_iA_i$ et $K=\sum_{i=1}^{k}|\lambda_i|A_i$. La matrice K est symétrique positive, car somme de matrices symétriques positives. Pour tout $X\in\mathbb{R}^n$, on a

$$|{}^tXHX|=\left|\sum_{i=1}^{k}\lambda_i\,{}^tXA_iX\right|\leqslant\sum_{i=1}^{n}|\lambda_i|\,{}^tXA_iX\leqslant{}^tXKX.$$

Si K est inversible, et donc définie positive, l'inégalité $|\det H|\leqslant\det K$ équivaut à $|\det(K^{-1}H)|\leqslant 1$. D'après le théorème de pseudo-réduction simultanée, il existe $P\in GL_n(\mathbb{R})$ et $D=\mathrm{Diag}(\lambda_1,\dots,\lambda_n)$, diagonale réelle telles que

$$K={}^tPP\quad\text{et}\quad H={}^tPDP.$$

On a alors $K^{-1}H=P^{-1}\,{}^tP^{-1}\,{}^tPDP=P^{-1}DP$ qui est semblable à D, donc diagonalisable dans $\mathcal{M}_n(\mathbb{R})$. Cela montre déjà que $\mathrm{Sp}(K^{-1}H)\subset\mathbb{R}$. Soit $\lambda\in\mathrm{Sp}(K^{-1}H)$ et X un vecteur propre associé. On a $HX=\lambda KX$ et donc ${}^tXHX=\lambda\,{}^tXHX$. On en déduit

$$|\lambda|\,{}^t\mathrm{XKX} = |\,{}^t\mathrm{XHX}| \leqslant {}^t\mathrm{XKX}$$

et donc $|\lambda| \leqslant 1$ car ${}^t\mathrm{XKX} > 0$. Les valeurs propres de $\mathrm{K}^{-1}\mathrm{H}$ ont une valeur absolue inférieure à 1 donc $|\det(\mathrm{K}^{-1}\mathrm{H})| \leqslant 1$.

Le résultat lorsque $\det \mathrm{K} = 0$ s'obtient par un passage à la limite. On remplace H par $\mathrm{H} + \lambda \mathrm{I}_n$ ($\lambda > 0$) et K par $\mathrm{K} + \lambda \mathrm{I}_n$, ce qui revient à ajouter une $(k+1)$-ième matrice $\mathrm{A}_{k+1} = \mathrm{I}_n$. La matrice $\mathrm{K} + \lambda \mathrm{I}_n$ est définie positive et d'après ce qui précède, $|\det(\mathrm{H} + \lambda \mathrm{I}_n)| \leqslant \det(\mathrm{K} + \lambda \mathrm{I}_n)$. Le déterminant étant une fonction continue des coefficients, on obtient en faisant tendre λ vers 0, $|\det \mathrm{H}| \leqslant \det \mathrm{K}$. ◁

Dans l'exercice suivant on donne une condition nécessaire et suffisante pour que deux formes quadratiques dont l'une est non dégénérée puissent se réduire simultanément. L'énoncé est formulé avec $\mathbb{R}$ pour corps de base, mais on peut prendre n'importe quel corps $\mathbb{K}$ de caractéristique différente de 2 à la place.

3.34. Diagonalisation simultanée de deux formes quadratiques

Soit A et B deux matrices symétriques de $\mathcal{M}_n(\mathbb{R})$. On suppose A inversible et on note q_A et q_B les formes quadratiques associées à A et B. Montrer que les deux propositions suivantes sont équivalentes :

(*i*) il existe une base de $\mathbb{R}^n$ orthogonale pour q_A et pour q_B ;

(*ii*) $\mathrm{A}^{-1}\mathrm{B}$ est diagonalisable.

(École polytechnique)

Solution.

• (*i*) $\Longrightarrow$ (*ii*) est le sens facile. Par hypothèse, il existe une matrice inversible P telle que $\mathrm{D} = {}^t\mathrm{PAP}$ et $\mathrm{D}' = {}^t\mathrm{PBP}$ soient toutes les deux diagonales. On a alors

$$\mathrm{D}\ {}^1\mathrm{D}' = \mathrm{P}^{-1}\mathrm{A}^{-1}\,{}^t\mathrm{P}^{-1}\,{}^t\mathrm{PBP} = \mathrm{P}^{-1}\mathrm{A}^{-1}\mathrm{BP}$$

et $\mathrm{A}^{-1}\mathrm{B}$ est semblable à la matrice diagonale $\mathrm{D}^{-1}\mathrm{D}'$.

• Montrons la réciproque (*ii*) $\Longrightarrow$ (*i*). Soit λ une valeur propre de $\mathrm{A}^{-1}\mathrm{B}$. Si X est un vecteur propre associé à λ on a $\mathrm{BX} = \lambda \mathrm{AX}$ et donc $q_\mathrm{B}(\mathrm{X}) = \lambda q_\mathrm{A}(\mathrm{X})$. Autrement dit, la forme quadratique q_B est proportionnelle à q_A sur l'espace propre associé à λ. Une base de cet espace propre qui est orthogonale pour q_A l'est donc *a fortiori* aussi pour q_B. Pour conclure il suffit d'observer que les sous-espaces propres de $\mathrm{A}^{-1}\mathrm{B}$ sont orthogonaux deux à deux pour les deux formes quadratiques. En effet, soit λ et μ deux valeurs propres distinctes de $\mathrm{A}^{-1}\mathrm{B}$ et X, Y deux vecteurs propres associés. On a donc $\mathrm{BX} = \lambda \mathrm{AX}$ et $\mathrm{BY} = \mu \mathrm{AY}$. Ainsi.

$$\lambda\,{}^t\mathrm{XAY} = {}^t(\lambda\mathrm{AX})\mathrm{Y} = {}^t(\mathrm{BX})\mathrm{Y} = {}^t\mathrm{XBY} = \mu\,{}^t\mathrm{XAY}$$

et comme $\lambda \neq \mu$ on a ${}^t\mathrm{XAY} = 0$ et par suite ${}^t\mathrm{XBY} = 0$. Il suffit donc de prendre une base orthogonale pour q_A dans chaque espace propre de $\mathrm{A}^{-1}\mathrm{B}$: la base obtenue en réunissant ces bases est une base de E qui est orthogonale à la fois pour q_A et pour q_B. D'où le résultat. $\lhd$

Nous savons que (i) est vérifié lorsque A *est définie positive. On déduit donc de l'exercice que si* A, B *sont deux matrices symétriques réelles avec* A *définie positive, alors* $\mathrm{A}^{-1}\mathrm{B}$ *est diagonalisable (et donc* AB *également). C'est une question très classique (voir exercice 2.13).*

Mentionnons une autre condition suffisante de coréduction due à Milnor : si q et q' sont deux formes quadratiques sur $\mathbb{R}^n$ avec $n \geqslant 3$ dont les cônes isotropes ne se coupent qu'en 0, alors il existe une base de $\mathbb{R}^n$ orthogonale pour q et pour q'. Nous renvoyons le lecteur intéressé à l'article de Nicolas Tosel, Convexité et formes quadratiques réelles *dans la RMS 116-3 de mars 2006, page 11 pour un éclairage géométrique de ce résultat.*

Les exercices qui suivent concernent le groupe orthogonal d'une forme quadratique q sur un espace vectoriel E *de dimension finie. Comme dans le cas euclidien bien connu du lecteur ce groupe, noté* $\mathrm{O}(q)$, *est l'ensemble des automorphismes de* E *qui conservent la forme polaire b de q :*

$$\mathrm{O}(q) = \{u \in \mathrm{GL(E)},\quad \forall (x,y) \in \mathrm{E}^2,\quad b(u(x),u(y)) = b(x,y)\}.$$

C'est aussi $\{u \in \mathrm{GL(E)},\ q \circ u = q\}$ *comme on le voit à l'aide d'une formule de polarisation. Il est bien entendu immédiat de vérifier qu'il s'agit d'un sous-groupe de* GL(E). *Notons que q pouvant être dégénérée, il faut mettre l'inversibilité de u dans la définition (dans le cas non dégénéré, euclidien par exemple, la conservation du produit scalaire impose l'inversibilité). L'énoncé suivant montre que le commutant du groupe orthogonal d'une forme quadratique non dégénérée est toujours réduit aux homothéties.*

3.35. Commutant du groupe orthogonal d'une forme quadratique non dégénérée

Soit q une forme quadratique sur un espace vectoriel réel E de dimension finie, de forme polaire ϕ. On suppose que a est un élément de E tel que $q(a) \neq 0$. On pose $\mathrm{A} = \mathbb{R}a$ et $\mathrm{B} = \{x \in \mathrm{E},\ \phi(a,x) = 0\}$.

1. Montrer que $\mathrm{E} = \mathrm{A} \overset{\perp}{+} \mathrm{B}$.

2. Soit $G = \{u \in GL(E),\ \forall (x,y) \in E^2,\ \phi(u(x),u(y)) = \phi(x,y)\}$. Montrer que la symétrie par rapport à A parallèlement à B est dans G.

3. Soit $G' = \{v \in \mathcal{L}(E),\ \forall u \in G,\ v \circ u = u \circ v\}$ le commutant de G. Montrer que si $v \in G'$ alors $v(a)$ s'écrit $\lambda_a a$. Si $b \in E$ est tel que $q(b) \neq 0$, il existe de même un réel λ_b tel que $v(b) = \lambda_b b$. Montrer que $\lambda_b = \lambda_a$.

4. Déterminer G'.

(**École polytechnique**)

▷ **Solution.**

1. L'application $x \longmapsto \phi(a,x)$ est une forme linéaire sur E. Par hypothèse, elle est non nulle et B, son noyau, est donc un hyperplan de E. Comme $a \notin B$, la droite $\mathbb{R}a$ forme un supplémentaire de B. D'où le résultat. Bien entendu, B est simplement l'orthogonal de A au sens de ϕ.

2. Notons s la symétrie par rapport à A parallèlement à B. Soit x, y deux vecteurs de E. On écrit $x = \lambda a + b$ et $y = \lambda' a + b'$ avec $(b,b') \in B^2$. On a alors $s(x) = \lambda a - b$ et $s(y) = \lambda' a - b'$. En développant par bilinéarité, on obtient

$$\phi(s(x),s(y)) = \phi(\lambda a - b, \lambda' a - b') = \lambda\lambda'\phi(a,a) + \phi(b,b') = \phi(x,y).$$

Comme s est inversible on a bien $s \in G$.

De manière plus générale, si F *et* G *sont deux sous-espaces supplémentaires de* E *orthogonaux pour* q*, alors la symétrie par rapport à* F *parallèlement à* G *est dans* G.

3. Soit $v \in G'$. Comme v commute en particulier avec s, il doit laisser stable le sous-espace propre de s relatif à la valeur propre 1 c'est-à-dire la droite $\mathbb{R}a$ donc $v(a)$ s'écrit $\lambda_a a$.

Soit b tel que $q(b) \neq 0$. Si b est colinéaire à a, le résultat est évident : on peut écrire $b = ta$ avec $t \in \mathbb{R}$ et alors $v(b) = tv(a) = t\lambda_a a = \lambda_a b$.

On suppose donc que (a,b) est libre. Il existe un réel $t \neq 0$ tel que $\phi(b + ta, b + ta) \neq 0$ (c'est un polynôme du second degré en t dont le coefficient dominant est $\phi(a,a) = q(a) \neq 0$). Pour un tel t on pose $c = b + ta$. On a alors d'une part $v(c) = \lambda_c c = \lambda_c(b + ta)$ et d'autre part, $v(c) = v(b + ta) = v(b) + tv(a) = \lambda_b b + t\lambda_a a$. Comme (a,b) est libre on peut identifier : $\lambda_c = \lambda_b = \lambda_a$. D'où le résultat demandé.

4. Soit $v \in G'$. D'après la question précédente, il existe un réel λ tel que $v(x) = \lambda x$ pour tout $x \in E$ tel que $q(x) \neq 0$. Considérons maintenant un vecteur $x \in E$ tel que $q(x) = 0$. On peut toujours trouver $t \in \mathbb{R}$ tel que $q(ta + x) = \phi(ta + x, ta + x) \neq 0$. On a alors $v(ta + x) = \lambda(ta + x)$ ce qui donne, $v(x) = \lambda(ta + x) - v(ta) = \lambda x$. Donc, v est une homothétie. La réciproque étant triviale, on a ainsi $G' = \mathbb{R}\,\mathrm{Id}_E$. ◁

Le groupe orthogonal de la forme quadratique nulle est évidemment GL(E) *tout entier. Dans l'exercice qui suit on montre que le groupe orthogonal d'une forme quadratique non dégénérée est minimal au sens de l'inclusion. Pour cela on fera usage du résultat de l'exercice précédent.*

3.36. Minimalité du groupe orthogonal d'une forme non dégénérée

Soit E un espace vectoriel réel de dimension finie. On se donne q et q' deux formes quadratiques sur E, q étant non dégénérée.

1. Montrer qu'il existe $u \in \mathcal{L}(\mathrm{E})$ tel que, pour tout couple $(x, y) \in \mathrm{E}^2$, $b'(x, y) = b(u(x), y) = b(x, u(y))$, b et b' désignant les formes polaires de q et q'.

2. On suppose que $\mathrm{O}(q') \subset \mathrm{O}(q)$ où $\mathrm{O}(q)$ désigne le groupe orthogonal de q à savoir $\{u \in \mathrm{GL(E)},\ q \circ u = q\}$.

Montrer que $\mathrm{O}(q) = \mathrm{O}(q')$.

(École polytechnique)

▷ **Solution.**

1. Cette première question généralise un résultat bien connu dans le cas euclidien : la forme quadratique q' se représente à travers b à l'aide d'un endomorphisme auto-adjoint u. La preuve est d'ailleurs identique. Fixons $x \in \mathrm{E}$. L'application $b'_x : y \longmapsto b'(x, y)$ est une forme linéaire sur E. Comme b est non dégénérée, l'application $z \longmapsto b_z$ (où b_z est la forme linéaire $y \longmapsto b(z, y)$) est un isomorphisme de E sur E^*. Il existe donc un unique vecteur, que l'on note $u(x)$, tel que $b'_x = b_{u(x)}$, c'est-à-dire $b'(x, y) = b(u(x), y)$ pour tout $y \in \mathrm{E}$. Il est ensuite aisé de vérifier que l'application u ainsi définie est un endomorphisme de E. Comme b et b' sont symétriques on a, pour tout couple (x, y) de E^2,

$$b(x, u(y)) = b(u(y), x) = b'(y, x) = b'(x, y) = b(u(x), y).$$

2. Soit $v \in \mathrm{O}(q')$ et $(x, y) \in \mathrm{E}^2$. On a $b'(v(x), v(y)) = b'(x, y)$ ce qui conduit à

$$b((u \circ v)(x), v(y)) = b(u(x), y) = b((v \circ u)(x), v(y)),$$

car v est aussi élément de $\mathrm{O}(q)$. Comme cela vaut pour tout couple (x, y) et comme v est surjective et b non dégénérée, on en déduit que $u \circ v = v \circ u$. Ainsi u commute avec tous les éléments de $\mathrm{O}(q')$ et l'exercice précédent montre que u est forcément une homothétie. Il en découle que q' est proportionnelle à q. De plus, q' est non nulle car sinon $\mathrm{O}(q)$ serait égal à GL(E), ce qui n'est pas le cas (par exemple $2\,\mathrm{Id}_\mathrm{E} \notin \mathrm{O}(q)$, puisque $q \neq 0$). On a donc $q' = \alpha q$ avec $\alpha \in \mathbb{R}^*$ et donc clairement $\mathrm{O}(q) = \mathrm{O}(q')$. ◁

Formulé géométriquement, l'énoncé suivant fait bien appel aux formes quadratiques : en effet, un ellipsoïde de $\mathbb{R}^n$ de centre l'origine est défini par une inéquation du type $q(x) \leqslant 1$ où q est une forme quadratique définie positive. On y utilisera le résultat de l'exercice 3.31.

3.37. Ellipsoïde de John-Loewner

Soit K un compact d'intérieur non vide de $\mathbb{R}^n$. Montrer qu'il existe un unique ellipsoïde centré en 0 de volume minimal contenant K.

(École normale supérieure)

▷ **Solution.**

On munit $\mathbb{R}^n$ de sa structure euclidienne usuelle. Comme nous venons de le dire en préambule, un ellipsoïde plein centré en 0 de $\mathbb{R}^n$ a une équation du type $q(x) \leqslant 1$, où q est une forme quadratique définie positive. On notera Q (resp. Q_+, resp. Q_{++}) l'ensemble des formes quadratiques (resp. positives, resp. définies positives) de $\mathbb{R}^n$ et pour $q \in \mathrm{Q}_{++}$, on pose $\mathcal{E}_q = \{x \in \mathbb{R}^n.\ \ q(x) \leqslant 1\}$.

Commençons par calculer le volume V_q de $\mathcal{E}_q$. Il existe une base orthonormale $\mathcal{B} = (e_1, \ldots, e_n)$ dans laquelle q s'écrit $q(x) = \sum_{i=1}^{n} a_i x_i^2$. On obtient

$$\mathrm{V}_q = \iint \cdots \int_{a_1 x_1^2 + \cdots + a_n x_n^2 \leqslant 1} \mathrm{d}x_1 \ldots \mathrm{d}x_n.$$

On considère le changement de variables défini par $x_i = \dfrac{t_i}{\sqrt{a_i}}$ (c'est bien un $\mathcal{C}^1$-difféomorphisme) dont le jacobien est $\dfrac{1}{\sqrt{a_1 \ldots a_n}}$. On observe que si S est la matrice de q dans une base orthonormale quelconque de $\mathbb{R}^n$, il existe $\mathrm{P} \in \mathrm{O}_n(\mathbb{R})$ telle que $\mathrm{S} = \mathrm{P}\,\mathrm{Diag}(a_1, \ldots, a_n)\,{}^t\mathrm{P}$. On a donc $\det \mathrm{S} = \det \mathrm{Diag}(a_1, \ldots, a_n) = a_1 \ldots a_n$. Ce déterminant ne dépend donc pas de la base orthonormale de $\mathbb{R}^n$ choisie. Notons-le $\mathrm{D}(q)$. Le changement de variables dans l'intégrale multiple donne

$$\mathrm{V}_q = \iint \cdots \int_{x_1^2 + \cdots x_n^2 \leqslant 1} \frac{\mathrm{d}x_1 \ldots \mathrm{d}x_n}{\sqrt{\mathrm{D}(q)}} = \frac{\mathrm{V}_0}{\sqrt{\mathrm{D}(q)}},$$

où V_0 est le volume de la boule unité pour la norme euclidienne canonique.

Le problème peut donc se reformuler ainsi : il s'agit de montrer que si K est un compact d'intérieur non vide de $\mathbb{R}^n$, il existe une unique

forme quadratique $q \in \mathrm{Q}_{++}$ telle que $\mathrm{D}(q)$ soit maximal et que pour tout $x \in \mathrm{K}$, $q(x) \leqslant 1$.

On munit l'espace Q de la norme N définie par $\mathrm{N}(q) = \sup_{\|x\| \leqslant 1} |q(x)|$. Il est alors naturel de considérer l'ensemble

$$\mathcal{A} = \{q \in \mathrm{Q}_+, \ \ \forall x \in \mathrm{K}, \ \ q(x) \leqslant 1\}$$

puis de chercher à maximiser D sur ce domaine. Montrons que $\mathcal{A}$ est un compact convexe non vide de Q.

• $\mathcal{A}$ est convexe : soit q et q' dans $\mathcal{A}$ et $\lambda \in [0,1]$. Pour tout $x \in \mathbb{R}^n$,

$$(\lambda q + (1-\lambda)q')(x) = \lambda q(x)) + (1-\lambda)q'(x) \geqslant 0$$

et $\lambda q + (1-\lambda)q'$ est une forme quadratique positive. De plus, si $x \in \mathrm{K}$,

$$(\lambda q + (1-\lambda)q')(x) \leqslant \lambda q(x) + (1-\lambda)q'(x) \leqslant \lambda + 1 - \lambda = 1$$

ce qui entraîne $\lambda q + (1-\lambda)q' \in \mathcal{A}$.

• $\mathcal{A}$ est fermé : si $(q_n)_{n\in\mathbb{N}}$ est une suite de $\mathcal{A}$ convergente dans Q vers q, on a pour tout $x \in \mathbb{R}^n$, $|q(x) - q_n(x)| \leqslant \mathrm{N}(q - q_n)\|x\|$ donc $\lim_{n\to+\infty} q_n(x) = q(x)$. On en déduit que

$$\forall x \in \mathbb{R}^n, \ \ q(x) = \lim_{n\to+\infty} q_n(x) \geqslant 0 \quad \text{et } \forall x \in \mathrm{K}, \ \ q(x) = \lim_{n\to+\infty} q_n(x) \leqslant 1,$$

donc $q \in \mathcal{A}$.

• Montrons que $\mathcal{A}$ est borné. Comme K est d'intérieur non vide, il existe $a \in \mathrm{K}$ et $r > 0$ tel que $\mathrm{B}(a,r) \subset \mathrm{K}$. Soit $q \in \mathcal{A}$. Si $\|x\| \leqslant r$ alors $a + x \in \mathrm{K}$ donc $q(a+x) \leqslant 1$; d'autre part, on a $q(-a) = q(a) \leqslant 1$. Par l'inégalité de Minkowski, on obtient

$$\sqrt{q(x)} = \sqrt{q(x+a-a)} \leqslant \sqrt{q(x+a)} + \sqrt{q(-a)} \leqslant 2$$

donc $q(x) \leqslant 4$. Si $\|x\| \leqslant 1$, $|q(x)| = q(x) = \frac{1}{r^2} q(rx) \leqslant \frac{4}{r^2}$ ce qui montre que $\mathrm{N}(q) \leqslant \frac{4}{r^2}$.

• Montrons enfin que $\mathcal{A}$ est non vide. Puisque K est compact, il est borné : soit $\mathrm{M} > 0$ tel que pour tout $x \in \mathrm{K}$, $\|x\| \leqslant \mathrm{M}$. Alors si q est définie par $q(x) = \frac{\|x\|^2}{\mathrm{M}^2}$, on a, pour tout $x \in \mathrm{K}$, $q(x) \leqslant 1$.

Bilan : $\mathcal{A}$ est un compact convexe non vide de Q.

L'application déterminant est continue donc $q \longmapsto \mathrm{D}(q)$ est continue sur le compact $\mathcal{A}$. Elle atteint donc son maximum sur $\mathcal{A}$, en q_0. Comme $\mathcal{A}$ contient $x \longmapsto \frac{\|x\|^2}{\mathrm{M}^2}$ qui est définie positive, on a $\mathrm{D}(q_0) > 0$ donc $q_0 \in \mathrm{Q}_{++}$. Nous venons de prouver qu'il existe un ellipsoïde $\mathcal{E}_{q_0}$ de volume minimal qui contient K.

Il reste à prouver que cet ellipsoïde est unique *i.e.* que q_0 est unique. Supposons qu'il existe $q \in \mathcal{A}$ tel que $D(q) = D(q_0)$ et $q \neq q_0$. Soit S et S_0 les matrices de q et q_0 dans la base canonique de $\mathbb{R}^n$. Comme $\mathcal{A}$ est convexe, $\frac{1}{2}(q + q_0)$ appartient à $\mathcal{A}$ et d'après l'exercice 3.31 (convexité logarithmique du déterminant sur l'ensemble des matrices symétriques définies positives). on a

$$D\left(\frac{1}{2}(q+q_0)\right) = \det\left(\frac{1}{2}(S+S_0)\right) > (\det S)^{\frac{1}{2}}(\det S_0)^{\frac{1}{2}} \geqslant \det S_0 \geqslant D(q_0).$$

ce qui contredit la maximalité de $D(q_0)$.

Conclusion. Il existe un unique ellipsoïde centré en 0 de volume minimal contenant K. ◁

Le résultat reste vrai dans un espace vectoriel réel de dimension finie quelconque, avec la même démonstration (il suffit de munir E *d'une structure euclidienne quelconque). Un corollaire important est donné par la première question de l'exercice suivant qui montre qu'un sous-groupe compact du groupe linéaire est forcément un sous-groupe du groupe orthogonal pour une structure euclidienne de* E. *Notons que le cas particulier des groupes finis se traite beaucoup plus simplement (voir l'exercice 4.40 du chapitre Géométrie).*

3.38. Sous-groupes compacts maximaux de GL(E)

Soit E un $\mathbb{R}$-espace vectoriel de dimension finie $n \geqslant 1$.

1. Soit G un sous-groupe compact de GL(E). Montrer qu'il existe un produit scalaire sur E, associé à une forme quadratique q, tel que $G \subset O(q)$.

2. En déduire que les sous-groupes compacts de GL(E) qui sont maximaux pour l'inclusion sont exactement les groupes orthogonaux $O(q)$ où q est une forme quadratique définie positive.

3. *Application.* Soit N une norme sur E et G_N le groupe des isométries de E pour cette norme. Montrer que si G_N opère de façon transitive sur la sphère de centre 0 et de rayon 1, alors la norme est euclidienne.

(École normale supérieure)

▷ **Solution.**

1. Munissons E d'une norme euclidienne $\| \ \|$ quelconque et notons B la boule fermée de centre 0 et de rayon 1 pour cette norme. Considérons l'ensemble $K = \{g(x),\ g \in G \ \text{ et } \ x \in B\} = \bigcup_{g \in G} g(B)$. Il s'agit d'un compact de E en tant qu'image du compact $K \times B$ par l'application continue $(g, x) \longmapsto g(x)$. Il est d'intérieur non vide puisqu'il contient B et

invariant par tout élément de G. On considère (voir l'exercice précédent dont on reprend les notations) l'unique ellipsoïde $\mathcal{E}_q$ de volume minimal contenant K où q est une forme quadratique définie positive. Pour tout $g \in \mathrm{G}$, la forme quadratique q' définie par $q'(x) = q(g(x))$ est définie positive et $\mathrm{K} \subset \mathcal{E}_{q'}$ car $g(\mathrm{K}) = \mathrm{K}$. Mais d'autre part on a, puisque G est compact, $|\det g| = 1$. En effet l'application det est bornée sur G donc sur $\{g^p,\ p \in \mathbb{Z}\}$. On a donc $\mathrm{D}(q') = \mathrm{D}(q)$ et par unicité $q' = q$. Ainsi $g \in \mathrm{O}(q)$ pour tout $g \in \mathrm{G}$, donc $\mathrm{G} \subset \mathrm{O}(q)$.

2. Soit G un sous-groupe compact de GL(E) maximal pour l'inclusion. D'après la première question, il existe une forme quadratique q définie positive telle que $\mathrm{G} \subset \mathrm{O}(q)$ et comme $\mathrm{O}(q)$ est compact, on a $\mathrm{G} = \mathrm{O}(q)$ par maximalité.

Réciproquement, si $\mathrm{G} = \mathrm{O}(q)$ avec q définie positive, alors G est compact et si G′ est un sous-groupe compact de GL(E) tel que $\mathrm{G} \subset \mathrm{G}'$ il existe, d'après ce qui précède, une forme quadratique définie positive q' telle que $\mathrm{G}' \subset \mathrm{O}(q')$. On a donc $\mathrm{O}(q) \subset \mathrm{O}(q')$ et d'après le résultat de l'exercice 3.36, cela implique $\mathrm{O}(q) = \mathrm{O}(q')$ et donc $\mathrm{G} = \mathrm{G}'$. Le sous-groupe $\mathrm{G} = \mathrm{O}(q)$ est compact maximal.

3. Par définition $\mathrm{G_N}$ est le groupe des automorphismes u de E vérifiant $\mathrm{N}(u(x)) = \mathrm{N}(x)$ pour tout $x \in \mathrm{E}$. Il est clair que $\mathrm{G_N}$ est fermé et borné puisque $|||u||| = 1$ pour tout $u \in \mathrm{G_N}$ où la norme triple est associée à N. Il s'agit donc d'un sous-groupe compact du groupe linéaire GL(E). La première question nous donne une structure euclidienne associée à une forme quadratique définie positive q telle que $\mathrm{G} \subset \mathrm{O}(q)$. Notons $\mathrm{S_N} = \{x \in \mathrm{E},\ \mathrm{N}(x) = 1\}$ la sphère unité de E et soit x_0 un vecteur quelconque de cette sphère. Par hypothèse, si $x \in \mathrm{S_N}$ il existe $g \in \mathrm{G_N}$ tel que $x = g(x_0)$. On a alors

$$q(x) = q(g(x_0)) = q(x_0)$$

car $g \in \mathrm{O}(q)$. Autrement dit, q est constante sur $\mathrm{S_N}$. Si $x \in \mathrm{E}$ est non nul, le vecteur $\dfrac{x}{\mathrm{N}(x)}$ est dans $\mathrm{S_N}$ et $q\left(\dfrac{x}{\mathrm{N}(x)}\right) = q(x_0)$ donc par homogénéité $q(x) = q(x_0)\mathrm{N}(x)^2$. On en déduit que N est positivement colinéaire à la norme euclidienne $\sqrt{q}$: c'est donc aussi une norme euclidienne. ◁

3.39. Endomorphisme auto-adjoint sur un espace quadratique complexe

1. Une matrice symétrique A de $\mathcal{M}_n(\mathbb{C})$ est-elle forcément diagonalisable ?

2. Soit E un espace vectoriel complexe de dimension finie et b une forme bilinéaire symétrique non dégénérée sur E.

Par analogie avec la structure euclidienne, définir l'adjoint u^* d'un endomorphisme u de E.

3. Trouver une condition nécessaire et suffisante sur E pour que tout endomorphisme auto-adjoint de E soit diagonalisable.

(École normale supérieure)

▷ **Solution.**

1. La réponse est non si $n \geqslant 2$. Prenons par exemple une matrice symétrique de taille $(2,2)$, $A = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$. Son polynôme caractéristique est $X^2 - (a+c)X + ac - b^2$ dont le discriminant est

$$\Delta = (a+c)^2 - 4ac + 4b^2 = (a-c)^2 + 4b^2.$$

Si on travaille sur le corps des réels on voit que Δ ne peut être nul que si $a = c$ et $b = 0$ et dans ce cas A est une matrice scalaire. Mais cela n'est plus le cas sur $\mathbb{C}$. Prenons par exemple $a = 1$, $c = -1$ et $b = i$ c'est-à-dire $A = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$. On a $\Delta = 0$ donc A admet une seule valeur propre mais comme A n'est pas scalaire elle n'est pas diagonalisable. Il est immédiat d'en déduire un contre-exemple pour toute taille $n \geqslant 2$.

2. On procède comme dans le cas euclidien. Soit $y \in E$. L'application $x \longmapsto b(u(x), y)$ est une forme linéaire sur E et comme b est non dégénérée il existe un unique vecteur $u^*(y)$ tel que $b(u(x), y) = b(u^*(y), x)$ pour tout x de E. Cela définit une application $u^* : E \longrightarrow E$ et il est immédiat de vérifier qu'elle est linéaire.

3. Choisissons une base $\mathcal{B}$ de E orthonormée pour b. Soit u un endomorphisme de E et A la matrice de u dans la base $\mathcal{B}$. La matrice de u^* dans $\mathcal{B}$ est la transposée de A. Ainsi, tout endomorphisme auto-adjoint de E est diagonalisable si, et seulement si, toute matrice symétrique de $\mathcal{M}_n(\mathbb{C})$ est diagonalisable (où $n = \dim E$). La question 1 montre que cela n'est le cas que si $n = 1$. ◁

La démonstration effectuée dans l'exercice suivant est à rapprocher de celle de l'exercice 2.66.

3.40. Réduction des matrices symétriques complexes

Soit M une matrice symétrique complexe. Montrer qu'il existe une matrice unitaire O telle que ${}^t\overline{O}MO$ soit diagonale, à termes diagonaux réels positifs ou nuls.

(École normale supérieure)

▷ **Solution.**

On va démontrer la propriété par récurrence sur n. Pour $n = 1$, il existe $m \in \mathbb{C}$ tel que $M = (m)$ et un réel θ tel que $m = |m|e^{i\theta}$. On obtient $(m) = (e^{i\frac{\theta}{2}})(|m|)(e^{i\frac{\theta}{2}})$, qui est la décomposition cherchée.

On suppose que la propriété est vraie au rang $n-1$ et on considère M symétrique de $\mathcal{M}_n(\mathbb{C})$. Si O est une matrice unitaire ayant la propriété voulue et X_1 le premier vecteur de la base canonique de $\mathbb{C}^n$, il existe $\lambda \in \mathbb{R}_+$ tel que ${}^tOMOX_1 = \lambda X_1$ et donc $MOX_1 = \lambda\overline{O}X_1 = \lambda\overline{OX_1}$. On va donc chercher un vecteur X tel que $MX = \lambda\overline{X}$. Si cela est réalisé, on a alors $\overline{M}MX = \lambda\overline{MX} = \lambda^2 X$ et X est un vecteur propre de $\overline{M}M$.

Il faut remarquer qu'en fait $\overline{M}M$ est hermitienne positive. En effet, on a $(\overline{M}M)^* = M^*(\overline{M})^* = \overline{M}M$. car M est symétrique, et pour tout $X \in \mathbb{C}^n$, $X^*\overline{M}MX = (MX)^*(MX) \geqslant 0$. Les valeurs propres de $\overline{M}M$ sont donc réelles et positives.

Soit donc $\mu \in \mathbb{R}_+$ une valeur propre de $\overline{M}M$. X un vecteur propre associé.

Si la famille $(X, \overline{MX})$ est $\mathbb{R}$-liée, il existe $\lambda \in \mathbb{R}$ tel que $\overline{MX} = \lambda X$ (on peut montrer que $\lambda^2 = \mu$). Si $\lambda < 0$, on remarque que $\overline{Mi}X = -\lambda i X$.

Si la famille $(X, \overline{MX})$ est $\mathbb{R}$-libre, on pose $Y = \lambda X + \overline{MX}$, où $\lambda = \sqrt{\mu}$. Le vecteur Y n'est pas nul et on obtient alors

$$\overline{M}Y = \lambda\overline{MX} + \overline{M}MX = \lambda\overline{MX} + \mu X = \lambda Y.$$

Dans tous les cas, on a déterminé un vecteur X non nul et $\lambda \in \mathbb{R}_+$ tel que $\overline{MX} = \lambda X$ et donc $MX = \lambda\overline{X}$. En le multipliant par une constante réelle, on peut supposer de plus que X est de norme (hermitienne) 1. Soit U une matrice unitaire telle que $UX_1 = X$ et $N = {}^tUMU$. La matrice N est également symétrique et

$$NX_1 = {}^tUMUX_1 = {}^tUMX = \lambda\,{}^tU\overline{X} = \lambda\overline{U^{-1}X} = \lambda\overline{X_1} = \lambda X_1.$$

Ainsi, il existe une matrice symétrique de $N_1 \in \mathcal{M}_{n-1}(\mathbb{C})$ telle que

$$N = \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & N_1 & \\ 0 & & & \end{pmatrix}.$$

On applique l'hypothèse de récurrence à N_1. Il existe une matrice O_1 unitaire de taille $n-1$ telle que ${}^tO_1N_1O_1 = D_1$, matrice diagonale à coefficients positifs ou nuls. On pose $O' = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & O_1 & \\ 0 & & & \end{pmatrix}$. La ma-

trice O' est unitaire et $D = {}^tO'NO' = \begin{pmatrix} \lambda & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & D_1 & \\ 0 & & & \end{pmatrix}$ est diagonale à coefficients positifs. Finalement, on obtient

$$D = {}^tO'NO' = {}^tO'\,{}^tUMUO' = {}^t(UO')M(UO'),$$

ce qui est le résultat voulu, car $O = UO'$ est unitaire. ◁

Il s'agit d'une « pseudo-réduction » : la matrice M *est congruente à une matrice diagonale à coefficients positifs, la matrice de passage étant unitaire.*

Le résultat obtenu peut encore s'énoncer ainsi : pour toute forme quadratique q *sur* $\mathbb{C}^n$*, il existe une base orthonormale pour le produit hermitien ordinaire dans laquelle* $q(x) = \sum_{i=1}^{n} a_i x_i^2$*, où* $a_1, \dots, a_n$ *sont des réels positifs.*

Nous terminons ce chapitre avec divers exercices touchant aux formes quadratiques et qui sont relativement inclassables.

3.41. Une décomposition

Soit $n \in \mathbb{N}^*$ et $(x, y, z) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$. On pose

$$M(x,y,z) = \begin{pmatrix} x & y_1 & y_2 & \dots & \dots & y_n \\ y_1 & z_1 & 0 & \dots & \dots & 0 \\ y_2 & 0 & z_2 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ y_{n-1} & 0 & \dots & 0 & z_{n-1} & 0 \\ y_n & 0 & \dots & \dots & 0 & z_n \end{pmatrix}.$$

1. Calculer $\det M(x, y, z)$.

2. On pose $S = \{M(x, y, z) \text{ positive}\}$, $S_0 = \{M(x, 0, 0),\ x \geqslant 0\}$ et pour tout $1 \leqslant i \leqslant n$,

$$S_i = \{M(x, (0, \dots, y_i, \dots, 0), (0, \dots, z_i, \dots, 0)),\ x \geqslant 0,\ z_i \geqslant 0,\ xz_i = y_i^2\}.$$

Montrer que $S = S_0 + S_1 + \cdots + S_n$.

(École polytechnique)

▷ **Solution.**

1. Supposons le produit $z_1 z_2 \dots z_n$ non nul. On peut alors se ramener au calcul du déterminant d'une matrice triangulaire. Pour $2 \leqslant j \leqslant n+1$, on multiplie la $j^{\text{ième}}$ colonne de $\mathrm{M}(x,y,z)$ par $-\dfrac{y_j}{z_j}$ et on l'ajoute à la première colonne. On obtient

$$\det \mathrm{M}(x,y,z) = \begin{vmatrix} x - \displaystyle\sum_{j=1}^{n} \frac{y_j^2}{z_j} & y_1 & \dots & y_n \\ 0 & z_1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & z_n \end{vmatrix}.$$

On a alors $\det \mathrm{M}(x,y,z) = \left(x - \displaystyle\sum_{j=1}^{n} \frac{y_j^2}{z_j} \right) \prod_{i=1}^{n} z_i$, et donc

$$\boxed{\det \mathrm{M}(x,y,z) = x \prod_{i=1}^{n} z_i - \sum_{i=1}^{n} y_i^2 \prod_{j \neq i} z_j}.$$

L'application $(x,y,z) \longmapsto \det \mathrm{M}(x,y,z)$ est continue car polynomiale, et $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^{*n}$ est dense dans $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$. On en déduit que la formule précédente est vérifiée pour tout (x,y,z) de $\mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n$.

2. On va procéder par double inclusion. Vérifions, pour commencer, que chaque S_i (pour $0 \leqslant i \leqslant n$) est inclus dans S. Soit $\mathrm{M} = \mathrm{M}(x,y,z)$, q la forme quadratique dont la matrice dans la base canonique de $\mathbb{R}^{n+1}$ est M et $t = (t_0, t_1, \dots, t_n) \in \mathbb{R}^{n+1}$.

Si M est dans S_0, alors $q(t) = x t_1^2 \geqslant 0$ et M est positive.

Si M est dans S_i, pour $1 \leqslant i \leqslant n$, alors $q(t) = x t_0^2 + 2 y_i t_0 t_i + z_i t_i^2$.

- Si $x = 0$, alors $y_i = 0$, par définition de S_i et $q(t) = z_i t_i^2 \geqslant 0$.
- Si $x \neq 0$, alors

$$\begin{aligned} q(t) &= x \left(t_0^2 + 2 \frac{y_i}{x} t_0 t_i \right) + z_i t_i^2 \\ &= x \left(t_0 + \frac{y_i z_i}{x} \right)^2 + \underbrace{\left(z_i - \frac{y_i^2}{x} \right)}_{=0} t_i^2 \\ &= x \left(t_0 + \frac{y_i z_i}{x} \right)^2 \geqslant 0, \end{aligned}$$

ce qui montre que M est positive. Chaque S_i est bien inclus dans S. Comme l'application

$$(x,y,z) \in \mathbb{R} \times \mathbb{R}^n \times \mathbb{R}^n \longmapsto \mathrm{M}(x,y,z) \in \mathcal{M}_{n+1}(\mathbb{R})$$

est linéaire, $\{M(x,y,z),(x,y,z)\in\mathbb{R}\times\mathbb{R}^n\times\mathbb{R}^n\}$ est un sous-espace vectoriel de $\mathcal{M}_{n+1}(\mathbb{R})$. De plus, une somme de matrices symétriques positives étant symétrique positive, on en déduit $S_0+\cdots+S_n\subset S$.

Montrons maintenant que toute matrice de S s'écrit comme somme de matrices appartenant à $S_0,S_1,\ldots,S_n$. Soit $M=M(x,y,z)\in S$ et q la forme quadratique de $\mathbb{R}^{n+1}$ associée dans la base canonique $(e_0,e_1,\ldots,e_n)$.

Pour $1\leqslant i\leqslant n$, les coefficients y_i et z_i qu'il nous faut mettre dans l'élément de S_i sont ceux qui se trouvent dans la matrice M. On a simplement le choix du coefficient en position $(1,1)$. La matrice de S_0 sera alors fixée de manière unique : il nous faudra vérifier qu'elle convient.

Remarquons que $q(e_0)=x\geqslant 0$, et pour $1\leqslant i\leqslant n$, $z_i=q(e_i)\geqslant 0$. De plus $q_{|\mathrm{Vect}(e_0,e_i)}$ est positive donc son déterminant est positif : $xz_i-y_i^2\geqslant 0$. On en déduit que si $z_i=0$, alors $y_i=0$.

Considérons alors $I=\{i\in[\![1,n]\!],z_i\neq 0\}$. Pour $i\in I$, posons $x_i=\dfrac{y_i^2}{z_i}$ et $M_i=M(x_i,(0,\ldots,y_i,\ldots,0),(0,\ldots,z_i,\ldots,0))$. Sachant que $z_i>0$ et étant donné le choix de x_i, M_i est dans S_i. Pour $i\notin I$, on pose $M_i=0$, qui est encore dans S_i.

On a alors $\displaystyle\sum_{i=1}^{n}M_i=M\left(\sum_{i\in I}x_i,y,z\right)$. Cette égalité résulte de la linéarité de l'application $(x,y,z)\longmapsto M(x,y,z)$ et du fait que $y_i=z_i=0$ si $i\notin I$.

Pour obtenir $\displaystyle M=\sum_{i=0}^{n}M_i$, il faut prendre $\displaystyle M_0=M\left(x-\sum_{i\in I}x_i,0,0\right)$. Il reste à montrer que cette matrice est dans S_0, *i.e.* que $\displaystyle x-\sum_{i\in I}x_i\geqslant 0$.

Considérons la matrice M' obtenue en enlevant à M les lignes et les colonnes d'indice i tel que $1\leqslant i\leqslant n$ et $i\notin I$. La matrice M' est la restriction de q à $\mathrm{Vect}(e_i)_{i\in\{0\}\cup I}$ et elle est donc positive. Son déterminant est positif. Il a la même forme que celui de M et on trouve

$$\begin{aligned}\det M' &= x\prod_{i\in I}z_i-\sum_{i\in I}y_i^2\prod_{j\in I,j\neq i}z_j\\ &\quad -x\prod_{j\in I}z_j-\sum_{i\in I}x_i\prod_{j\in I}z_j \text{ par définition de } x_i\\ &=\left(\prod_{j\in I}z_j\right)\left(x-\sum_{i\in I}x_i\right).\end{aligned}$$

Ce déterminant est positif et, de plus, pour tout $j\in I$, on a $z_j>0$. On en déduit $\displaystyle x-\sum_{i\in I}x_i\geqslant 0$, ce qui permet de conclure la démonstration. ◁

L'exercice suivant est un joli problème de recherche d'extremum.

3.42. Maximum d'une forme quadratique sur un compact

Soit $\mathrm{K} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n,\ \forall i,\ x_i \geqslant 0,\ \sum_{i=1}^{n} x_i = 1\}$.

1. Déterminer la borne supérieure sur K de la forme quadratique définie par $q(x_1, \ldots, x_n) = \sum_{i \neq j} x_i x_j$.

2. Même question pour la forme quadratique q_{A} définie par $q_{\mathrm{A}}(x_1, \ldots, x_n) = \sum_{(i,j) \in \mathrm{A}} x_i x_j$, où A est une partie symétrique non vide de $[\![1, n]\!]^2$.

(École polytechnique)

▷ **Solution.**

1. Nous avons pour $(x_1, \ldots, x_n) \in \mathrm{K}$,

$$q(x_1, \ldots, x_n) = \left(\sum_{i=1}^{n} x_i\right)^2 - \sum_{i=1}^{n} x_i^2 = 1 - \sum_{i=1}^{n} x_i^2 \leqslant 1 - \frac{1}{n},$$

car, par l'inégalité de Cauchy-Schwarz, $1 = \left(\sum_{i=1}^{n} x_i\right)^2 \leqslant n \sum_{i=1}^{n} x_i^2$. La borne supérieure de q sur K est donc $1 - \frac{1}{n}$: c'est un maximum atteint pour le vecteur dont toutes les coordonnées valent $\frac{1}{n}$ et seulement en ce point (cas d'égalité dans l'inégalité de Cauchy-Schwarz).

2. Notons que la borne supérieure est toujours un maximum, car q_{A} est continue et K est compact. Par ailleurs, si $(x_1, \ldots, x_n) \in \mathrm{K}$, on a

$$q_{\mathrm{A}}(x_1, \ldots, x_n) \leqslant \sum_{1 \leqslant i,j \leqslant n} x_i x_j = \left(\sum_{k=1}^{n} x_k\right)^2 = 1$$

de sorte que le maximum est toujours inférieur ou égal à 1. Cela permet de conclure dans le cas où la partie A contient un élément (k, k) de la diagonale D du carré $\mathrm{C} = [\![1, n]\!]^2$ car on a alors $q_{\mathrm{A}}(e_k) = 1$, où e_k désigne le k-ième vecteur de la base canonique de $\mathbb{R}^n$. Nous supposerons donc dans la suite que A ne contient aucun terme de la diagonale. Soit q le cardinal maximal d'une partie I de $[\![1, n]\!]$ telle que $\mathrm{I} \times \mathrm{I} \setminus \mathrm{D} \subset \mathrm{A}$ (comme A est non vide on a $q \geqslant 1$). D'après les calculs vus dans la première question on a

$$q_{\mathrm{A}}\left(\frac{1}{q} \sum_{i \in \mathrm{I}} e_i\right) = 1 - \frac{1}{q}$$

donc $\max_{\mathrm{K}} q_{\mathrm{A}} \geqslant 1 - \dfrac{1}{q}$. Nous allons montrer qu'il s'agit bien du maximum de q_{A} sur K. Soit $(x_1, \ldots, x_n)$ un point de K où le maximum de q_{A} est atteint et supposons par l'absurde qu'il existe deux indices $h \neq k$ tels que $x_h x_k \neq 0$ et $(h,k) \notin \mathrm{A}$. On peut alors écrire

$$q_{\mathrm{A}}(x_1, \ldots, x_n) = \alpha x_h + \beta x_k + \gamma$$

où les coefficients α, β, γ ne dépendent que des réels x_i avec $i \notin \{h,k\}$. On va un peu perturber le n-uplet $(x_1, \ldots, x_n)$. Posons $y_h = x_h + \varepsilon$, $y_k = x_k - \varepsilon$ et $y_i = x_i$ pour tous les autres indices i. Il est clair que pour $|\varepsilon|$ assez petit $(y_1, \ldots, y_n) \in \mathrm{K}$. On a

$$q_{\mathrm{A}}(y_1, \ldots, y_n) = \alpha(x_h + \varepsilon) + \beta(x_k - \varepsilon) + \gamma = q_{\mathrm{A}}(x_1, \ldots, x_n) + \varepsilon(\alpha - \beta).$$

On en déduit que $\alpha = \beta$ sans quoi le maximum ne serait pas atteint en q_{A}. On peut alors remplacer le point x de K par $x' = (x'_1, \ldots, x'_n)$ où $x'_h = x_h + x_k$ et $x'_k = 0$ les autres coordonnées étant inchangées. Le maximum de q_{A} est encore atteint en x'. En itérant cela on peut donc obtenir un point $(z_1, \ldots, z_n) \in \mathrm{K}$ en lequel q_{A} atteint son maximum et tel que $(i,j) \in \mathrm{A}$ pour tout couple (i,j) tel que $z_i z_j \neq 0$. Autrement dit, si on pose $\mathrm{I} = \{i \in [\![1,n]\!],\ z_i \neq 0\}$, on a $\mathrm{I} \times \mathrm{I} \setminus \mathrm{D} \subset \mathrm{A}$. On a alors

$$q_{\mathrm{A}}(z_1, \ldots, z_n) = \sum_{(i,j) \in \mathrm{I}^2, i \neq j} z_i z_j \leqslant 1 - \frac{1}{|\mathrm{I}|} \leqslant 1 - \frac{1}{q}$$

ce qui prouve le résultat. $\lhd$

Les liens entre les formes quadratiques et l'arithmétique sont très riches même si très peu abordés dans ce chapitre. Voici toutefois deux exercices qui s'inscrivent dans ce thème.

3.43. Forme quadratique et réseau

Soit $q : \mathbb{R}^2 \longrightarrow \mathbb{R}$ définie par $q(x,y) = ax^2 + 2bxy + cy^2$ une forme quadratique définie positive. On pose $e = \inf_{(x,y) \in \mathbb{Z}^2 \setminus \{(0,0)\}} q(x,y)$.

1. Montrer que e est atteint.

2. Montrer qu'il existe $u, v \in \mathbb{Z}^2$ tels que $\mathbb{Z}^2 = \mathbb{Z}u + \mathbb{Z}v$ et $q(u) = e$.

3. Prouver que $3e^2 \leqslant 4(ac - b^2)$. Peut-il y avoir égalité ?

(**École normale supérieure**)

▷ **Solution.**

1. On a, pour tout $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $q(x,y) > 0$. On en déduit que $\inf_{\|(x,y)\|=1} q(x,y) = m > 0$, car ce minimum est atteint, puisque le cercle unité de $\mathbb{R}^2$ est compact. Par homogénéité, on obtient, pour tout $(x,y) \in \mathbb{R}^2$, $q(x,y) \geqslant m(x^2+y^2)$. Soit N un entier naturel tel que $\mathrm{N} \geqslant \sqrt{\dfrac{c+1}{m}}$. D'après ce qui précède, on a, si $(x,y) \in \mathbb{Z}^2$ et $|x| > \mathrm{N}$ ou $|y| > \mathrm{N}$, $q(x,y) \geqslant c+1$. On a donc $c = \inf_{\substack{(x,y)\in\mathbb{Z}^2\setminus\{(0,0)\} \\ |x|\leqslant \mathrm{N}, |y|\leqslant \mathrm{N}}} q(x,y)$. Cette borne inférieure est atteinte car l'ensemble est fini. On en déduit que $c > 0$.

2. Soit $u = (\alpha,\beta) \in \mathbb{Z}^2$ tel que $q(u) = c$. Notons $d = \mathrm{pgcd}(\alpha,\beta)$ et α', β' les entiers tels que $\alpha = d\alpha'$ et $\beta = d\beta'$. On a

$$c = q(u) = d^2 q(\alpha',\beta') \geqslant d^2 c$$

et donc $d^2 \leqslant 1$, c'est-à-dire $d = 1$.

Ainsi α et β sont premiers entre eux et d'après le théorème de Bezout, il existe $(\gamma,\delta) \in \mathbb{Z}^2$ tel que $\alpha\gamma + \beta\delta = 1$. Posons $v = (-\gamma,\delta)$ et montrons que (u,v) convient. On a évidemment $\mathbb{Z}u + \mathbb{Z}v \subset \mathbb{Z}^2$. Soit $\mathrm{P} = \begin{pmatrix} \alpha & -\delta \\ \beta & \gamma \end{pmatrix} \in \mathcal{M}_2(\mathbb{Z})$. On a $\det \mathrm{P} = 1$, donc P est inversible et $\mathrm{P}^{-1} \in \mathcal{M}_2(\mathbb{Z})$. La matrice P^{-1} est la matrice de passage de (u,v) à la base canonique (e_1,e_2) de $\mathbb{R}^2$. Comme elle est à coefficients entiers, e_1 et e_2 appartiennent à $\mathbb{Z}u + \mathbb{Z}v$ et finalement $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}^2$.

3. Il existe $(b',c') \in \mathbb{R}^2$ tel que l'expression de q dans la base (u,v) est $cx^2 + 2b'xy + c'y^2$, puisque $q(u) = c$. On a $\begin{pmatrix} c & b' \\ b' & c' \end{pmatrix} = {}^t\mathrm{P} \begin{pmatrix} a & b \\ b & c \end{pmatrix} \mathrm{P}$ et donc $c'c - b'^2 = (\det \mathrm{P})^2 (ac - b^2) = (ac - b^2)$.

Pour tout $(x,y) \in \mathbb{Z}^2$, on obtient

$$q(xu+yv) = cx^2 + 2b'xy + c'y^2 = c\left(x + \frac{b'}{c}y\right)^2 + \left(c' - \frac{b'^2}{c}\right)y.$$

En prenant $y = 1$ et en choisissant $n \in \mathbb{Z}$ tel que $\left|n + \dfrac{b'}{c}\right| \leqslant \dfrac{1}{2}$, on a

$$c \leqslant q(nu+v) \leqslant \frac{1}{4}c + c' - \frac{b'^2}{c}$$

et donc $3c^2 \leqslant 4(c'c - b'^2)$. Comme $c'c - b'^2 = ac - b^2$, on a bien

$$\boxed{3c^2 \leqslant 4(ac - b^2)}.$$

D'après ce qui précède, il y a égalité s'il existe $n \in \mathbb{Z}$ tel que $n + \dfrac{b'}{e} = \dfrac{1}{2}$ et $3e^2 = 4(c'e - b'^2)$, ce qui donne $2b' = (2n+1)e$ et $c' = \dfrac{1}{4e}(3e^2 + b'^2) = e(n^2+n+1)$.

Réciproquement, si (u, v) est une base de $\mathbb{R}^2$ telle que $\mathbb{Z}u + \mathbb{Z}v = \mathbb{Z}^2$, si $e > 0$, $n \in \mathbb{Z}$ et si q est la forme quadratique définie par

$$q(xu + yv) = ex^2 + (2n+1)exy + (n^2+n+1)ey^2,$$

on écrit $q(xu+yv) = e\left(x + \dfrac{2n+1}{2}y\right)^2 + \dfrac{3e}{4}y^2$.

Soit $(x, y) \in \mathbb{Z}^2 \setminus \{(0,0)\}$. Si y est pair, on $q(xu+yv) \geqslant e$, avec égalité si $x = 1$ et $y = 0$ et si y est impair, on a $q(xu+yv) \geqslant \dfrac{e}{4} + \dfrac{3e}{4} \geqslant e$. Le minimum de q sur $\mathbb{Z}^2 \setminus \{(0,0)\}$ est e et si l'expression de q dans la base canonique de $\mathbb{R}^2$ est $q(x,y) = ax^2 + 2bxy + cy^2$, on a

$$4(ac - b^2) = 4\left((n^2+n+1)e^2 - \frac{(2n+1)^2}{4}\right) = 3e^2. \quad \lhd$$

L'exercice suivant a été traité dans un cas particulier dans le tome 1 d'algèbre (exercice 4.36 dans la première édition et 4.37 dans la seconde édition) à savoir le cas où $n = 3$ et $A = I_3$. La démonstration est essentiellement la même.

3.44. Lemme de Davenport-Cassels

Soit $A \in \mathcal{M}_n(\mathbb{Z})$ une matrice symétrique définie positive et q_A la forme quadratique associée de $\mathbb{R}^n$. On suppose que, pour tout $x \in \mathbb{Q}^n$, il existe $y \in \mathbb{Z}^n$ tel que $q_A(x-y) < 1$.

Montrer que $\mathbb{N}^* \cap q_A(\mathbb{Q}^n) = \mathbb{N}^* \cap q_A(\mathbb{Z}^n)$.

(**École normale supérieure**)

$\rhd$ **Solution.**

On démontre que $\mathbb{N}^* \cap q_A(\mathbb{Q}^n) \subset \mathbb{N}^* \cap q_A(\mathbb{Z}^n)$, l'autre inclusion étant évidente. On note φ la forme polaire de q_A. Soit α dans $\mathbb{N}^* \cap q_A(\mathbb{Q}^n)$ et $x \in \mathbb{Q}^n$ tel que $q_A(x) = \alpha$. On écrit $x = \dfrac{1}{d}z$, avec $z \in \mathbb{Z}^n$ et d entier $d \geqslant 1$. On choisit x et z pour que d soit minimal. Par hypothèse, il existe $y \in \mathbb{Z}^n$ tel que $q_A(x-y) < 1$.

Si $q_A(x-y) = 0$ alors $x = y \in \mathbb{Z}^n$ et $\alpha \in q_A(\mathbb{Z}^n)$. Supposons donc $q_A(x-y) > 0$ et montrons qu'on aboutit à une contradiction.

On a $q_A(x-y) = q_A(x) + q_A(y) - 2\varphi(x,y)$ et $dq_A(x-y)$ est entier car

$q_A(x)$ est entier par hypothèse, $q_A(y)$ l'est également car $y \in \mathbb{Z}^n$ et A est à coefficients entiers et enfin $d\varphi(x,y) = \varphi(dx,y) = \varphi(z,y)$ est entier car z et y sont dans $\mathbb{Z}^n$. Ainsi, il existe un entier a tel que $q_A(x-y) = \dfrac{a}{d}$ et $0 < a < d$.

On pose $u = y + t(x-y)$ et on cherche le réel t tel que $q_A(u) = q_A(x)$. Comme $q_A(u) = q_A(y) + 2t\varphi(y, x-y) + t^2 q_A(x-y)$, il faut

$$t^2 q_A(x-y) + 2t\varphi(y, x-y) + q_A(y) - q_A(x) = 0.$$

Une solution évidente de cette équation du second degré est 1, cas où $u = x$. L'autre est donc $\dfrac{q_A(y) - q_A(x)}{q_A(x-y)}$. On obtient

$$\begin{aligned} u &= y + \frac{q_A(y) - q_A(x)}{q_A(x-y)}(x-y) = y + \frac{q_A(y) - q_A(x)}{\frac{a}{d}}\left(\frac{1}{d}z - y\right) \\ &= y + \frac{q_A(y) - q_A(x)}{a}(z - dy). \end{aligned}$$

Ainsi, puisque $q_A(x)$ et $q_A(y)$ sont entiers, u s'écrit $\dfrac{1}{a}v$, où $v \in \mathbb{Z}^n$. Comme $a = q(u) = q(x)$ et $0 < a < d$, cela contredit le choix de x. ◁

3.45. Image d'un cône

Soit q une forme quadratique sur $\mathbb{R}^n$ de matrice A dans la base canonique. On pose $C = (\mathbb{R}_+)^n$ et on note ∂C la frontière de C. On suppose que $q(\partial C) \subset \mathbb{R}_+$ et que tout vecteur propre de A qui est dans C est associé à une valeur propre positive. Montrer que $q(C) \subset \mathbb{R}_+$.

(École normale supérieure)

▷ **Solution.**

Posons $\Gamma = C \cap S$, où $S = \{x \in \mathbb{R}^n,\ \|x\| = 1\}$. Démontrer que q est positive sur C revient à démontrer que q est positive sur Γ. L'ensemble Γ est compact et comme q est continue, elle possède un minimum m sur Γ qui est atteint en un point u.

Si $x_0 \in \partial C$, alors $m = q(x_0) \geqslant 0$ par hypothèse. Supposons désormais que x_0 appartient à $(\mathbb{R}_+^*)^n$. On a alors

$$m = \min_{(\mathbb{R}_+^*)^n \cap S} q(x) = \min_{(\mathbb{R}_+^*)^n} \frac{q(x)}{\|x\|^2}.$$

La fonction f définie par $f(x) = \dfrac{q(x)}{\|x\|^2}$ est différentiable sur l'ouvert

$(\mathbb{R}_+^*)^n$. Au point u où elle atteint son minimum, sa différentielle est donc nulle. On a pour $1 \leqslant i \leqslant n$,

$$\begin{aligned}\frac{\partial f}{\partial x_i}(x) &= \frac{1}{\|x\|^4}\left(\|x\|^2 \frac{\partial q}{\partial x_i}(x) - 2x_i q(x)\right) \\ &= \frac{1}{\|x\|^2}\left(2\sum_{j=1}^{n} a_{ij}x_j - \frac{2x_i q(x)}{\|x\|^2}\right).\end{aligned}$$

Si $u = (u_1, \ldots, u_n)$, la condition $\dfrac{\partial f}{\partial x_i}(u) = 0$ donne $\displaystyle\sum_{j=1}^{n} a_{ij}u_j = \frac{q(u)}{\|u\|^2}u_i$ pour tout $i \in [\![1, n]\!]$. On obtient $Au = \dfrac{q(u)}{\|u\|^2}u$. Ainsi u est un vecteur propre de A qui est dans C. Par hypothèse, la valeur propre correspondante $\dfrac{q(u)}{\|u\|^2}$ est positive. On en déduit $m = q(u) \geqslant 0$. ◁

Le raisonnement précédent montre qu'étant donné une forme quadratique quelconque sur $\mathbb{R}^n$ de matrice A, *tout point de $\mathbb{R}^n \setminus \{0\}$ où la fonction $x \longmapsto \dfrac{q(x)}{\|x\|}$ atteint son minimum est vecteur propre de q et donc de* A. *Un tel minimum existe toujours car c'est aussi le minimum de la fonction sur la sphère unité. C'est une méthode possible pour montrer que toute matrice symétrique réelle possède une valeur propre (réelle), première étape dans la démonstration du théorème spectral.*

Chapitre 4

Géométrie affine et euclidienne

Nous allons commencer ce chapitre par quelques exercices de géométrie affine réelle plane. Les espaces affines ayant disparu des programmes des classes préparatoires lors de la dernière réforme, les énoncés sont désormais donnés dans le cadre des espaces vectoriels (un espace vectoriel E *est naturellement muni d'une structure d'espace affine de direction* E *lui-même), voire directement dans* $\mathbb{R}^n$*. Cette restriction a certes l'inconvénient de créer une légère confusion entre les points et les vecteurs mais n'apporte pas vraiment de perte de généralité puisque tout espace affine s'identifie canoniquement à un hyperplan affine d'un espace vectoriel*[1]*. Nous avons choisi de modifier les énoncés les plus anciens pour respecter ce nouveau programme. Nos lecteurs ayant étudié les espaces affines n'auront aucun mal à les reformuler éventuellement dans un contexte plus général.*

4.1. Partage équitable

On considère 1000 points dans le plan $\mathbb{R}^2$. Montrer qu'il existe une droite ayant (au sens strict) 500 points d'un côté et 500 points de l'autre.

(École polytechnique)

▷ **Solution.**

Imaginons que l'on fasse glisser une droite sur le plan, en venant de l'infini et en gardant toujours la même direction. Lorsque la droite traverse le nuage de points, ceux-ci vont passer d'un côté de la droite à l'autre. Si la direction de la droite est bien choisie, il n'y a qu'un seul point à la fois qui traverse la droite et il suffit de s'arrêter lorsque l'on a passé le 500-ième point !

On peut formaliser cela un peu plus. On munit le plan $\mathbb{R}^2$ de sa structure euclidienne et on note $(A_i)_{1\leqslant i\leqslant 1000}$ les points. Les vecteurs unitaires $\dfrac{\overrightarrow{A_iA_j}}{\|\overrightarrow{A_iA_j}\|}$, $i \neq j$, sont en nombre fini et le cercle unité du plan est infini. On choisit un vecteur unitaire u qui n'est pas dans l'ensemble précédent

1. Le lecteur intéressé lira par exemple le chapitre 3 de l'ouvrage *Géométrie 1* de MARCEL BERGER, Nathan. 1990.

et on considère un repère $\mathrm{R} = (\mathrm{O}, u, v)$. Les ordonnées y_i des points A_i dans ce repère sont deux à deux distinctes. On peut renuméroter les points pour avoir $y_1 < y_2 < \cdots < y_{1000}$. Il suffit de choisir un réel α entre y_{500} et y_{501} et de prendre la droite d'équation $y = \alpha$ dans le repère R. ◁

On regroupe dans l'exercice suivant deux versions affines du théorème de Pappus qui ont été posées séparément. Signalons que Pappus était un mathématicien grec qui a vécu au IV^{e} *siècle.*

4.2. Théorème de Pappus

Soient D et D′ deux droites affines sécantes d'un plan vectoriel réel, des points distincts A, B et C de D\D′, et des points distincts A′, B′, C′ de D′\D.

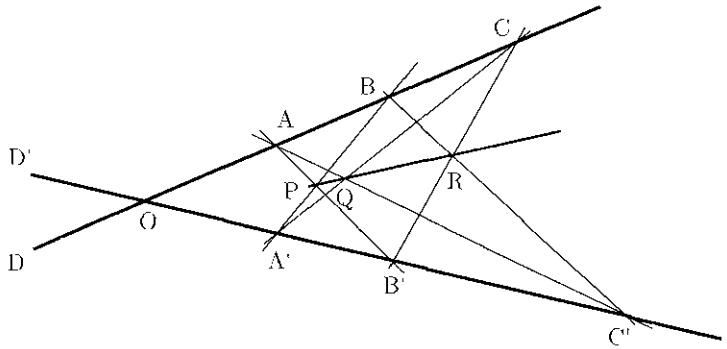
1. On suppose que (AB′) et (A′B) se coupent en P, (BC′) et (B′C) en Q et (CA′) et (C′A) en R. Montrer que P, Q, R sont alignés.

2. On suppose maintenant que (AB′) est parallèle à (A′B) et que (BC′) est parallèle à (B′C). Montrer que (AC′) est parallèle à (A′C).

(École polytechnique)

▷ **Solution.**

1. Prenons comme repère $\mathrm{R} = (\mathrm{O}, i, j)$ où O est le point d'intersection de D et D′, où i dirige D et j dirige D′. On pourrait prendre $i = \overrightarrow{\mathrm{OA}}$ mais cela ferait jouer un rôle particulier au point A et cette perte de symétrie impliquerait finalement plus de calculs.



Posons $\mathrm{A} = (a, 0)$, $\mathrm{B} = (b, 0)$, $\mathrm{C} = (c, 0)$ et de même $\mathrm{A}' = (0, a')$, $\mathrm{B}' = (0, b')$ et $\mathrm{C}' = (0, c')$ dans le repère choisi. On écrit l'équation de la

droite (AB′). Un point $M = (x, y)$ est sur cette droite si, et seulement si, $\det_{(i,j)}(\overrightarrow{AB'}, \overrightarrow{AM}) = 0$, ce qui donne $b'x + ay = ab'$. Pour trouver l'équation de la droite (A′B) il n'y a pas de calcul à faire : il suffit d'échanger les lettres a et b pour obtenir $a'x + by = a'b$. On résout le petit système obtenu, par exemple avec les formules de Cramer, pour obtenir les coordonnées du point P :

$$P = \left(\frac{ab(b' - a')}{bb' - aa'}, \frac{a'b'(b - a)}{bb' - aa'} \right).$$

La non nullité de $bb' - aa'$ est précisément la condition pour que les deux droites (AB′) et (A′B) se coupent. Il n'y a pas de calcul à faire pour obtenir les points Q et R : il suffit d'effectuer la permutation circulaire (a, b, c) deux fois de suite. On a donc

$$Q = \left(\frac{bc(c' - b')}{cc' - bb'}, \frac{b'c'(c - b)}{cc' - bb'} \right) \text{ et } R = \left(\frac{ac(a' - c')}{aa' - cc'}, \frac{a'c'(a - c)}{aa' - cc'} \right).$$

Rappelons que trois points $M_i = (x_i, y_i)$ $(i = 1, 2, 3)$ du plan sont alignés si, et seulement si,

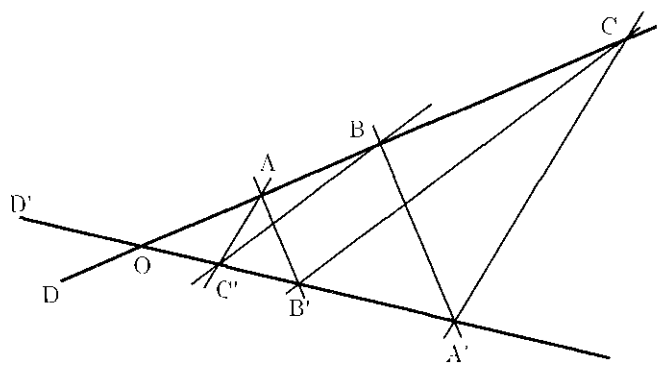
$$\begin{vmatrix} 1 & 1 & 1 \\ x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \end{vmatrix} = 0.$$

En sortant les dénominateurs par multilinéarité, on est donc amené à prouver que

$$\begin{vmatrix} bb' - aa' & cc' - bb' & aa' - cc' \\ ab(b' - a') & bc(c' - b') & ac(a' - c') \\ a'b'(b - a) & b'c'(c - b) & a'c'(a - c) \end{vmatrix} = 0.$$

C'est le cas puisque l'on a $cc'C_1 + aa'C_2 + bb'C_3 = 0$, où C_1, C_2, C_3 sont les trois colonnes de la matrice.

2. Voici la figure dans ce cas :



On garde les mêmes notations. Comme on l'a vu, la condition de parallélisme de (AB′) et (A′B) est $aa' = bb'$. On a de même $bb' = cc'$ car (BC′) est parallèle à (B′C). Par transitivité, il vient $aa' = cc'$ ce qui signifie exactement que (AC′) est parallèle à (A′C). ◁

La distinction des deux cas s'explique par le fait que le théorème de Pappus est un résultat de géométrie projective : sans rentrer dans les détails, le cas où les trois paires de droites sont parallèles est celui où elles se coupent sur la «droite à l'infini» et les points d'intersection sont encore alignés. Le lecteur pourra montrer que le résultat de l'exercice reste encore vrai dans le cas où les droites D *et* D′ *sont parallèles.*

Dans certaines questions, l'utilisation des coordonnées barycentriques peut se révéler plus commode que le recours aux coordonnées cartésiennes. Rappelons que si (A, B, C) *est un triangle d'un plan vectoriel (disons réel), on appelle coordonnées barycentriques d'un point* M *de ce plan dans la base affine* (A, B, C) *l'unique triplet* (a, b, c) *tel que* $a+b+c = 1$ *et tel que* M *soit le barycentre de* (A, a), (B, b) *et* (C, c). *Il est bien entendu très facile de passer des coordonnées barycentriques aux coordonnées cartésiennes par exemple dans le repère* $(A, \overrightarrow{AB}, \overrightarrow{AC})$. *Mais l'intérêt des premières est qu'elles font jouer un rôle parfaitement symétrique aux trois points* A, B, C.

La première question de l'énoncé suivant donne une condition d'alignement de trois points donnés en coordonnées barycentriques. La suite est consacrée au classique théorème de Ménélaüs.

4.3. Théorème de Ménélaüs

Soit P un plan vectoriel et A, B, C trois points non alignés de ce plan.

1. Soit M, M′ et M″ trois points de coordonnées barycentriques respectives (α, β, γ), $(\alpha', \beta', \gamma')$ et $(\alpha'', \beta'', \gamma'')$ dans la base affine (A, B, C) ; exprimer à l'aide d'un déterminant la condition pour que ces trois points soient alignés.

2. On suppose maintenant que P, Q, R sont trois points sur les côtés respectifs du triangle ABC ; trouver une condition pour que P, Q et R soient alignés.

(École polytechnique)

▷ **Solution.**

1. On revient simplement aux coordonnées cartésiennes dans le repère $(A, \overrightarrow{AB}, \overrightarrow{AC})$. On a $\overrightarrow{AM} = \alpha\overrightarrow{AB} + \beta\overrightarrow{AC}$ et $\overrightarrow{AM'} = \alpha'\overrightarrow{AB} + \beta'\overrightarrow{AC}$. On en déduit que

$$\overrightarrow{MM'} = (\alpha' - \alpha)\overrightarrow{AB} + (\beta' - \beta)\overrightarrow{AC}$$

et de même

$$\overrightarrow{MM''} = (\alpha'' - \alpha)\overrightarrow{AB} + (\beta'' - \beta)\overrightarrow{AC}.$$

Ainsi M,M', M'' sont alignés si, et seulement si, $\det_{(\overrightarrow{AB},\overrightarrow{AC})}(\overrightarrow{MM'}, \overrightarrow{MM''}) = 0$ soit $\begin{vmatrix} \alpha' - \alpha & \beta' - \beta \\ \alpha'' - \alpha & \beta'' - \beta \end{vmatrix} = 0$. On peut mettre ce déterminant sous une forme plus symétrique en remarquant qu'il vaut

$$\begin{vmatrix} \alpha & \beta & 1 \\ \alpha' & \beta' & 1 \\ \alpha'' & \beta'' & 1 \end{vmatrix} = \begin{vmatrix} \alpha & \beta & \alpha + \beta + \gamma \\ \alpha' & \beta' & \alpha' + \beta' + \gamma' \\ \alpha'' & \beta'' & \alpha'' + \beta'' + \gamma'' \end{vmatrix} - \begin{vmatrix} \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \\ \alpha'' & \beta'' & \gamma'' \end{vmatrix}.$$

La condition cherchée est donc $\begin{vmatrix} \alpha & \beta & \gamma \\ \alpha' & \beta' & \gamma' \\ \alpha'' & \beta'' & \gamma'' \end{vmatrix} = 0$.

2. Supposons que les points P,Q et R appartiennent respectivement à (BC), (CA) et (AB). Leurs coordonnées barycentriques sont respectivement de la forme $(0, \beta, \gamma)$, $(\alpha', 0, \gamma')$ et $(\alpha'', \beta'', 0)$. D'après la question précédente, les trois points sont alignés si, et seulement si,

$$\begin{vmatrix} 0 & \beta & \gamma \\ \alpha' & 0 & \gamma' \\ \alpha'' & \beta'' & 0 \end{vmatrix} = \alpha'\beta''\gamma + \alpha''\beta\gamma' = 0.$$

Si les points P, Q, R sont distincts des sommets du triangle, cette condition s'écrit $\dfrac{\gamma}{\beta}\dfrac{\alpha'}{\gamma'}\dfrac{\beta''}{\alpha''} = -1$, soit encore

$$\frac{\overline{PB}}{\overline{PC}}\frac{\overline{QC}}{\overline{QA}}\frac{\overline{RA}}{\overline{RB}} = -1.$$

C'est cette équivalence qui constitue le théorème de Ménélaüs. ◁

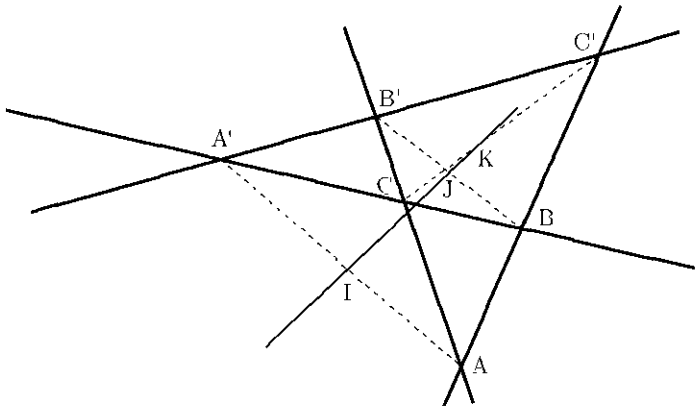
4.4. Quadrilatère complet

On se donne quatre droites affines d'un plan vectoriel en position générale. On appelle diagonale un segment joignant un point d'intersection de deux des droites avec le point d'intersection des deux autres droites. Montrer que les milieux des diagonales sont alignés.

(École polytechnique)

▷ **Solution.**

Notons [AA′], [BB′] et [CC′] les trois diagonales comme sur la figure suivante.



Dire que les droites sont en position générale signifie que tous les points d'intersection existent et sont deux à deux distincts. On va utiliser les coordonnées barycentriques dans la base affine (A, B, C). Écrivons les coordonnées de tous les points : A = (1, 0, 0), B = (0, 1, 0), C = (0, 0, 1) et A′ = (0, α, 1 − α), B′ = (1 − β, 0, β), et C′ = (γ, 1 − γ, 0). Comme les points A′, B′, C′ sont alignés, α, β, γ sont liés par la relation ci-après (voir la question 1 de l'exercice précédent) :

$$\begin{vmatrix} 0 & \alpha & 1-\alpha \\ 1-\beta & 0 & \beta \\ \gamma & 1-\gamma & 0 \end{vmatrix} = 0.$$

Les milieux I, J, K des diagonales [AA′], [BB′] et [CC′] ont pour coordonnées barycentriques respectives $\mathrm{I} = \left(\frac{1}{2}, \frac{\alpha}{2}, \frac{1-\alpha}{2}\right)$, $\mathrm{J} = \left(\frac{1-\beta}{2}, \frac{1}{2}, \frac{\beta}{2}\right)$ et $\mathrm{K} = \left(\frac{\gamma}{2}, \frac{1-\gamma}{2}, \frac{1}{2}\right)$ L'alignement des trois points équivaut donc à la nullité du déterminant

$$\begin{vmatrix} 1 & \alpha & 1-\alpha \\ 1-\beta & 1 & \beta \\ \gamma & 1-\gamma & 1 \end{vmatrix}.$$

Cela revient à dire que −1 est une valeur propre de la matrice

$$\mathrm{M} = \begin{pmatrix} 0 & \alpha & 1-\alpha \\ 1-\beta & 0 & \beta \\ \gamma & 1-\gamma & 0 \end{pmatrix}.$$

Or, nous savons que 0 est valeur propre de M et il est visible que 1 est aussi valeur propre (avec pour vecteur propre $(1,1,1)$). Comme la trace de M est nulle, -1 est aussi valeur propre et le résultat en découle. ◁

Le problème suivant, posé en 1893 par Sylvester, est étonnant : la formulation de la question est purement affine et pourtant sa résolution nécessite des arguments euclidiens (le résultat n'est pas vrai sur un corps quelconque).

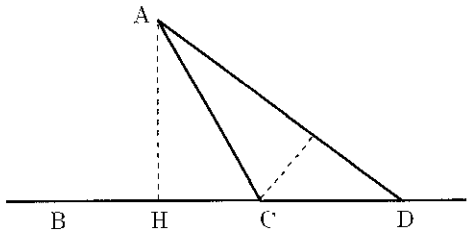
4.5. Théorème de Sylvester-Gallai

On considère n points du plan $\mathbb{R}^2$ qui ont la propriété suivante : toute droite passant par deux des points en contient un troisième. Montrer que tous les points sont alignés.

(École polytechnique)

▷ **Solution.**

Une autre formulation du problème est la suivante : si les points ne sont pas tous alignés, il existe une droite du plan qui contient exactement deux des points. C'est sous cette forme que nous allons le résoudre. Notons X notre ensemble de points et $\mathcal{D}$ l'ensemble (fini) de toutes les droites passant par deux points de X. Parmi tous les couples (A, Δ) vérifiant $A \in X$, $\Delta \in \mathcal{D}$ et $A \notin \Delta$ choisissons-en un tel que la distance de A à Δ soit minimale. Nous allons prouver que la droite Δ convient. Par hypothèse, il y a au moins deux points B, C de X qui sont sur Δ. Supposons par l'absurde qu'il y ait un troisième point D de X sur Δ. Quitte à permuter nos trois points, on peut très bien supposer que B, C, D sont dans cet ordre. Notons H le projeté orthogonal de A sur Δ. Deux des trois points B, C, D au moins sont sur la même demi-droite définie par H. Supposons que ce soient C et D. Voici la figure :



La contradiction apparaît bien sur la figure : la distance du point C à la droite $(AD) \in \mathcal{D}$ est visiblement strictement inférieure à la distance

AH (elle est inférieure à la distance de H à (AD) elle-même inférieure à AH). ◁

Il n'est pas certain que Sylvester ait disposé d'une solution à sa question. La première preuve publiée (près de 40 ans après la question) est due à Gallai, ce qui explique le nom du théorème. L'idée de la solution donnée ci-dessus est due à L.M. Kelly.

Les nombreux énoncés qui vont suivre concernent la géométrie affine euclidienne. Nous les avons classés selon plusieurs thèmes : géométrie du triangle, utilisation des nombres complexes, cercles, coniques euclidiennes, problèmes d'extrema.

4.6. Aire d'un triangle

Soit trois droites non parallèles deux à deux du plan euclidien $\mathbb{R}^2$ définies par leur équation dans le repère orthonormé canonique

$$ax + by + c = 0, \quad a'x + b'y + c' = 0, \quad a''x + b''y + c'' = 0.$$

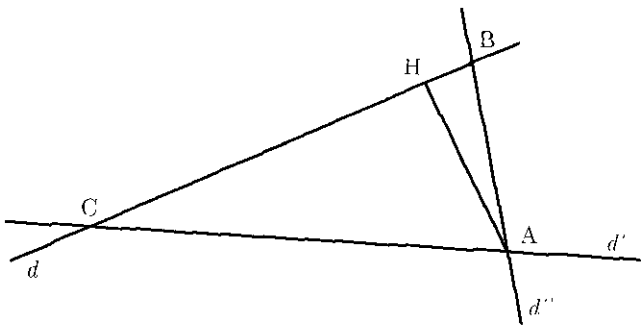
Soit S l'aire du triangle formé par ces trois droites. Montrer que

$$2S = \frac{\begin{vmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{vmatrix}^2}{|(ab' - a'b)(ab'' - a''b)(a'b'' - a''b')|}.$$

(École polytechnique)

▷ **Solution.**

On note d, d' et d'' les trois droites, $\{A\} = d' \cap d''$, $\{B\} = d'' \cap d$ et $\{C\} = d \cap d'$ et enfin Δ le déterminant $\begin{vmatrix} a & b & c \\ a' & b' & c' \\ a'' & b'' & c'' \end{vmatrix}$.



En résolvant le système constitué des équations de d' et d'', on trouve les coordonnées de A,

$$x_{\mathrm{A}} = \frac{b'c'' - b''c'}{a'b'' - a''b'}, \quad y_{\mathrm{A}} = \frac{a''c' - a'c''}{a'b'' - a''b'}.$$

On en déduit la distance de A à la droite d,

$$\mathrm{AH} = \frac{|ax_{\mathrm{A}} + by_{\mathrm{A}} + c|}{\sqrt{a^2+b^2}} = \frac{|a(b'c''-b''c') + b(a''c'-a'c'') + c(a'b''-a''b')|}{\sqrt{a^2+b^2}\,|a'b'' - a''b'|},$$

où H est le projeté orthogonal de A sur (BC). On identifie au numérateur le développement du déterminant Δ selon sa première ligne. Ainsi,

$$\mathrm{AH} = \frac{|\Delta|}{\sqrt{a^2+b^2}\,|a'b'' - a''b'|}.$$

On va maintenant calculer la distance BC. Le vecteur $(b, -a)$ est un vecteur directeur de la droite d donc il existe $t \in \mathbb{R}$ tel que $\overrightarrow{\mathrm{BC}} = (bt, -ta)$ et donc $x_{\mathrm{C}} = x_{\mathrm{B}} + bt$, $y_{\mathrm{C}} = y_{\mathrm{B}} - ta$. Le point C appartient à d' donc on a

$$a'x_{\mathrm{C}} + b'y_{\mathrm{C}} + c' = a'x_{\mathrm{B}} + b'y_{\mathrm{B}} + c' + t(-ab' + a'b) = 0.$$

On en déduit que $t = \dfrac{a'x_{\mathrm{B}} + b'y_{\mathrm{B}} + c'}{ab' - a'b}$ puis que

$$\mathrm{BC} = |t|\,\sqrt{a^2+b^2} = \frac{|a'x_{\mathrm{B}} + b'y_{\mathrm{B}} + c'|\sqrt{a^2+b^2}}{|ab' - a'b|}.$$

Or, le même calcul que celui qui a été effectué précédemment pour le point A montre que $|a'x_{\mathrm{B}} + b'y_{\mathrm{B}} + c'| = \dfrac{|\Delta|}{|ab'' - a''b|}$. On obtient donc finalement

$$2\mathrm{S} = \mathrm{AH}\cdot\mathrm{BC} = \frac{\Delta^2}{|ab'' - a''b||ab' - a'b||a'b'' - a''b'|}. \quad \lhd$$

Ce n'est pas une formule très commode. Dans le même genre, le lecteur pourra démontrer la formule de Héron : si a, b, c sont les trois côtés d'un triangle alors son aire est égale à $\sqrt{p(p-a)(p-b)(p-c)}$ où $p = \dfrac{a+b+c}{2}$ est le demi-périmètre du triangle.

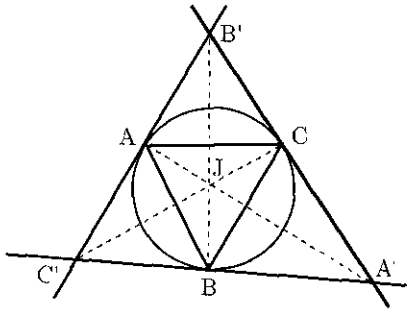
4.7. Point de Gergonne

Soit ABC un triangle de cercle circonscrit Γ. Les tangentes à Γ en A et B se coupent en C'. On définit de même A' et B'. Montrer que les droites (AA'), (BB') et (CC') sont concourantes.

(École polytechnique)

▷ **Solution.**

Voici pour commencer la figure :



On va démontrer que les trois droites sont concourantes en un point J que l'on va exprimer comme barycentre de (A', B', C'). On pose

$$\alpha = A'B = A'C, \quad \beta = B'C = B'A \quad \text{et} \quad \gamma = C'A = C'B.$$

Ces trois nombres sont strictement positifs. Le point A étant à l'intérieur du segment $[B'C']$, on a

$$\frac{\overline{AB'}}{\overline{AC'}} = -\frac{AB'}{AC'} = -\frac{\beta}{\gamma}$$

et donc $\frac{1}{\beta}\overrightarrow{AB'} + \frac{1}{\gamma}\overrightarrow{AC'} = \vec{0}$. Ainsi A est le barycentre des points pondérés $\left(B', \frac{1}{\beta}\right)$ et $\left(C', \frac{1}{\gamma}\right)$. On montre de même que B est le barycentre de $\left(A', \frac{1}{\alpha}\right)$ et $\left(C', \frac{1}{\gamma}\right)$ et C celui de $\left(A', \frac{1}{\alpha}\right)$ et $\left(B', \frac{1}{\beta}\right)$.

Considérons le barycentre J de $\left(\left(A', \frac{1}{\alpha}\right), \left(B', \frac{1}{\beta}\right), \left(C', \frac{1}{\gamma}\right)\right)$. Par associativité du barycentre, on trouve que G est le barycentre de $\left(\left(A', \frac{1}{\alpha}\right), \left(A, \frac{1}{\beta} + \frac{1}{\gamma}\right)\right)$, de $\left(\left(B', \frac{1}{\beta}\right), \left(B, \frac{1}{\alpha} + \frac{1}{\gamma}\right)\right)$ et de $\left(\left(C', \frac{1}{\gamma}\right), \left(C, \frac{1}{\alpha} + \frac{1}{\beta}\right)\right)$. On en déduit que J appartient aux droites (AA'), (BB') et (CC'). Ces trois droites sont donc concourantes en J. ◁

Le cercle Γ *est le cercle inscrit dans le triangle* A'B'C' *et les points* A, B, C *sont les points de contact entre les côtés du triangle et ce cercle inscrit. Le point de concours des droites* (AA'), (BB') *et* (CC') *est appelé le point de Gergonne du triangle* A'B'C'.

4.8. Quadrilatère formé par les sommets et l'orthocentre d'un triangle

Soit A, B, C, D quatre points du plan, trois à trois non alignés, d'isobarycentre G. Montrer l'équivalence entre

(*i*) un des quatre points A, B, C, D est l'orthocentre du triangle formé par les trois autres ;

(*ii*) G est à égale distance des 6 milieux de [AB], [AC], [AD], [BC], [BD], [CD].

(École polytechnique)

▷ **Solution.**

Notons I, J, K, L, M, N les milieux respectifs des côtés [AB], [AC], [AD], [BC], [BD], [CD].

• Remarquons pour commencer que si l'un des quatre points A, B, C, D est l'orthocentre du triangle formé par les trois autres, alors chacun des quatre points A, B, C, D est l'orthocentre du triangle formé par les trois autres. En effet si par exemple A est l'orthocentre du triangle BCD, on a

$$\overrightarrow{AB} \cdot \overrightarrow{CD} = \overrightarrow{AC} \cdot \overrightarrow{BD} = \overrightarrow{AD} \cdot \overrightarrow{BC} = 0.$$

En écrivant ces égalités sous la forme

$$\overrightarrow{BA} \cdot \overrightarrow{CD} = \overrightarrow{BC} \cdot \overrightarrow{AD} = \overrightarrow{BD} \cdot \overrightarrow{AC} = 0,$$

on voit que B est l'orthocentre du triangle BCD et il en est de même pour les deux autres points.

Ainsi (*i*) équivaut à : A est l'orthocentre du triangle BCD.

• Par associativité du barycentre, G est le milieu du segment [I, N] donc est équidistant de I et N. De même, il est équidistant de J et M et de K et L. On en déduit que G est équidistant des six points si, et seulement si, GI = GJ et GI = GK.

Cherchons à quelle condition, GI = GJ. On a

$$\overrightarrow{AG} = \frac{1}{4}(\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{AD}), \quad \overrightarrow{AI} = \frac{1}{2}\overrightarrow{AB}, \quad \overrightarrow{AJ} = \frac{1}{2}\overrightarrow{AC}.$$

On en déduit

$$\overrightarrow{IG} = \frac{1}{4}(-\overrightarrow{AB} + \overrightarrow{AC} + \overrightarrow{AD}) = \frac{1}{4}(\overrightarrow{BC} + \overrightarrow{AD})$$

$$\overrightarrow{JG} = \frac{1}{4}(\overrightarrow{AB} - \overrightarrow{AC} + \overrightarrow{AD}) = \frac{1}{4}(-\overrightarrow{BC} + \overrightarrow{AD}).$$

Cela montre que GI = GJ équivaut à $\overrightarrow{AD} \cdot \overrightarrow{BC} = 0$. On montre de même que GI = GK équivaut à $\overrightarrow{AC} \cdot \overrightarrow{BD} = 0$

Finalement (ii) équivaut à $\overrightarrow{AD} \cdot \overrightarrow{BC} = \overrightarrow{AC} \cdot \overrightarrow{BD} = 0$, c'est-à-dire à A orthocentre du triangle BCD, car si un point appartient à deux des hauteurs c'est l'orthocentre. Ainsi (ii) équivaut à (i). ◁

Le corps des nombres complexes permet d'algébriser la géométrie euclidienne plane. Les trois exercices suivants fournissent des exemples de cela. Le premier est très classique.

4.9. Caractérisation des triangles équilatéraux

Montrer qu'il existe un polynôme $P(X, Y, Z) \in \mathbb{R}[X, Y, Z]$ tel que, pour $(a, b, c) \in \mathbb{C}^3$, $P(a, b, c) = 0$ si, et seulement si, le triangle (a, b, c) est équilatéral.

(École polytechnique)

▷ **Solution.**

Le triangle (a, b, c) est équilatéral direct si, et seulement si, c est l'image de b par la rotation de centre a et d'angle $\frac{\pi}{3}$. L'expression analytique de cette rotation est $z \longmapsto e^{\frac{i\pi}{3}}(z - a) + a$ ou encore en posant classiquement $j = e^{\frac{2i\pi}{3}}$, $z \longmapsto -j^2(z-a)+a$. Donc (a, b, c) est équilatéral direct si, et seulement si, $c = -j^2(b - a) + a$ soit encore $c + ja + j^2 b = 0$. De même le triangle (a, b, c) est équilatéral indirect si, et seulement si, $c + j^2 a + jb = 0$. On peut conclure que le triangle est équilatéral si, et seulement si, la quantité

$$(c + ja + j^2 b)(c + j^2 a + jb) = a^2 + b^2 + c^2 - ab - ac - bc$$

est nulle. Le polynôme P recherché est donc

$$P(X, Y, Z) = X^2 + Y^2 + Z^2 - XY - XZ - YZ.$$

Il est symétrique ce qui est tout à fait normal. ◁

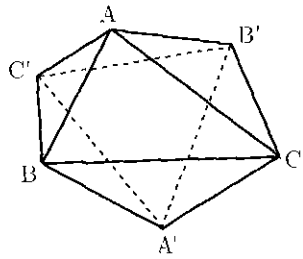
4.10. Un triangle équilatéral

Soit ABC un triangle du plan euclidien. On construit A′ tel que A et A′ soient de part et d'autre de la droite (BC), avec $\widehat{BA'C} = \frac{2\pi}{3}$ et A′B = A′C. On construit de manière analogue B′ et C′. Montrer que le triangle A′B′C′ est équilatéral.

(École polytechnique)

Solution.

On oriente le plan de façon à ce que le triangle ABC soit direct et on rapporte le plan à un repère orthonormal direct (O, u, v). On note a, b, c, a', b', c' les affixes des points A, B, C, A', B', C'.



Le triangle A'CB est alors aussi direct et B est l'image de C par la rotation de centre A' et d'angle $\dfrac{2\pi}{3}$, ce qui donne

$$b - a' = e^{\frac{2i\pi}{3}}(c - a') = j(c - a')$$

et donc $a' = \dfrac{b - cj}{1 - j}$. On obtient de même $b' = \dfrac{c - aj}{1 - j}$ et $c' = \dfrac{a - bj}{1 - j}$ en permutant les lettres. On en déduit que

$$b' - a' = \frac{c - aj - b + cj}{1 - j} = \frac{-aj - b - cj^2}{1 - j} \quad \text{et}$$
$$c' - a' = \frac{a - bj - b + cj}{1 - j} = \frac{a + bj^2 + cj}{1 - j}.$$

car $1 + j + j^2 = 0$. Ainsi,

$$c' - a' = -j^2(b' - a') = e^{i\frac{\pi}{3}}(b' - a').$$

Le point C' est donc l'image de B' dans la rotation de centre A' et d'angle $\dfrac{\pi}{3}$. Le triangle A'B'C' est équilatéral (direct). ◁

On peut démontrer de plus que les droites (AA'), (BB') *et* (CC') *sont concourantes en un point nommé point de Torricelli du triangle* ABC. *Ce point est étudié dans l'exercice 4.34.*

L'utilisation des nombres complexes est particulièrement efficace pour le résultat suivant (théorème de Ptolemée) lorsqu'on ne connaît pas les inversions. Il donne une condition métrique, et non pas angulaire, pour que quatre points du plan soient cocycliques.

4.11. Théorème de Ptolémée

Soient A, B, C, D quatre points distincts formant dans cet ordre un polygone convexe d'un plan euclidien. Montrer que ces points sont cocycliques si, et seulement si, AC.BD = AB.CD + AD.BC.

(École polytechnique)

▷ **Solution.**

On se place dans le plan complexe et on note a, b, c, d les affixes des points A, B, C, D. On a alors AC.BD $= |c-a||d-b|$, AB.CD $= |b-a||d-c|$ et AD.BC $= |d-a||c-b|$. La clé est de remarquer que

$$(c-a)(d-b) = (b-a)(d-c) + (d-a)(c-b).$$

Par inégalité triangulaire on a donc AC.BD $\leqslant$ AB.CD + AD.BC et il y a égalité si, et seulement si,

$$|(b-a)(d-c) + (d-a)(c-b)| = |(b-a)(d-c)| + |(d-a)(c-b)|.$$

Or deux nombres complexes non nuls z et z' vérifient $|z+z'| = |z|+|z'|$ si, et seulement si, ils ont le même argument.
Ainsi l'égalité AC.BD = AB.CD + AD.BC est vérifiée si, et seulement si, $\arg(b-a)(d-c) \equiv \arg(d-a)(c-b)\ [2\pi]$ soit encore

$$\arg\left(\frac{b-a}{d-a}\right) \equiv \arg\left(\frac{c-b}{d-c}\right) \equiv \arg\left(\frac{b-c}{d-c}\right) + \pi \quad [2\pi].$$

Cette condition se traduit en termes d'angles par

$$(\widehat{\overrightarrow{AB}, \overrightarrow{AD}}) \equiv (\widehat{\overrightarrow{CB}, \overrightarrow{CD}}) + \pi \quad [2\pi].$$

Comme les points A et C sont de part et d'autre du segment [BD] puisque le quadrilatère est supposé convexe, il s'agit exactement de la condition pour que les points A, B, C, D soient cocycliques. ◁

Les exercices suivants concernent des cercles sous divers aspects. Dans le premier, on rappelle la définition de la puissance d'un point par rapport à un cercle et la condition de cocyclicité de quatre points qui en résulte.

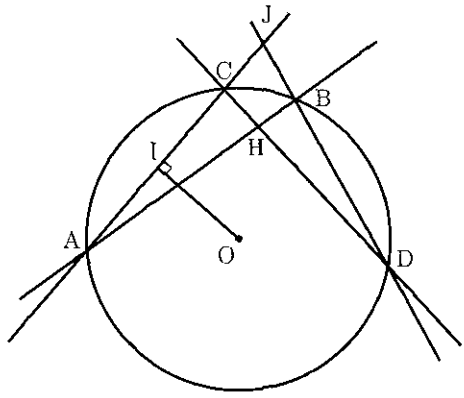
4.12. Puissance d'un point par rapport à un cercle

Soient A, B, C, D quatre points distincts d'un plan euclidien. On suppose que les droites (AC) et (BD) se coupent en J et que les droites (AB) et (CD) se coupent en H. Montrer que $\overrightarrow{JA}\cdot\overrightarrow{JC} = \overrightarrow{JB}\cdot\overrightarrow{JD}$ si et seulement si $\overrightarrow{HA}\cdot\overrightarrow{HB} = \overrightarrow{HC}\cdot\overrightarrow{HD}$.

(École polytechnique)

▷ **Solution.**

Montrons que la condition $\overrightarrow{JA} \cdot \overrightarrow{JC} = \overrightarrow{JB} \cdot \overrightarrow{JD}$ équivaut à A, B, C, D cocycliques.



• Si les quatre points sont cocycliques sur un cercle $\mathcal{C}$, les deux produits scalaires $\overrightarrow{JA} \cdot \overrightarrow{JC}$ et $\overrightarrow{JB} \cdot \overrightarrow{JD}$ sont égaux à la puissance du point J par rapport au cercle $\mathcal{C}$, c'est-à-dire à $JO^2 - R^2$, où O est le centre du cercle et R son rayon. En effet, en appelant I le milieu du segment [AC], on obtient

$$\overrightarrow{JA} \cdot \overrightarrow{JC} = (\overrightarrow{JI} + \overrightarrow{IA}) \cdot (\overrightarrow{JI} + \overrightarrow{IC}) = (\overrightarrow{JI} + \overrightarrow{IA}) \cdot (\overrightarrow{JI} - \overrightarrow{IA}) = JI^2 - IA^2.$$

La droite (OI) est la médiatrice du segment [AC], donc est orthogonale à (AC) et d'après le théorème de Pythagore,

$$\overrightarrow{JA} \cdot \overrightarrow{JC} = JO^2 - OI^2 - IA^2 = JO^2 - OA^2 = JO^2 - R^2.$$

On obtient de même $\overrightarrow{JB} \cdot \overrightarrow{JD} = JO^2 - R^2$ et on l'égalité voulue.

• Supposons réciproquement que $\overrightarrow{JA} \cdot \overrightarrow{JC} = \overrightarrow{JB} \cdot \overrightarrow{JD}$. Si J n'est égal ni à A, ni à C, les points A, B et C ne sont pas alignés. Soit $\mathcal{C}$ le cercle circonscrit au triangle ABC. La droite (JB) recoupe le cercle $\mathcal{C}$ en D′ tel que

$$\overrightarrow{JA} \cdot \overrightarrow{JC} = \overrightarrow{JB} \cdot \overrightarrow{JD} = \overrightarrow{JB} \cdot \overrightarrow{JD'}.$$

Comme les points J,B D et D′ sont alignés, on a $\overrightarrow{JB} \cdot \overrightarrow{JD} = \overrightarrow{JB} \cdot \overrightarrow{JD'}$ et $D = D'$ puisque $J \neq B$.

Si J = A ou C, on a alors $\overrightarrow{JB} \cdot \overrightarrow{JD} = \overrightarrow{JA} \cdot \overrightarrow{JC} = 0$, donc $\overrightarrow{JB} \cdot \overrightarrow{JD} = 0$. On en déduit que J = B ou J = D. Mais, deux des points sont égaux, car confondus avec J, ce qui est contraire à l'hypothèse.

L'équivalence annoncée est ainsi démontrée.

De la même façon, on a $\overrightarrow{HA} \cdot \overrightarrow{HB} = \overrightarrow{HC} \cdot \overrightarrow{HD}$ si et seulement si A, B, C et D sont cocycliques. Cela démontre l'équivalence demandée. ◁

4.13. Suite de cercles

Soit, pour tout n de $\mathbb{N}$, C_n un cercle de $\mathbb{R}^2$ euclidien, de rayon $r_n > 0$, et D_n le disque fermé de frontière C_n. On suppose que la suite (r_n) est croissante majorée et que les cercles C_n sont deux à deux disjoints. On suppose de plus que, pour tout n de $\mathbb{N}^*$, $(0,0) \in D_n$ et $(1,0) \notin D_n$.

1. Montrer que pour tout entier naturel n, il existe un unique x_n appartenant à $[0,1]$ tel que $(x_n, 0) \in C_n$.

2. Montrer que la suite (x_n) converge.

3. Montrer que la tangente à C_n en $(x_n, 0)$ converge vers une droite à préciser.

(École normale supérieure)

▷ **Solution.**

1. Le point $O = (0,0)$ est à l'intérieur du disque D_n et le point $A = (1,0)$ à l'extérieur du disque. Le segment $[O, A]$ coupe donc le cercle C_n en un unique point B_n. Les points du segment $[O, A]$ sont les points $(0, x)$ avec $x \in [0,1]$, donc il existe $x_n \in [0,1]$ tel que $B_n = (x_n, 0)$. Ce réel x_n est unique car le segment $[0, A]$ coupe C_n en un point unique.

2. Les cercles C_n étant disjoints, soit les disques D_n et D_{n+1} sont disjoints, soit l'un est inclus dans l'autre. Comme ils contiennent tous $(0,0)$, on est dans le deuxième cas. La suite (r_n) étant croissante, on a $r_n \leqslant r_{n+1}$ et $D_n \subset D_{n+1}$. On en déduit que $D_n \cap [0, A] \subset D_{n+1} \cap [0, A]$, *i.e.* $[0, B_n] \subset [0, B_{n+1}]$ et donc $x_n \leqslant x_{n+1}$. La suite (x_n) est croissante et majorée par 1 donc elle converge. On note ℓ sa limite, $\ell \in [0,1]$.

3. Pour tout entier naturel n, notons ω_n le centre du cercle C_n. La suite (r_n) est croissante et majorée donc converge. Soit n et p deux entiers tels que $n \geqslant p$. Comme $D_p \subset D_n$, on a $\|\omega_n - \omega_p\| \leqslant r_n - r_p$. La suite (ω_n) est donc de Cauchy et, comme $\mathbb{R}^2$ est complet, elle converge. On note ω sa limite. La tangente en $B_n = (x_n, 0)$ à C_n est la droite passant par B_n orthogonale à $\overrightarrow{\omega_n B_n}$. Elle converge vers la droite contenant $L = (\ell, 0)$ et orthogonale à $\overrightarrow{\omega L}$. ◁

L'exercice suivant rappelle l'étude classique des lignes de niveau d'une fonction scalaire de Leibniz.

4.14. Condition de cocyclicité de quatre points

Soit A, B, C et D quatre points non alignés d'un plan euclidien.

1. Montrer que A, B, C et D sont cocycliques si, et seulement si, il existe a, b, c, d réels non tous nuls tels que

$$a\mathrm{MA}^2 + b\mathrm{MB}^2 + c\mathrm{MC}^2 + d\mathrm{MD}^2 = 0$$

pour tout point M du plan.

2. Étudier, pour (a, b, c, d) quelconque, les lignes de niveau de l'application

$$\mathrm{M} \longmapsto a\mathrm{MA}^2 + b\mathrm{MB}^2 + c\mathrm{MC}^2 + d\mathrm{MD}^2.$$

(École polytechnique)

Solution.

1. • Supposons qu'il existe quatre réels non tous nuls tels que l'application $f : \mathrm{M} \longmapsto a\mathrm{MA}^2 + b\mathrm{MB}^2 + c\mathrm{MC}^2 + d\mathrm{MD}^2$ soit identiquement nulle. Pour tout point O du plan, on obtient

$$\begin{aligned} f(\mathrm{M}) &= a(\overrightarrow{\mathrm{MO}}+\overrightarrow{\mathrm{OA}})^2 + b(\overrightarrow{\mathrm{MO}}+\overrightarrow{\mathrm{OB}})^2 + c(\overrightarrow{\mathrm{MO}}+\overrightarrow{\mathrm{OC}})^2 + d(\overrightarrow{\mathrm{MO}}+\overrightarrow{\mathrm{OD}})^2 \\ &= (a+b+c+d)\mathrm{MO}^2 + 2\overrightarrow{\mathrm{MO}} \cdot \overrightarrow{u} + f(\mathrm{O}), \end{aligned}$$

où $\overrightarrow{u} = a\overrightarrow{\mathrm{OA}} + b\overrightarrow{\mathrm{OB}} + c\overrightarrow{\mathrm{OC}} + d\overrightarrow{\mathrm{OD}}$.

Si $a+b+c+d \neq 0$, on $f(\mathrm{M}) \sim (a+b+c+d)\mathrm{MO}^2$ quand OM tend vers $+\infty$, ce qui contredit le fait que f est identiquement nulle. Donc $a+b+c+d = 0$.

On a alors, pour tout M,

$$f(\mathrm{M}) = 2\overrightarrow{\mathrm{MO}} \cdot \overrightarrow{u} = 0,$$

ce qui nécessite d'avoir $\overrightarrow{u} = \overrightarrow{0}$. Les coefficients a, b, c, d ne sont pas tous nuls. On suppose par exemple $a \neq 0$. Comme $a+b+c+d = 0$, le vecteur $\overrightarrow{u}$ ne dépend pas de O : $\overrightarrow{u} = b\overrightarrow{\mathrm{AB}}+c\overrightarrow{\mathrm{AC}}+d\overrightarrow{\mathrm{AD}}$. Puisque $b+c+d = -a \neq 0$, le point A est le barycentre de la famille $((\mathrm{B}, b), (\mathrm{C}, c), (\mathrm{D}, d))$. Si les points B, C, D sont alignés, il en est de même des quatre points A, B, C, D, ce qui contredit l'hypothèse. Ainsi les points B, C, D ne sont pas alignés. On considère le centre O du cercle circonscrit à BCD. On note R son rayon ($\mathrm{R} > 0$). On a alors

$$\begin{aligned} f(\mathrm{O}) &= a\mathrm{OA}^2 + b\mathrm{OB}^2 + c\mathrm{OC}^2 + d\mathrm{OD}^2 \\ &= a\mathrm{OA}^2 + (b+c+d)\mathrm{R}^2 = a(\mathrm{OA}^2 - \mathrm{R}^2) = 0. \end{aligned}$$

Comme $a \neq 0$, on a nécessairement $\mathrm{OA} = \mathrm{R}$. Les quatre points A, B, D et D sont cocycliques sur le cercle de centre O et de rayon R.

• Supposons réciproquement que les points A, B, C et D sont cocycliques sur un cercle de centre O de rayon R. Pour quatre réels a, b, c, d tels que $a+b+c+d = 0$, on obtient comme précédemment, pour tout M du plan,

$$f(\mathrm{M}) = (a+b+c+d)\mathrm{MO}^2 + 2\overrightarrow{\mathrm{MO}} \cdot \overrightarrow{u} + f(\mathrm{O}) = 2\overrightarrow{\mathrm{MO}} \cdot \overrightarrow{u},$$

car $a+b+c+d=0$ et $f(\mathrm{O})=(a+b+c+d)\mathrm{R}^2=0$. Pour que f soit identiquement nulle, il faut choisir a, b, c, d tels que $\overrightarrow{u}=\overrightarrow{0}$. Comme $a+b+c+d=0$, on a encore $\overrightarrow{u}=b\overrightarrow{\mathrm{AB}}+c\overrightarrow{\mathrm{AC}}+d\overrightarrow{\mathrm{AD}}$. Comme les vecteurs $\overrightarrow{\mathrm{AB}}$, $\overrightarrow{\mathrm{AC}}$ et $\overrightarrow{\mathrm{AD}}$ forment une famille liée, on peut trouver b, c, d non tous nuls tels que $b\overrightarrow{\mathrm{AB}}+c\overrightarrow{\mathrm{AC}}+d\overrightarrow{\mathrm{AD}}=\overrightarrow{0}$. Pour ce choix de b, c, d et $a=-(b+c+d)$, l'application f est identiquement nulle.

2. Pour O et M quelconques, on peut toujours écrire

$$f(\mathrm{M})=(a+b+c+d)\mathrm{MO}^2+2\overrightarrow{\mathrm{MO}}\cdot\overrightarrow{u}+f(\mathrm{O}).$$

• Si $a+b+c+d\neq 0$, on prend pour O le barycentre de $((\mathrm{A},a),(\mathrm{B},b),(\mathrm{C},c),(\mathrm{D},d))$. Alors $\overrightarrow{u}=\overrightarrow{0}$ et, pour tout point M du plan

$$f(\mathrm{M})=(a+b+c+d)\mathrm{MO}^2+f(\mathrm{O}).$$

Les lignes de niveau non vides sont des cercles de centre O.

• Si $a+b+c+d=0$, le vecteur $\overrightarrow{u}$ ne dépend pas de O et, pour tout point M du plan,

$$f(\mathrm{M})=2\overrightarrow{\mathrm{MO}}\cdot\overrightarrow{u}+f(\mathrm{O}).$$

Si $\overrightarrow{u}\neq\overrightarrow{0}$, les lignes de niveau de f sont des droites orthogonales au vecteur $\overrightarrow{u}$. Si $\overrightarrow{u}=\overrightarrow{0}$, l'application f est constante. ◁

4.15. Problème angulaire

Dans le plan euclidien, on se donne trois points distincts A, B et C. Déterminer les points D tels que

$$(\widehat{\overrightarrow{\mathrm{DB}},\overrightarrow{\mathrm{DC}}})\equiv(\widehat{\overrightarrow{\mathrm{AC}},\overrightarrow{\mathrm{AB}}})\ [\pi]\quad\text{et}\quad(\widehat{\overrightarrow{\mathrm{DC}},\overrightarrow{\mathrm{DA}}})\equiv(\widehat{\overrightarrow{\mathrm{BA}},\overrightarrow{\mathrm{BC}}})\ [\pi].$$

(École polytechnique)

▷ **Solution.**

Si les points A, B, C sont alignés, le point D doit vérifier les égalités

$$(\widehat{\overrightarrow{\mathrm{DB}},\overrightarrow{\mathrm{DC}}})\equiv 0\ [\pi]\quad\text{et}\quad(\widehat{\overrightarrow{\mathrm{DC}},\overrightarrow{\mathrm{DA}}})\equiv 0\ [\pi]$$

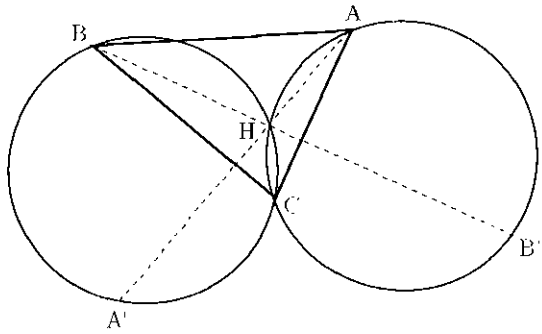
et le point D doit appartenir à la droite (AB) privée des points A, B, C.

On suppose dorénavant que les points A, B, C ne sont pas alignés. L'égalité $(\widehat{\overrightarrow{\mathrm{DB}},\overrightarrow{\mathrm{DC}}})\equiv(\widehat{\overrightarrow{\mathrm{AC}},\overrightarrow{\mathrm{AB}}})\ [\pi]$ ne traduit pas la cocyclicité des quatre points A, B, C, D, car l'ordre des points B, C n'est pas le même à gauche et à droite : le point D serait sur le cercle circonscrit au triangle ABC si on avait $(\widehat{\overrightarrow{\mathrm{DB}},\overrightarrow{\mathrm{DC}}})\equiv(\widehat{\overrightarrow{\mathrm{AB}},\overrightarrow{\mathrm{AC}}})\ [\pi]$.

Pour se ramener à des conditions de cocyclicité, il nous suffit d'introduire A′ le symétrique de A par rapport à (BC) et B′ le symétrique de B par rapport à (CA). En effet, les conditions s'écrivent alors

$$(\widehat{\overrightarrow{DB},\overrightarrow{DC}}) \equiv (\widehat{\overrightarrow{A'B},\overrightarrow{A'C}})\ [\pi] \quad \text{et} \quad (\widehat{\overrightarrow{DC},\overrightarrow{DA}}) \equiv (\widehat{\overrightarrow{B'C},\overrightarrow{B'A}})\ [\pi].$$

Le point D doit donc appartenir aux cercles $\mathcal{C}_1$ et $\mathcal{C}_2$ circonscrits aux triangles A′BC et B′CA respectivement. Ces deux cercles sont forcément distincts : il s'agirait sinon du cercle circonscrit au triangle ABC et les segments [BC] et [AC] devraient être deux diamètres de ce cercle, ce qui est impossible. Ils se coupent donc en C et un autre point, éventuellement confondu avec C.



Montrons que ce deuxième point d'intersection des cercles est l'orthocentre H du triangle ABC. Si $H \neq B, C$, on a, modulo π,

$$\begin{aligned}(\widehat{\overrightarrow{HB},\overrightarrow{HC}}) &\equiv (\widehat{\overrightarrow{HB},\overrightarrow{AC}}) + (\widehat{\overrightarrow{AC},\overrightarrow{AB}}) + (\widehat{\overrightarrow{AB},\overrightarrow{HC}}) \\ &\equiv \frac{\pi}{2} + (\widehat{\overrightarrow{AC},\overrightarrow{AB}}) + \frac{\pi}{2} \equiv (\widehat{\overrightarrow{AC},\overrightarrow{AB}}) \equiv (\widehat{\overrightarrow{A'B},\overrightarrow{A'C}}).\end{aligned}$$

On trouve de même si $H \neq C, A$, $(\widehat{\overrightarrow{HC},\overrightarrow{HA}}) \equiv (\widehat{\overrightarrow{BA},\overrightarrow{BC}})$, donc si le triangle n'est pas rectangle, H appartient à $\mathcal{C}_1$ et $\mathcal{C}_2$.

Si par exemple H = A, on obtient

$$(\widehat{\overrightarrow{HB},\overrightarrow{HC}}) \equiv (\widehat{\overrightarrow{AB},\overrightarrow{AC}}) \equiv \frac{\pi}{2} \equiv (\widehat{\overrightarrow{A'B},\overrightarrow{A'C}})$$

et la conclusion est la même. Le cas H = B est identique.

Si H = C, les cercles $\mathcal{C}_1$ et $\mathcal{C}_2$ sont symétriques par rapport au point C qui appartient aux deux cercles. Ils sont donc tangents en C et n'ont qu'un point d'intersection, le point C.

Conclusion. Si le triangle ABC n'est pas rectangle, le problème a une seule solution, l'orthocentre H du triangle. Si le triangle est rectangle, il n'a pas de solution. ◁

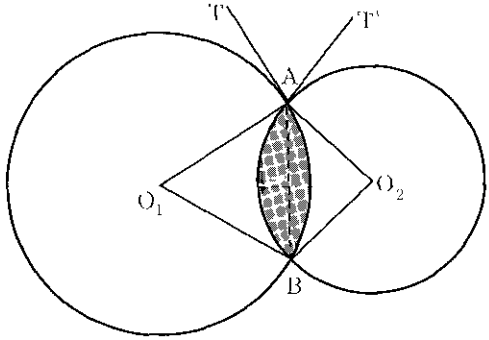
4.16. Intersection de deux disques

Deux disques fermés de rayons respectifs R_1 et R_2 ont en commun un ménisque d'aire $\mathcal{A}$. Déterminer un équivalent de $\mathcal{A}$ lorsque l'angle des deux cercles en leurs points communs tend vers 0.

(**École normale supérieure**)

▷ **Solution.**

Notons $\mathcal{C}_1$ et $\mathcal{C}_2$ les cercles, O_1 et O_2 leurs centres, A et B leurs points communs, (AT) et (AT′) les tangentes à $\mathcal{C}_1$ et $\mathcal{C}_2$ en A et $\theta = \widehat{\mathrm{TAT}'}$. Posons enfin $\alpha = \widehat{\mathrm{AO_1B}}$ et $\beta = \widehat{\mathrm{AO_1B}}$. On suppose que θ tend vers 0.



Pour $i = 1, 2$, on considère l'aire $\mathcal{A}_i$ de la partie du disque de centre O_i de rayon R_i à l'intérieur du secteur angulaire $\widehat{\mathrm{AO_iB}}$. On note S est l'aire du quadrilatère O_1AO_2B. On a alors

$$\mathcal{A}_1 + \mathcal{A}_2 = \mathcal{A} + \mathrm{S}.$$

On obtient $\mathcal{A}_1 = \dfrac{1}{2}\alpha R_1^2$, $\mathcal{A}_2 = \dfrac{1}{2}\beta R_2^2$ et, en écrivant que S est la somme des aires des triangles O_1AB et O_2AB, $\mathrm{S} = \dfrac{1}{2}\sin\alpha\, R_1^2 + \dfrac{1}{2}\sin\beta\, R_2^2$, ce qui donne finalement

$$\mathcal{A} = \mathcal{A}_1 + \mathcal{A}_2 - \mathrm{S} = \frac{R_1^2}{2}(\alpha - \sin\alpha) + \frac{R_2^2}{2}(\beta - \sin\beta).$$

On remarque que, d'après les propriétés des angles au centre, on a

$$\alpha + \beta = 2\left(\pi - \widehat{\mathrm{BAT}}\right) + 2\left(\pi - \widehat{\mathrm{BAT}'}\right) = 2\theta.$$

Ainsi quand θ tend vers 0, α et β tendent vers 0. On en déduit que

$$\mathcal{A} = \frac{R_1^2}{2}\left(\frac{\alpha^3}{6} + o(\alpha^3)\right) + \frac{R_2^2}{2}\left(\frac{\beta^3}{6} + o(\beta^3)\right) = \frac{R_1^2}{12}\alpha^3 + \frac{R_2^2}{12}\beta^3 + o(\theta^3).$$

Mais la loi des sinus dans le triangle O_1AO_2 donne $\dfrac{R_2}{\sin\alpha/2} = \dfrac{R_1}{\sin\beta/2}$. On a donc, quand θ tend vers 0, $\dfrac{R_2}{\alpha} \sim \dfrac{R_1}{\beta}$ et comme $\alpha+\beta = 2\theta$, $\alpha\left(1+\dfrac{R_1}{R_2}\right) \sim 2\theta$, soit $\alpha \sim \dfrac{2R_2}{R_1+R_2}\theta$. On a de même $\beta \sim \dfrac{2R_1}{R_1+R_2}\theta$. On obtient finalement

$$\mathcal{A} \sim \frac{R_1^2}{12}\cdot\frac{8R_2^3}{(R_1+R_2)^3}\theta^3 + \frac{R_1^2}{12}\cdot\frac{8R_1^3}{(R_1+R_2)^3}\theta^3 \sim \boxed{\frac{2(R_1R_2)^2}{3(R_1+R_2)^2}\theta^3}. \lhd$$

Les exercices suivants concernent les coniques.

4.17. Ellipses semblables

Montrer que deux ellipses d'un plan euclidien se déduisent l'une de l'autre par une similitude si, et seulement si, elles ont la même excentricité.

(École polytechnique)

$\rhd$ **Solution.**

L'intervention de l'excentricité nous incite bien entendu à utiliser la définition monofocale des coniques.

• Soit $\mathcal{E}$ une ellipse d'excentricité e, de foyer F et de directrice associée d. Par définition, $\mathcal{E}$ est l'ensemble des points M tels que $MF = eMH$, où H est le projeté orthogonal de M sur d.

Soit s une similitude de rapport k, F′ et d' les images de F et d par s. Pour un point M quelconque, de projeté orthogonal H sur d, on pose $M' = s(M)$ et $H' = s(H)$. Alors, une similitude conservant l'orthogonalité, H′ est le projeté orthogonal de M′ sur d'. Par définition d'une similitude, on a $M'F' = kMF$ et $M'H' = kMH$. Si $M \in \mathcal{E}$ on a $MF = eMH$ et en multipliant cela par k il vient $M'F' = eM'H'$ de sorte que M′ appartient à l'ellipse $\mathcal{E}'$ d'excentricité e, de foyer F′ et de directrice associée d'. On a donc $s(\mathcal{E}) \subset \mathcal{E}'$. En appliquant ce qui précède à la similitude s^{-1}, on obtient $s^{-1}(\mathcal{E}') \subset \mathcal{E}$, c'est-à-dire $\mathcal{E}' \subset s(\mathcal{E})$ et finalement $\mathcal{E}' = s(\mathcal{E})$. L'image d'une ellipse par une similitude est donc une ellipse de même excentricité.

• Soit réciproquement $\mathcal{E}$ et $\mathcal{E}'$ deux ellipses d'excentricité e, de foyers F et F′ de directrices associées d et d'. D'après ce qui précède, pour qu'une similitude s transforme $\mathcal{E}$ en $\mathcal{E}'$, il suffit que $s(F) = F'$ et $s(d) = d'$. En appelant K (resp. K′) le projeté orthogonal de F (resp. F′) sur d (resp. d'), cela équivaut à $s(F) = F'$ et $s(K) = K'$. Il existe une unique

similitude directe (et aussi une unique similitude indirecte) transformant un couple de points distincts en un autre couple de points distincts (pour le montrer, on peut utiliser l'expression complexe d'une similitude qui est $z \longmapsto az + b$ pour une similitude directe et $z \longmapsto a\overline{z} + b$ pour une similitude indirecte, où $(a, b) \in \mathbb{C}^* \times \mathbb{C}$). Il existe donc une similitude transformant $\mathcal{E}$ en $\mathcal{E}'$.

Conclusion. Deux ellipses se déduisent l'une de l'autre par une similitude si, et seulement si, elles ont la même excentricité. ◁

Avec la même démonstration, on montre plus généralement que deux coniques quelconques se déduisent l'une de l'autre par une similitude si et seulement si elles ont la même excentricité. Ainsi deux paraboles quelconques sont toujours semblables.

4.18. Rayon de courbure d'une ellipse

On considère une ellipse $\mathcal{E}$ de $\mathbb{R}^2$ de foyers F_1, F_2. Soit R le rayon de courbure en un point M de l'ellipse et ω l'angle entre la tangente en M et la droite MF_1. On pose $r_1 = \mathrm{F}_1\mathrm{M}$ et $r_2 = \mathrm{F}_2\mathrm{M}$. Montrer que

$$\frac{2}{\mathrm{R}} = \left(\frac{1}{r_1} + \frac{1}{r_2}\right) \sin\omega.$$

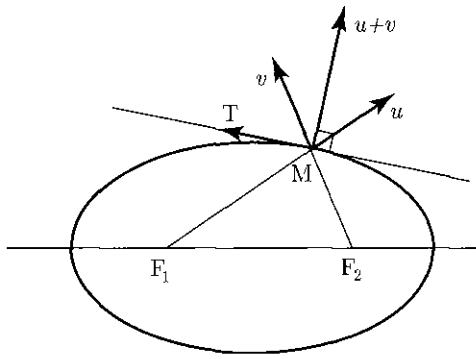
(École polytechnique)

▷ **Solution.**

On se donne un paramétrage normal $s \longmapsto \mathrm{M}$ de $\mathcal{E}$. D'après la définition bifocale de l'ellipse, la fonction $s \longmapsto \mathrm{F}_1\mathrm{M}+\mathrm{F}_2\mathrm{M}$ est constante. Sa dérivée est donc nulle. Comme le gradient de la fonction $\mathrm{M} \longmapsto \mathrm{F}_1\mathrm{M}$ en un point M distinct de F_1 est le vecteur $u = \dfrac{\overrightarrow{\mathrm{F}_1\mathrm{M}}}{\mathrm{F}_1\mathrm{M}}$, la dérivée de $s \longmapsto \mathrm{F}_1\mathrm{M}$ est $s \longmapsto \langle u, \mathrm{T}\rangle$ où $\mathrm{T} = \dfrac{\mathrm{dM}}{\mathrm{d}s}$ est le vecteur tangent unitaire. En posant $v = \dfrac{\overrightarrow{\mathrm{F}_2\mathrm{M}}}{\mathrm{F}_2\mathrm{M}}$ on a donc

$$\langle u + v, \mathrm{T}\rangle = 0.$$

On retrouve ainsi, un résultat bien connu : le vecteur $u + v$ est normal à l'ellipse en M, autrement dit la normale en M est la bissectrice intérieure de l'angle $\widehat{\mathrm{F}_1\mathrm{M}\mathrm{F}_2}$.



Le rayon de courbure géométrique R est défini par $\dfrac{\mathrm{d}\mathrm{T}}{\mathrm{d}s} = \dfrac{\mathrm{N}}{\mathrm{R}}$ où N est le vecteur normal principal (N est unitaire et pointe dans la concavité de l'ellipse). Dérivons la fonction nulle $s \longmapsto \langle u+v, \mathrm{T}\rangle$. Il vient,

$$\frac{1}{\mathrm{R}}\langle u+v, \mathrm{N}\rangle + \langle \frac{\mathrm{d}u}{\mathrm{d}s} + \frac{\mathrm{d}v}{\mathrm{d}s}, \mathrm{T}\rangle = 0.$$

On a $\overrightarrow{\mathrm{F_1M}} = \mathrm{F_1M}\, u$, donc en dérivant,

$$\mathrm{T} = \mathrm{F_1M}\frac{\mathrm{d}u}{\mathrm{d}s} + \langle u, \mathrm{T}\rangle u.$$

En posant $r_1 = \mathrm{F_1M}$ on a $\dfrac{\mathrm{d}u}{\mathrm{d}s} = \dfrac{1}{r_1}(\mathrm{T} - \langle u, \mathrm{T}\rangle u)$ et donc

$$\langle \frac{\mathrm{d}u}{\mathrm{d}s}, \mathrm{T}\rangle = \frac{1}{r_1} - \frac{1}{r_1}\langle u, \mathrm{T}\rangle^2 = \frac{\sin^2\omega}{r_1}$$

car par définition $|\langle u, \mathrm{T}\rangle| = \cos\omega$. On a de même $\langle \dfrac{\mathrm{d}v}{\mathrm{d}s}, \mathrm{T}\rangle = \dfrac{\sin^2\omega}{r_2}$. Par ailleurs on a $\langle u, \mathrm{N}\rangle = \langle v, \mathrm{N}\rangle = \cos\left(\omega + \dfrac{\pi}{2}\right) = -\sin\omega$.

On obtient donc finalement

$$\frac{2\sin\omega}{\mathrm{R}} = \left(\frac{1}{r_1} + \frac{1}{r_2}\right)\sin^2\omega,$$

ce qui donne la relation voulue en simplifiant par $\sin\omega$ (qui n'est pas nul). ◁

Le fait que la tangente à l'ellipse en un point M *est la bissectrice extérieure de l'angle* $\widehat{\mathrm{F_1MF_2}}$ *est une propriété classique et importante. Elle permet par exemple de montrer immédiatement le résultat suivant qui a aussi été proposé à l'oral : dans un billard elliptique, une bille dont la trajectoire passe par le foyer* $\mathrm{F_1}$ *va, après rebond en un point* M *du bord, passer par l'autre foyer* $\mathrm{F_2}$.

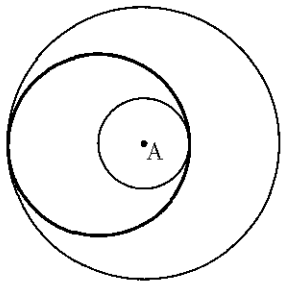
4.19. Cercles tangents à une ellipse

Soit $\mathcal{E}$ une ellipse de $\mathbb{R}^2$. Déterminer les points A pour lesquels il existe quatre points distincts de $\mathcal{E}$ en lesquels un cercle de centre A est tangent à l'ellipse.

(École normale supérieure)

▷ **Solution.**

Notons que l'ensemble à déterminer n'est jamais vide puisqu'il contient toujours le centre O de l'ellipse : on peut trouver un cercle de centre O tangent à l'ellipse en deux sommets opposés, et un autre tangent à l'ellipse aux deux autres sommets. Si l'ellipse est un cercle, il s'agit de la seule solution : si le point A est différent du centre du cercle $\mathcal{E}$, il n'existe que deux cercles de centre A tangents à $\mathcal{E}$, chacun en un seul point.



Nous supposerons donc désormais que l'ellipse $\mathcal{E}$ n'est pas un cercle et, quitte à changer de repère orthonormé, que l'ellipse est donnée par la représentation paramétrique $\begin{cases} x = a\cos t \\ y = b\sin t \end{cases}$, avec $a > b > 0$. Soit $A = (\alpha, \beta)$ un point du plan. La question équivaut clairement à chercher le nombre de points M de l'ellipse tels que la normale à l'ellipse en M passe par A et plus précisément de déterminer pour quels points A il y en a quatre.

Au point de paramètre t, la normale à $\mathcal{E}$ a pour vecteur directeur $(b\cos t, a\sin t)$. Un point de cette normale s'écrit donc

$$(a\cos t, b\sin t) + \lambda(b\cos t, a\sin t) = \big((a+\lambda b)\cos t, (b+\lambda a)\sin t\big),$$

où $\lambda \in \mathbb{R}$. Le point A est le centre d'un cercle tangent à $\mathcal{E}$ au point de paramètre t s'il existe $\lambda \in \mathbb{R}$ tel que $\begin{cases} \alpha = (a+\lambda b)\cos t \\ \beta = (b+\lambda a)\sin t \end{cases}$. Pour un réel λ fixé (distinct de $-b/a$ et $-a/b$), il existe t (unique modulo 2π) tel que ce système soit vérifié si, et seulement si,

$$\left(\frac{\alpha}{a+\lambda b}\right)^2 + \left(\frac{\beta}{b+\lambda a}\right)^2 = 1.$$

De plus, deux valeurs distinctes de λ qui vérifient cette équation conduisent à deux valeurs distinctes de t modulo 2π, donc à deux points distincts de l'ellipse.

Avant de poursuivre, examinons le cas $a+\lambda b=0$. On a alors $\alpha=0$ et notre point A est sur l'axe des ordonnées, axe qui est la normale à l'ellipse aux sommets $(0,b)$ et $(0,-b)$. Pour avoir d'autres points solutions on doit pouvoir trouver t vérifiant $\beta=\left(b-\dfrac{a^2}{b}\right)\sin t$ et $\sin t\neq\pm1$. C'est le cas si et seulement si $-\dfrac{c^2}{b}<\beta<\dfrac{c^2}{b}$ avec $c^2=a^2-b^2$ et on a deux solutions en t distinctes modulo 2π. Les points de cet intervalle ouvert de l'axe des ordonnées sont donc solution. On obtient un intervalle analogue sur l'axe des abscisses dans le cas $\lambda=-\dfrac{b}{a}$, à savoir l'intervalle des points $(\alpha,0)$ avec $-\dfrac{c^2}{a}<\alpha<\dfrac{c^2}{a}$.

Dans le cas général, le problème se ramène donc à déterminer à quelle condition la fonction f définie sur $\mathbb{R}\setminus\left\{-\dfrac{a}{b},-\dfrac{b}{a}\right\}$ par

$$f(\lambda)=\left(\frac{\alpha}{a+\lambda b}\right)^2+\left(\frac{\beta}{b+\lambda a}\right)^2-1$$

admet 4 zéros distincts. Notons déjà qu'il ne peut pas y avoir plus de 4 solutions puisque la recherche des zéros de f conduit à une équation polynomiale de degré 4. Étudions les variations de f. Sa dérivée est définie par $f'(\lambda)=-\dfrac{2\alpha^2b}{(a+b\lambda)^3}-\dfrac{2\beta^2a}{(b+a\lambda)^3}$. Elle s'annule si, et seulement si, $\dfrac{a+b\lambda}{b+a\lambda}=-\left(\dfrac{\alpha^2b}{\beta^2a}\right)^{1/3}$ et donc en une unique valeur $\lambda_0=\dfrac{-a-bu}{b+au}$, en posant $u=\left(\dfrac{\alpha^2b}{\beta^2a}\right)^{1/3}$. Comme $f'(\lambda)$ est positif (resp. négatif) pour $\lambda<-\dfrac{a}{b}$ $\left(\text{resp. } \lambda>-\dfrac{b}{a}\right)$, car somme de termes de même signe et que f a pour limite $+\infty$ en $-\dfrac{a}{b}$ et $-\dfrac{b}{a}$, on obtient le tableau de variation suivant :

x	$-\infty$		$-a/b$		λ_0		$-b/a$		$+\infty$
$f'(x)$	$+$		$\|\|$	$-$	0	$+$	$\|\|$	$-$	
$f(x)$	-1	$\nearrow$	$+\infty \;\|\|\; +\infty$	$\searrow$	$f(\lambda_0)$	$\nearrow$	$+\infty \;\|\|\; +\infty$	$\searrow$	-1

La fonction f s'annule une fois sur chacun des intervalles $\left]-\infty,-\dfrac{a}{b}\right[$ et $\left]-\dfrac{b}{a},+\infty\right[$. Elle s'annule donc quatre fois si, et seulement si, $f(\lambda_0)<0$.

On trouve $a + b\lambda_0 = \dfrac{c^2 u}{b + au}$ et $b + a\lambda_0 = -\dfrac{c^2}{b + au}$, où $c = \sqrt{a^2 - b^2}$. En remplaçant u par son expression, on obtient

$$b + au = \frac{b^{1/3}}{\beta^{2/3}} \left((a\alpha)^{2/3} + (b\beta)^{2/3} \right).$$

On en déduit que

$$\begin{aligned} f(\lambda_0) &= \frac{(b + au)^2}{c^4} \left(\frac{\alpha^2}{u^2} + \beta^2 \right) - 1 \\ &= \frac{b^{2/3}}{\beta^{4/3} c^4} \left((a\alpha)^{2/3} + (b\beta)^{2/3} \right)^2 \left(\frac{\alpha^{2/3} \beta^{4/3} a^{2/3}}{b^{2/3}} + \beta^2 \right) - 1 \\ &= \frac{1}{c^4} \left((a\alpha)^{2/3} + (b\beta)^{2/3} \right)^3 - 1. \end{aligned}$$

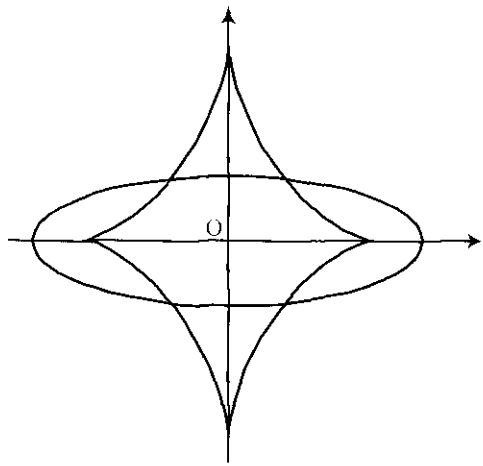
La condition $f(\lambda_0) < 0$ s'écrit donc $(a\alpha)^{2/3} + (b\beta)^{2/3} < c^{4/3}$. On obtient la partie du plan limitée par la courbe d'équation $(a\alpha)^{2/3} + (b\beta)^{2/3} = c^{4/3}$. En écrivant cette équation

$$\left((a\alpha)^{1/3} \right)^2 + \left((b\beta)^{1/3} \right)^2 = \left(c^{2/3} \right)^2,$$

on voit que la courbe qu'elle définit peut se paramétrer sous la forme $(a\alpha)^{1/3} = c^{2/3} \cos t$, $(b\beta)^{1/3} = c^{2/3} \cos t$, ou encore

$$\begin{cases} \alpha = \dfrac{c^2}{a} \cos^3 t \\ \beta = \dfrac{c^2}{b} \sin^3 t. \end{cases}$$

On obtient une *astroïde*, qui n'est autre que la développée de l'ellipse. ◁



Ce résultat s'explique. Donnons-nous, plus généralement, un arc birégulier $s \longmapsto \mathrm{M}(s) = (x(s), y(s))$ *paramétré par l'abscisse curviligne* s. *Un point* $\mathrm{A} = (\mathrm{X}, \mathrm{Y})$ *est sur la normale à l'arc au point* $\mathrm{M}(s)$ *si et seulement si* $\overrightarrow{\mathrm{AM}(s)}$ *est orthogonal à* $\mathrm{M}'(s)$ *ce qui équivaut à*

$$f(s, \mathrm{X}, \mathrm{Y}) = x'(s)(\mathrm{X} - x(s)) + y'(s)(\mathrm{Y} - y(s)) = 0.$$

Imaginons que nous ayons une solution $(s_0, \mathrm{X}_0, \mathrm{Y}_0)$. *Si la dérivée partielle de* f *selon* s *en ce point n'est pas nulle, le théorème des fonctions implicites permet de dire que l'équation* $f(s, \mathrm{X}, \mathrm{Y}) = 0$ *admet une unique solution* s *au voisinage de* s_0 *lorsque* (X, Y) *est au voisinage de* $(\mathrm{X}_0, \mathrm{Y}_0)$. *Or un petit calcul facile montre que le système* $f(s_0, \mathrm{X}_0, \mathrm{Y}_0) = 0$ *et* $\dfrac{\partial f}{\partial s}(s_0, \mathrm{X}_0, \mathrm{Y}_0) = 0$ *a pour solution*

$$\mathrm{X}_0 = x(s_0) - \frac{y'(s_0)}{(x'y'' - x''y')(s_0)} \quad \text{et} \quad \mathrm{Y}_0 = y(s_0) + \frac{x'(s_0)}{(x'y'' - x''y')(s_0)}$$

c'est-à-dire le centre de courbure en $\mathrm{M}(s_0)$. *On en déduit donc que si le point* $\mathrm{A}_0 = (\mathrm{X}_0, \mathrm{Y}_0)$ *n'est pas sur la développée, le nombre de normales à la courbe passant par* $\mathrm{A} = (\mathrm{X}, \mathrm{Y})$ *reste constant lorsque* A *est au voisinage de* A_0. *Ce nombre reste donc constant sur les composantes connexes du complémentaire de la développée, ce que l'on observe dans l'exemple de l'exercice. Nous remercions Nicolas Tosel de nous avoir transmis cette remarque pertinente.*

4.20. Triangles d'aire maximale inscrits dans une ellipse

Quels sont les triangles inscrits dans une ellipse qui sont d'aire maximale ?

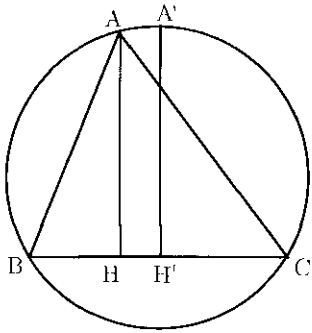
(**École normale supérieure**)

▷ **Solution.**

Soit $\begin{cases} x = a\cos t \\ y = b\sin t \end{cases}$, avec $a \geqslant b > 0$ une représentation paramétrique de l'ellipse $\mathcal{E}$, A, B et C trois points de $\mathcal{E}$. L'affinité orthogonale f dont l'axe est le grand axe de l'ellipse et le rapport $\dfrac{a}{b}$ transforme l'ellipse $\mathcal{E}$ en un cercle Γ de rayon a. Rappelons que f est tout simplement l'application linéaire de $\mathbb{R}^2$ associée à la matrice $\begin{pmatrix} 1 & 0 \\ 0 & a/b \end{pmatrix}$. Notons A_1, B_1 et C_1 les images de A, B et C par f. L'aire du triangle ABC, que l'on notera $\mu(\mathrm{ABC})$, est égale à $\dfrac{1}{2}\left|\det(\overrightarrow{\mathrm{AB}}, \overrightarrow{\mathrm{AC}})\right|$. On a

$$\det(\overrightarrow{A_1B_1}, \overrightarrow{A_1C_1}) = \det f \cdot \det(\overrightarrow{AB}, \overrightarrow{AC}) = \frac{a}{b} \det(\overrightarrow{AB}, \overrightarrow{AC}).$$

On en déduit que $\mu(A_1B_1C_1) = \frac{a}{b}\mu(ABC)$. Ainsi l'aire du triangle ABC est maximale si, et seulement si, l'aire de $A_1B_1C_1$ est maximale. On est donc ramené à traiter le problème pour le cercle Γ. Celui-ci étant compact et la fonction $(A, B, C) \in \Gamma^3 \longmapsto \mu(ABC)$ étant continue, son maximum est atteint. Soit ABC un triangle réalisant ce maximum. Supposons que ce triangle ne soit pas équilatéral. On a par exemple $AB \neq AC$. Soit A' le point d'intersection de Γ et de la médiatrice de [BC] du même côté que A par rapport à (BC). En notant H et H' les projetés respectifs de A et A' sur (BC), on a $A'H' > AH$ et donc $\mu(A'BC) > \mu(ABC)$, ce qui contredit la maximalité de l'aire de ABC.



Le triangle ABC est donc équilatéral. Or, tous les triangles équilatéraux inscrits dans le cercle Γ ont pour côté $2a\cos\frac{\pi}{6} = a\sqrt{3}$ et ont donc la même aire. Ils réalisent donc tous l'aire maximale.

Les triangles inscrits dans l'ellipse $\mathcal{E}$ et d'aire maximale sont donc les images de ces triangles équilatéraux par f^{-1}. Le cercle Γ ayant comme représentation paramétrique $\begin{cases} x = a\cos t \\ y = a\sin t \end{cases}$, un triangle équilatéral correspond à trois points de paramètres t, $t + \frac{2\pi}{3}$, $t + \frac{4\pi}{3}$, où t est un réel quelconque.

Conclusion. Un triangle inscrit dans l'ellipse est d'aire maximale si et seulement s'il existe t tel que les trois sommets aient pour paramètres t, $t + \frac{2\pi}{3}$, $t + \frac{4\pi}{3}$ respectivement. ◁

L'énoncé suivant pose la même question mais avec des rectangles au lieu de triangles.

4.21. Rectangles d'aire maximale inscrits dans une ellipse

Quels sont les rectangles inscrits dans une ellipse qui sont d'aire maximale ?

(École normale supérieure)

▷ **Solution.**

Soit $\begin{cases} x = a\cos\theta \\ y = b\sin\theta \end{cases}$, avec $a > 0$, $b > 0$ une représentation paramétrique de l'ellipse $\mathcal{E}$, et M_1, M_2, M_3 et M_4 quatre points distincts de $\mathcal{E}$. On note θ_i le paramètre de M_i. Le vecteur $\overrightarrow{M_1M_2}$ a pour coordonnées

$$(a(\cos\theta_2 - \cos\theta_1), b(\sin\theta_2 - \sin\theta_1)),$$

c'est-à-dire

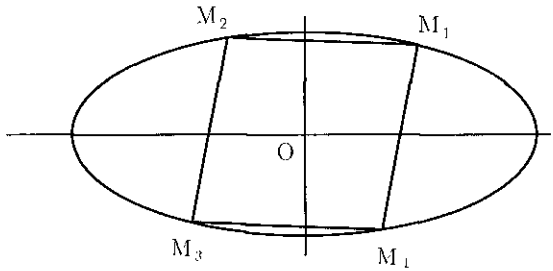
$$\left(-2a\sin\frac{\theta_2-\theta_1}{2}\sin\frac{\theta_2+\theta_1}{2}, 2b\sin\frac{\theta_2-\theta_1}{2}\cos\frac{\theta_2+\theta_1}{2}\right).$$

Ce vecteur est colinéaire à $u = \left(-a\sin\frac{\theta_2+\theta_1}{2}, b\cos\frac{\theta_2+\theta_1}{2}\right)$. De même, (M_3M_4) a pour vecteur directeur $v = \left(-a\sin\frac{\theta_3+\theta_4}{2}, b\cos\frac{\theta_3+\theta_4}{2}\right)$.

Le déterminant de ces deux vecteurs est $ab\sin\frac{\theta_1+\theta_2-\theta_3-\theta_4}{2}$. Ainsi les droites (M_1M_2) et (M_3M_4) sont parallèles si, et seulement si,

$$\theta_1 + \theta_2 \equiv \theta_3 + \theta_4 \pmod{2\pi}.$$

Il en découle que le quadrilatère $M_1M_2M_3M_4$ est un parallélogramme si, et seulement si, $\theta_1 + \theta_2 \equiv \theta_3 + \theta_4 \pmod{2\pi}$ et $\theta_1 + \theta_4 \equiv \theta_3 + \theta_2 \pmod{2\pi}$. On en déduit $\theta_1 \equiv \theta_3 \pmod{\pi}$ et $\theta_2 \equiv \theta_4 \pmod{\pi}$. Comme les points sont supposés distincts, on doit donc avoir $\theta_3 \equiv \theta_1 + \pi \pmod{2\pi}$ et $\theta_4 \equiv \theta_2 + \pi \pmod{2\pi}$. Autrement dit, les points sont deux à deux symétriques par rapport au centre de l'ellipse :



Réciproquement, si $\theta_3 \equiv \theta_1 + \pi \pmod{2\pi}$ et $\theta_4 \equiv \theta_2 + \pi \pmod{2\pi}$, alors $M_1M_2M_3M_4$ est un parallélogramme de centre O. Regardons maintenant à quelle condition ce parallélogramme est un rectangle. Il faut et il suffit pour cela que (M_1M_2) et (M_2M_3) soient orthogonales ce qui équivaut d'après ce qui précède à la nullité du produit scalaire des vecteurs u et $w = \left(-a\sin\frac{\theta_4+\theta_1}{2}, b\cos\frac{\theta_4+\theta_1}{2}\right) = \left(-a\cos\frac{\theta_2+\theta_1}{2}, -b\sin\frac{\theta_2+\theta_1}{2}\right)$, c'est-à-dire de $\frac{1}{2}(a^2-b^2)\sin(\theta_1+\theta_2)$. On discute alors comme suit.

• Si $\mathcal{E}$ est un cercle, $a = b$ et ce produit scalaire est toujours nul. Tout parallélogramme inscrit dans un cercle est un rectangle. L'aire d'un tel rectangle est $4\frac{1}{2}a^2|\sin(\theta_1-\theta_2)|$. Cette aire est maximale si $\theta_2 \equiv \theta_1 + \frac{\pi}{2} \pmod{\pi}$, c'est-à-dire si $M_1M_2M_3M_4$ est un carré. Elle vaut alors $2a^2$.

• Si $\mathcal{E}$ n'est pas un cercle, alors $a \neq b$ et $M_1M_2M_3M_4$ est un rectangle si et seulement si $\theta_2 \equiv -\theta_1 \pmod{\pi}$. Ainsi $M_1M_2M_3M_4$ est un rectangle si et seulement s'il existe $\theta_1 \in \mathbb{R}$ tel que les quatre points aient pour paramètre θ_1, $-\theta_1$, $\theta_1+\pi$ et $-\theta_1+\pi$ (autrement dit les axes du rectangle sont les axes de l'ellipse).

On a alors, en supposant que $\theta_2 = \pi - \theta_1$, $M_1M_2 = 2a|\sin\theta_1|$ et $M_1M_4 = 2b|\cos\theta_1|$. L'aire du rectangle vaut donc

$$\mathcal{A} = 4ab|\sin\theta_1\cos\theta_1| = 2ab|\sin(2\theta_1)|.$$

Elle est maximale si $|\sin(2\theta_1)| = 1$, c'est-à-dire si $2\theta_1 \equiv \frac{\pi}{2} \pmod{\pi}$ ou encore $\theta_1 \equiv \frac{\pi}{4} \pmod{\frac{\pi}{2}}$. Les différentes valeurs de θ_1 donnent le même rectangle. Il existe donc un seul rectangle inscrit dans l'ellipse $\mathcal{E}$, d'aire maximale. Il a pour sommets les points de coordonnées $\left(\pm\frac{a\sqrt{2}}{2}, \pm\frac{b\sqrt{2}}{2}\right)$. Il est d'aire $2ab$. ◁

4.22. Problème de Pappus

Soient A, B, C et D quatre points d'un plan euclidien. Trouver le lieu des points M tels que

$$d(M,(AB)) \times d(M,(CD)) = d(M,(AC)) \times d(M,(BD)).$$

(École polytechnique)

▷ **Solution.**

Notons $\mathcal{E}$ le lieu cherché. Il contient toujours les points A, B, C et D. On suppose que $A \neq B$, $A \neq C$, $D \neq B$ et $D \neq C$.

• Si les quatre points sont alignés, $\mathcal{E}$ est égal au plan tout entier.

• Si trois des points, par exemple A, B, C sont alignés, sans qu'ils soient tous alignés, le point M appartient à $\mathcal{E}$ si, et seulement si,

$$d(\mathrm{M},(\mathrm{AB}))\times d(\mathrm{M},(\mathrm{CD}))=d(\mathrm{M},(\mathrm{AB}))\times d(\mathrm{M},(\mathrm{BD})),$$

c'est-à-dire si et seulement si $\mathrm{M}\in(\mathrm{AB})$ ou si $d(\mathrm{M},(\mathrm{CD}))=d(\mathrm{M},(\mathrm{BD}))$. Dans ce cas, l'ensemble $\mathcal{E}$ est la réunion de la droite (AB) et des bissectrices de l'angle $\widehat{\mathrm{BDC}}$.

• Supposons que les points soient en position générale (trois à trois non alignés). On rapporte le plan à un repère orthonormal. Toute droite Δ possède une équation normale, c'est-à-dire de la forme $ax+by+c=0$, avec $a^2+b^2=1$. Pour tout point M de coordonnées (x,y), la distance de M à la droite Δ est égale à $|ax+by+c|$. Notons $f(x,y)=0$, $g(x,y)=0$, $h(x,y)=0$ et $k(x,y)=0$ des équations normales des droites (AB), (CD), (AC) et (BD). Alors M appartient à $\mathcal{E}$ si, et seulement si, $|f(x,y)||g(x,y)|=|h(x,y)||k(x,y)|$, *i.e.* $f(x,y)g(x,y)=\pm h(x,y)k(x,y)$. Cela montre que $\mathcal{E}$ est la réunion de deux coniques, éventuellement dégénérées. Les seuls points d'intersection entre les deux coniques sont A, B, C et D. ◁

4.23. Cocyclicité de quatre points d'une ellipse

Soit $\mathcal{E}$ l'ellipse de représentation paramétrique $\begin{cases} x=a\cos t \\ y=b\sin t \end{cases}$ avec $a>b>0$, M_1, M_2, M_3, M_4 quatre points de $\mathcal{E}$ distincts, de paramètres t_1, t_2, t_3, t_4. Montrer que les points M_1, M_2, M_3, M_4 sont cocycliques si, et seulement si,

$$t_1+t_2+t_3+t_4\in 2\pi\mathbb{Z}.$$

(École polytechnique)

▷ **Solution.**

• Supposons que les quatre points soient cocycliques. Ce sont les points d'intersection de $\mathcal{E}$ et d'un cercle d'équation $x^2+y^2+2\alpha x+2\beta y+\gamma=0$. Pour $t\in\{t_1,t_2,t_3,t_4\}$ on a donc

$$a^2\cos^2 t+b^2\sin^2 t+2\alpha a\cos t+2\beta b\sin t+\gamma=0.$$

Remplaçons le cosinus et le sinus à l'aide des formules d'Euler. Il vient

$$\frac{a^2}{4}(e^{2it}+e^{-2it}+2)+\frac{b^2}{4}(2-e^{2it}-e^{-2it})$$
$$+\alpha a(e^{it}+e^{-it})-i\beta b(e^{it}-e^{-it})+\gamma=0.$$

En multipliant par e^{2it} et en ordonnant les termes cela donne

$$\frac{a^2-b^2}{4}e^{4it}+(\alpha a-i\beta b)e^{3it}+\left(\frac{a^2+b^2}{2}+\gamma\right)e^{2it}+(\alpha a+i\beta b)e^{it}+\frac{a^2-b^2}{4}=0.$$

Autrement dit, les nombres complexes $z_k=e^{it_k}$ pour $k=1,\dots,4$ sont racines du polynôme

$$\mathrm{P}(\mathrm{X})=\frac{a^2-b^2}{4}\mathrm{X}^4+(\alpha a-i\beta b)\mathrm{X}^3+\left(\frac{a^2+b^2}{2}+\gamma\right)\mathrm{X}^2+(\alpha a+i\beta b)\mathrm{X}+\frac{a^2-b^2}{4}.$$

Or il est clair que le produit des racines de ce polynôme est égal à 1. On a donc $1=z_1z_2z_3z_4=e^{i(t_1+t_2+t_3+t_4)}$, d'où il découle $t_1+t_2+t_3+t_4\in 2\pi\mathbb{Z}$.

• Supposons réciproquement que $t_1+t_2+t_3+t_4\equiv 0 \pmod{2\pi}$. Considérons le cercle $\mathcal{C}$ circonscrit au triangle $\mathrm{M_1M_2M_3}$. Il coupe $\mathcal{E}$ en quatre points $\mathrm{M_1}$, $\mathrm{M_2}$, $\mathrm{M_3}$ et M'_4. En notant t'_4 le paramètre de M'_4, on obtient d'après ce qui précède $t_1+t_2+t_3+t'_4\equiv 0 \pmod{2\pi}$ et donc $t_4\equiv t'_4 \pmod{2\pi}$, c'est-à-dire $\mathrm{M_4}=\mathrm{M}'_4$. Ainsi les points $\mathrm{M_1}$, $\mathrm{M_2}$, $\mathrm{M_3}$ et $\mathrm{M_4}$ sont cocycliques sur $\mathcal{C}$. ◁

4.24. Cercle orthoptique d'une ellipse

Déterminer l'ensemble des points où passent deux tangentes orthogonales à une ellipse $\mathcal{E}$ donnée.

(École polytechnique)

▷ **Solution.**

Considérons un repère orthonormé dans lequel l'ellipse $\mathcal{E}$ a pour équation $\frac{x^2}{a^2}+\frac{y^2}{b^2}=1$, avec $(a,b)\in(\mathbb{R}_+^*)^2$. Par la règle du dédoublement, la tangente à l'ellipse au point de coordonnées (x_0,y_0) a pour équation $\frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1$. Soit Δ une droite d'équation $ux+vy=h$ avec (u,v) non nul. La droite Δ est tangente à l'ellipse s'il existe $(x_0,y_0)\in\mathcal{E}$ tel que

$$ux+vy=h \iff \frac{x_0x}{a^2}+\frac{y_0y}{b^2}=1.$$

Cela équivaut à $h\neq 0$ et $x_0=\frac{a^2u}{h}$, $y_0=\frac{b^2v}{h}$. Le point (x_0,y_0) ainsi

défini appartient à $\mathcal{E}$ si, et seulement si, $\frac{a^2u^2}{h^2} + \frac{b^2v^2}{b^2} = 1$, soit encore $a^2u^2 + b^2v^2 = h^2$. Réciproquement, tout triplet non nul (u, v, h) vérifiant cette relation définit une droite Δ qui est tangente à l'ellipse. On appelle cette relation *une équation tangentielle* de l'ellipse.

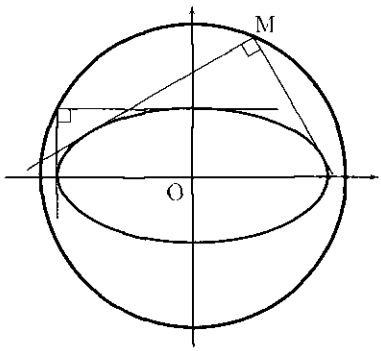
Soit $M = (\alpha, \beta)$ un point du plan. Une droite passant par M a pour équation $ux + vy = u\alpha + v\beta$ avec (u, v) non nul. Elle est tangente à $\mathcal{E}$ si, et seulement si, $a^2u^2 + b^2v^2 = (u\alpha + v\beta)^2$, c'est-à-dire

$$(a^2 - \alpha^2)u^2 - 2\alpha\beta uv + (b^2 - \beta^2)v^2 = 0.$$

• Supposons $b^2 - \beta^2 \neq 0$. Si (u, v) est un couple non nul qui vérifie l'équation précédente on a forcément $u \neq 0$ (la droite n'est pas verticale) et la pente $m = -\frac{v}{u}$ vérifie l'équation du second degré

$$(b^2 - \beta^2)m^2 + 2\alpha\beta m + a^2 - \alpha^2 = 0.$$

Cette équation admet deux solutions réelles si, et seulement si, son discriminant est positif ce qui équivaut à $\frac{\alpha^2}{a^2} + \frac{\beta^2}{b^2} - 1 \geqslant 0$. Géométriquement cette condition signifie que M doit être extérieur à l'ellipse $\mathcal{E}$. Dans le cas où M est sur l'ellipse il n'y a bien entendu qu'une seule tangente à l'ellipse passant par M : c'est le cas où l'équation admet une racine double. Plaçons nous dans le cas où il y a deux tangentes distinctes de pentes m_1 et m_2. Celles-ci sont orthogonales si, et seulement si, $m_1m_2 = -1$. Cela équivaut à $\frac{a^2 - \alpha^2}{b^2 - \beta^2} = -1$, c'est-à-dire à $\alpha^2 + \beta^2 = a^2 + b^2$. On vérifie réciproquement que tout point (α, β) de ce cercle avec $\beta \neq \pm b$ est bien extérieur à l'ellipse et convient.



• Si $b^2 - \beta^2 = 0$ alors il y a une tangente verticale à l'ellipse passant par M (cas $u = 0$, v quelconque non nul). On a une seconde tangente donnée par la relation $(a^2 - \alpha^2)u - 2\alpha\beta v = 0$, à savoir la droite de pente $\frac{a^2 - \alpha^2}{2\alpha\beta}$

(on a $\beta \neq 0$ et on peut supposer $\alpha \neq 0$ sans quoi M est sur l'ellipse). Les deux droites sont orthogonales si, et seulement si, $a^2 - c^2 = 0$. On trouve les quatre points du cercle précédent qui nous manquaient.

Conclusion. L'ensemble des points où passent deux tangentes orthogonales à une ellipse d'équation $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ est le cercle d'équation $x^2 + y^2 = a^2 + b^2$, appelé *cercle orthoptique.* ◁

Le même problème pour une hyperbole d'équation $\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$ *conduit à l'ensemble d'équation* $x^2+y^2 = a^2-b^2$. *Si* $a < b$ *c'est l'ensemble vide. Si* $a = b$, *i.e. si l'hyperbole est équilatère, on trouve un seul point, le centre de l'hyperbole mais il ne convient pas vraiment : les droites orthogonales sont les deux asymptotes et elles sont tangentes à l'hyperbole au point à l'infini. Enfin lorsque* $a > b$ *on obtient un cercle auquel il faut ôter les quatre points d'intersection avec les deux asymptotes (car pour ces points l'une des deux droites obtenues est parallèle à une asymptote et ne coupe pas l'hyperbole). Notons que la condition* $a > b$ *équivaut à* $2a^2 > a^2 + b^2$. *Comme* $\frac{a^2+b^2}{a^2}$ *est le carré de l'excentricité* e, *elle s'écrit aussi* $1 < e < \sqrt{2}$.

Pour une parabole, on trouve la directrice.

L'exercice suivant envisage le même problème pour un ellipsoïde de l'espace $\mathbb{R}^3$.

4.25. Sphère orthoptique d'un ellipsoïde

Soit S la surface d'équation $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$, où a, b, c sont des réels strictement positifs. Déterminer l'ensemble des points par lesquels passent trois plans tangents à S, deux à deux perpendiculaires.

(École polytechnique)

▷ **Solution.**

Nous allons démontrer que le lieu cherché est une sphère de centre 0. Par la règle du dédoublement, une équation du plan tangent à S au point $(x_0, y_0, z_0) \in$ S est

$$\frac{x_0 x}{a^2} + \frac{y_0 y}{b^2} + \frac{z_0 z}{c^2} = 1.$$

Le plan d'équation $ux + vy + wz = h$ avec (u, v, w) non nul est donc tangent à S si et seulement si $h \neq 0$ et $\frac{1}{h}(ua^2, vb^2, wc^2)$ appartient à S, c'est-à-dire si, et seulement si,

$$a^2u^2 + b^2v^2 + c^2w^2 = h^2.$$

Soit $X = (\alpha, \beta, \gamma) \in \mathbb{R}^3$. Un plan contenant le point X a une équation de la forme $ux + vy + wz = u\alpha + v\beta + w\gamma$. Il est tangent à S si, et seulement si, $a^2u^2 + b^2v^2 + c^2w^2 = (u\alpha + v\beta + w\gamma)^2$.

Considérons la forme quadratique définie par

$$q(u, v, w) = a^2u^2 + b^2v^2 + c^2w^2 - (u\alpha + v\beta + w\gamma)^2.$$

On cherche les points $X = (\alpha, \beta, \gamma)$ pour lesquels il existe trois vecteurs non nuls $U = (u, v, w)$, $U' = (u', v', w')$ et $U'' = (u'', v'', w'')$, deux à deux orthogonaux tels que $q(U) = q(U') = q(U'') = 0$. En effet, la première condition traduit le fait que les plans $\mathrm{Vect}(U, U')$, $\mathrm{Vect}(U, U'')$ et $\mathrm{Vect}(U', U'')$ sont deux à deux perpendiculaires, la deuxième le fait qu'ils sont tangents à S et contiennent X. En considérant des vecteurs de norme 1 colinéaires à U, U' et U'', cela équivaut à l'existence d'une base orthonormale (e_1, e_2, e_3) de $\mathbb{R}^3$ formée de vecteurs q-isotropes (*i.e.* tels que $q(e_1) = q(e_2) = q(e_3) = 0$).

• Si une telle base existe, la matrice de q dans cette base a une trace nulle. Cette trace ne dépend pas de la base orthonormale choisie, car les matrices d'une forme quadratique dans deux bases orthonormales sont semblables : si l'une est A, l'autre s'écrit ${}^tQAQ$, avec Q orthogonale et donc $Q^{-1}AQ$. On obtient, à partir de l'expression de q dans la base canonique de $\mathbb{R}^3$,

$$\mathrm{Tr}(q) = (a^2 - \alpha^2) + (b^2 - \beta^2) + (c^2 - \gamma^2) = 0$$

et donc $\alpha^2 + \beta^2 + \gamma^2 = a^2 + b^2 + c^2$: le point X appartient à la sphère S' de centre 0 et de rayon $\sqrt{a^2 + b^2 + c^2}$.

• Supposons réciproquement que X appartient à S', c'est-à-dire que $\mathrm{Tr}(A) = 0$, où A est la matrice de q dans la base canonique de $\mathbb{R}^3$. Nous avons démontré dans l'exercice 2.8 que toute matrice est orthogonalement semblable à une matrice dont tous les termes diagonaux sont égaux. Il existe $P \in O_3(\mathbb{R})$ telle que $B = {}^tPAP$ ait des termes diagonaux égaux. Comme B est semblable à A, on a $\mathrm{Tr}(B) = \mathrm{Tr}(A) = 0$, donc tous les termes diagonaux de B sont nuls. Soit $\mathcal{B} = (e_1, e_2, e_3)$ la base orthonormale de $\mathbb{R}^3$ telle que P soit la matrice de passage de la base canonique de $\mathbb{R}^3$ à $\mathcal{B}$. Alors B est la matrice de q dans la base $\mathcal{B}$ et on a donc $q(e_1) = q(e_2) = q(e_3) = 0$. Le point X appartient à l'ensemble cherché.

Conclusion. L'ensemble des points par lesquels passent trois plans tangents à l'ellipsoïde d'équation $\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$ et deux à deux perpendiculaires est la sphère de centre 0 et de rayon $\sqrt{a^2 + b^2 + c^2}$, appelée *sphère orthoptique* de l'ellipsoïde. ◁

Cette démonstration s'applique sans changement pour un ellipsoïde de $\mathbb{R}^n$.

Nous poursuivons cette série avec deux exercices assez calculatoires.

4.26. Centres des cercles circonscrits à une famille de triangles inscrits dans une parabole

On considère la parabole (P) d'équation $y^2 = 2px$. Une droite Δ variable passant par le foyer coupe (P) en A et B. Déterminer l'ensemble des centres des cercles circonscrits aux triangles OAB.

(École polytechnique)

▷ **Solution.**

Le foyer F a pour coordonnées $\left(\frac{p}{2}, 0\right)$. On écarte le cas où Δ est l'axe de la parabole, car alors elle ne coupe P qu'en un point. Il existe donc $m \in \mathbb{R}$ tel que Δ ait pour équation $x - \frac{p}{2} = my$. On note $\left(\frac{a^2}{2p}, a\right)$ et $\left(\frac{b^2}{2p}, b\right)$ les coordonnées de A et B. Alors a et b sont les solutions de

$$y^2 - 2pmy - p^2 = 0. \tag{E}$$

Soit (x, y) les coordonnées du centre du cercle circonscrit Ω au triangle OAB. Une équation de la médiatrice de [OA] est

$$\frac{a^2}{2p}\left(x - \frac{a^2}{4p}\right) + a\left(y - \frac{a}{2}\right) = 0,$$

soit $\frac{1}{2p}x + \frac{1}{a}y = \frac{a^2}{8p^2} + \frac{1}{2}$. De même, la médiatrice de [OB] a pour équation $\frac{1}{2p}x + \frac{1}{b}y = \frac{b^2}{8p^2} + \frac{1}{2}$. En soustrayant les deux équations, on obtient $\frac{a-b}{ab}y = \frac{b^2-a^2}{8p^2}$ et donc $-\frac{1}{ab}y = \frac{b+a}{8p^2}$ puisque $a \neq b$. Comme a, b sont les solutions de (E) on a $ab = -p^2$ et $b + a = 2mp$ donc $y = \frac{pm}{4}$.

En additionnant les deux équations, on obtient

$$\frac{1}{p}x + \frac{a+b}{ab}y = \frac{a^2+b^2}{8p^2} + 1,$$

soit $\frac{1}{p}x - \frac{2m}{p}y = \frac{2m^2+1}{4} + 1$, car $a^2+b^2 = (a+b)^2 - 2ab = 4p^2m^2 + 2p^2$. On en déduit que $x = pm^2 + \frac{5p}{4}$ et donc que

$$x = \frac{16y^2}{p} + \frac{5p}{4}.$$

Ainsi Ω appartient à une parabole. Comme m décrit $\mathbb{R}$, il en est de même de y et le lieu de Ω est la parabole en entier. Cette parabole a le même axe que (P), ce qui est normal par symétrie. $\lhd$

4.27. Centres des triangles équilatéraux inscrits dans une parabole

Déterminer le lieu des centres des triangles équilatéraux inscrits dans la parabole d'équation $y^2 = 2px$ du plan euclidien $\mathbb{R}^2$.

(École polytechnique)

$\rhd$ **Solution.**

Soit $\mathrm{M}_i = \left(\dfrac{y_i^2}{2p}, y_i\right)$ $(1 \leqslant i \leqslant 3)$ trois points distincts de la parabole, et $\mathrm{G} = (a,b) = \left(\dfrac{y_1^2 + y_2^2 + y_3^2}{6p}, \dfrac{y_1 + y_2 + y_3}{3}\right)$ le centre de gravité du triangle $\mathrm{M}_1\mathrm{M}_2\mathrm{M}_3$. Le triangle $\mathrm{M}_1\mathrm{M}_2\mathrm{M}_3$ est équilatéral si, et seulement si, G est aussi orthocentre du triangle, c'est-à-dire si, et seulement si, on a $\overrightarrow{\mathrm{GM}_1} \cdot \overrightarrow{\mathrm{M}_2\mathrm{M}_3} = 0$ et les autres égalités obtenues par permutation circulaire. On obtient

$$\left(\frac{y_1^2}{2p} - a\right)\left(\frac{y_3^2}{2p} - \frac{y_2^2}{2p}\right) + (y_1 - b)(y_3 - y_2) = 0$$

et après simplification par $y_3 - y_2$ et multiplication par $4p^2$,

$$\left(y_1^2 - 2pa\right)(y_3 + y_2) + 4p^2(y_1 - b) = \left(y_1^2 - 2pa\right)(3b - y_1) + 4p^2(y_1 - b) = 0,$$

car $y_1 + y_2 + y_3 = 3b$. Ainsi y_1, et par permutation circulaire y_2 et y_3, sont racines du polynôme

$$\mathrm{P} = \mathrm{X}^3 - 3b\mathrm{X}^2 + (2pa + 4p^2)\mathrm{X} + (6pab + 4p^2 b).$$

Les fonctions symétriques des racines vérifient

$$\sigma_1 = 3b \quad \text{et} \quad \sigma_2 = -(2pa + 4p^2).$$

Mais par définition de a,

$$y_1^2 + y_2^2 + y_3^2 - 6ap = \sigma_1^2 - 2\sigma_2 = 9b^2 + 4ap + 8p^2.$$

On a donc $a = \dfrac{9}{2p} b^2 + 4p$.

Réciproquement, soit $G = (a, b)$ tel que $a = \frac{9}{2p}b^2 + 4p$. Considérons les racines y_1, y_2, y_3 du polynôme P, où P est défini comme précédemment et admettons provisoirement qu'elles sont bien réelles. On définit les points $M_i = \left(\frac{y_i^2}{2p}, y_i\right)$. On a alors

$$\begin{cases} \dfrac{y_1 + y_2 + y_3}{3} = \dfrac{\sigma_1}{3} = b \\ \dfrac{y_1^2 + y_2^2 + y_3^2}{6p} = \dfrac{\sigma_1^2 - 2\sigma_2}{6p} = \dfrac{9b^2 + 4pa + 8p^2}{6p} = a, \end{cases}$$

car $a = \frac{9}{2p}b^2 + 4p$, donc G est le centre de gravité du triangle $M_1M_2M_3$. Et, d'après ce qui précède, comme y_1, y_2, y_3 sont racines de P, le triangle $M_1M_2M_3$ est bien équilatéral.

Il suffit maintenant de vérifier que, pour tout réel b, le polynome

$$P = X^3 - 3bX^2 - (9b^2 + 12p^2)X + (27b^3 + 28p^2b).$$

(on a remplacé a par son expression en fonction de b) possède trois racines distinctes. Le polynôme dérivé $P' = 3X^2 - 6bX - (9b^2 + 12p^2)$ possède deux racines distinctes $b - 2\sqrt{b^2 + p^2}$ et $b + 2\sqrt{b^2 + p^2}$. Le polynôme P a trois racines distinctes si, et seulement si,

$$P\left(b - 2\sqrt{b^2 + p^2}\right) P\left(b + 2\sqrt{b^2 + p^2}\right) < 0.$$

Le reste dans la division de P par P′ est $-8(b^2 + p^2)(X - 3b)$. On étudie donc le signe de

$$\left(8(b^2 + p^2)\right)^2(-2b - 2\sqrt{b^2 + p^2})(-2b + 2\sqrt{b^2 + p^2}) = -4\left(8(b^2 + p^2)\right)^2 p^2.$$

C'est bien un terme strictement négatif : le polynôme P a trois racines distinctes.

Conclusion. Le lieu cherché est la parabole d'équation $x = \frac{9}{2p}y^2 + 4p$, ou $y^2 = \frac{2p}{9}(x - 4p)$. Elle a même axe que la parabole initiale et son sommet est $(4p, 0)$. ◁

On peut se poser le même problème pour une ellipse ou une hyperbole.

• *Pour une ellipse d'équation* $\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$ *avec* $0 < b < a$, *le lieu des centres des triangles équilatéraux est l'ellipse d'équation*

$$\frac{(a^2 + 3b^2)^2x^2}{(a^2 - b^2)^2a^2} + \frac{(3a^2 + b^2)^2y^2}{(a^2 - b^2)^2b^2} = 1.$$

Dans le cas où $a = b$ *l'ellipse de départ est un cercle et on obtient seulement le point* $(0, 0)$.

• *Pour un hyperbole d'équation* $\dfrac{x^2}{a^2} - \dfrac{y^2}{b^2} = 1$, *le lieu des centres des triangles équilatéraux a pour équation*

$$\frac{(a^2-3b^2)^2x^2}{(a^2+b^2)^2a^2} - \frac{(3a^2-b^2)^2y^2}{(a^2+b^2)^2b^2} = 1.$$

Si $a = \dfrac{b}{\sqrt{3}}$, *on obtient la réunion de deux droites (d'équation* $x = \pm\dfrac{a}{2}$*), si* $a = b\sqrt{3}$, *l'ensemble vide, et dans le cas général* $a \notin \{\dfrac{b}{\sqrt{3}}, b\sqrt{3}\}$ *une hyperbole de mêmes axes que l'hyperbole initiale.*

L'énoncé suivant fait apparaître une quadrique.

4.28. Points équidistants de deux droites non coplanaires

Dans un espace euclidien de dimension 3, trouver l'ensemble des points équidistants de deux droites affines non coplanaires données.
(**École polytechnique**)

▷ **Solution.**

Soient D et D′ les deux droites, $H \in D$ et $H' \in D'$ les deux points tels que (HH′) soit la perpendiculaire commune à D et D′. On choisit un repère orthonormal $(0, i, j, k)$ de l'espace tel que le point O soit le milieu de [HH′], $(O, k) = (HH')$. Il existe $a \in \mathbb{R}^*$ tels que les droites D et D′ soient dans les plans $z = a$ et $z = -a$ respectivement. En choisissant de plus i et j tels que (O, i) et (O, j) soient les bissectrices des projetées de D et D′ sur le plan orthogonal à k contenant 0, il existe m tel que D et D′ aient pour équations respectives

$$D \begin{cases} y = mx \\ z = a \end{cases} \quad \text{et} \quad D' \begin{cases} y = -mx \\ z = -a \end{cases}.$$

Soit M de coordonnées (x, y, z). En appelant P le projeté de M sur le plan d'équation $z = a$, la distance d de M à D vérifie

$$d^2 = MP^2 + (d(P, D))^2 = (z-a)^2 + \frac{(y-mx)^2}{1+m^2}.$$

On obtient de même pour la distance d' de M à D′,

$$d'^2 = (z+a)^2 + \frac{(y+mx)^2}{1+m^2}.$$

On cherche les points M vérifiant

$$(z-a)^2+\frac{(y-mx)^2}{1+m^2}=(z+a)^2+\frac{(y+mx)^2}{1+m^2}.$$

Cela équivaut à $az+\dfrac{mxy}{1+m^2}=0$ soit encore $z=-\dfrac{m}{a(1+m^2)}xy$. L'ensemble des points M équidistants des droites D et D′ est donc un paraboloïde hyperbolique. ◁

Avant d'aborder le thème des problèmes d'extrema, voici un exercice qui n'utilise que les propriétés générales de la distance. Il fournit une petite introduction à la notion de diamètre transfini d'une partie du plan.

4.29. Diamètre transfini

Soit E une partie non vide de $\mathbb{R}^2$. On pose, pour $n\geqslant 2$ et $(a_1,\dots,a_n)\in \mathrm{E}^n$,

$$\delta_n(a_1\dots,a_n)=\left(\prod_{1\leqslant i<j\leqslant n} d(a_i,a_j)\right)^{\frac{2}{n(n-1)}}$$

(moyenne géométrique des distances entre les points pris deux à deux), puis

$$d_n(\mathrm{E})=\sup_{(a_1,\dots,a_n)\in \mathrm{E}^n}\delta_n(a_1,\dots,a_n)\in\mathbb{R}\cup\{+\infty\}.$$

1. Que dire de $d_n(\mathrm{E})$ suivant que E est bornée ou non bornée ? Si $\mathrm{E}\subset\mathrm{F}$, comparer $d_n(\mathrm{E})$ et $d_n(\mathrm{F})$.

2. On suppose E bornée. Montrer que la suite $(d_n(\mathrm{E}))_{n\geqslant 0}$ est décroissante et converge. *Sa limite est appelée le diamètre transfini de la partie* E.

3. Calculer $d_3(\mathrm{E})$ lorsque E est un segment.

(École polytechnique)

▷ **Solution.**

1. On remarque déjà que si E est un ensemble fini alors $d_n(\mathrm{E})=0$ dès que n est strictement plus grand que Card E. On supposera donc dans la suite que E est une partie infinie. Notons aussi que $d_2(\mathrm{E})$ est le diamètre usuel de la partie E.

• Si E est borné de diamètre $d_2(\mathrm{E})$, alors pour tout $(a,b)\in\mathrm{E}^2$, on a $d(a,b)\leqslant d_2(\mathrm{E})$. On en déduit que, si $n\geqslant 2$ et $(a_1,\dots,a_n)\in\mathrm{E}^n$, alors

$$\prod_{1\leqslant i<j\leqslant n} d(a_i,a_j)\leqslant d_2(\mathrm{E})^{\frac{n(n-1)}{2}}\quad\text{et}\quad \delta_n(a_1,\dots,a_n)\leqslant d_2(\mathrm{E}).$$

Ainsi, si E est borné, on a $d_n(\mathrm{E}) \leqslant d_2(\mathrm{E})$, pour tout $n \geqslant 2$.

• Supposons que E ne soit pas borné. Soit $\mathrm{K} > 0$. Montrons par récurrence que pour tout $n \geqslant 2$, il existe $(a_1, \ldots, a_n) \in \mathrm{E}^n$ tel que si $1 \leqslant i < j \leqslant n$ alors $d(a_i, a_j) \geqslant \mathrm{K}$. C'est clair pour $n = 2$. Supposons le résultat au rang n. La réunion des boules fermés de rayon K centrées en les a_k pour $1 \leqslant k \leqslant n$ ne recouvre pas E car E n'est pas borné. On peut donc trouver $a_{n+1} \in \mathrm{E}$ tel que $d(a_{n+1}, a_k) \geqslant \mathrm{K}$ pour tout $k \in [\![1, n]\!]$ ce qui donne un $(n+1)$-uplet vérifiant la propriété souhaitée.

Pour un tel choix de $a_1, \ldots, a_n$, on a alors $\delta_n(a_1, \ldots, a_n) \geqslant \mathrm{K}$, et donc $d_n(\mathrm{E}) \geqslant \mathrm{K}$. Comme K est arbitraire, on a donc $d_n(\mathrm{E}) = +\infty$ pour tout $n \geqslant 2$.

Soit E et F tels que $\mathrm{E} \subset \mathrm{F}$. Si $(a_1, \ldots, a_n) \in \mathrm{E}$ alors $(a_1, \ldots, a_n) \in \mathrm{F}$ et $\delta_n(a_1, \ldots, a_n) \leqslant d_n(\mathrm{F})$. On en déduit que $d_n(\mathrm{E}) \leqslant d_n(\mathrm{F})$ pour tout $n \geqslant 2$.

2. Montrons que la suite $(d_n(\mathrm{E}))_{n \geqslant 2}$ est décroissante. Soit $n \geqslant 2$ et $(a_1, \ldots, a_{n+1}) \in \mathrm{E}^{n+1}$. Pour $k \in [\![1, n+1]\!]$, considérons le produit

$$\mathrm{P}_k = \prod_{\substack{1 \leqslant i < j \leqslant n+1 \\ i = k,\, j \neq k}} d(a_i, a_j).$$

On a, par définition,

$$\mathrm{P}_k = \delta_n(a_1, \ldots, \widehat{a_k}, \ldots, a_{n+1})^{\frac{n(n-1)}{2}} \leqslant d_n(\mathrm{E})^{\frac{n(n-1)}{2}}.$$

Considérons le produit $\mathrm{P}_1 \ldots \mathrm{P}_{n+1}$. Il est égal à $\left(\prod_{1 \leqslant i < j \leqslant n+1} d(a_i, a_j) \right)^{n-1}$, car chaque élément $d(a_i, a_j)$ apparaît autant de fois que l'on peut choisir k dans $[\![1, n+1]\!] \setminus \{i, j\}$. De ce qui précède, il résulte que

$$\left(\prod_{1 \leqslant i < j \leqslant n+1} d(a_i, a_j) \right)^{n-1} = \mathrm{P}_1 \ldots \mathrm{P}_{n+1} \leqslant d_n(\mathrm{E})^{\frac{n(n-1)(n+1)}{2}}$$

puis $\prod_{1 \leqslant i < j \leqslant n+1} d(a_i, a_j) \leqslant d_n(\mathrm{E})^{\frac{n(n+1)}{2}}$, soit $\delta_{n+1}(a_1, \ldots, a_{n+1}) \leqslant d_n(\mathrm{E})$. Ceci étant vrai pour tout $(a_1, \ldots, a_{n+1}) \in \mathrm{E}^{n+1}$, on en déduit par passage à la borne supérieure que

$$d_{n+1}(\mathrm{E}) \leqslant d_n(\mathrm{E}).$$

Conclusion. La suite $(d_n(\mathrm{E})_{n \geqslant 2}$ décroît. Comme elle est minorée par 0, elle converge.

3. On suppose que E est le segment $[m, n]$. Soit $(a, b, c) \in \mathrm{E}^3$. On peut supposer que m, a, b, c, n sont alignés dans cet ordre. On a, par

définition : $\delta_3(a,b,c) = (d(a,b)d(b,c)d(a,c))^{\frac{1}{3}}$. Puisque $b \in [a,c]$, on a $d(a,b)+d(b,c) = d(a,c)$. On en déduit que

$$d(a,b)d(b,c) = \frac{1}{4}\left[(d(a,c))^2 - (d(a,b)-d(b,c))^2\right] \leqslant \frac{1}{4}(d(a,c))^2,$$

avec égalité si b est le milieu de $[a,c]$. On obtient donc

$$\delta_3(a,b,c) \leqslant \left(\frac{1}{4}(d(a,c))^3\right)^{1/3} \leqslant \left(\frac{1}{4}(d(m,n))^3\right)^{1/3},$$

avec égalité si $a = m, c = n$ et b est le milieu de $[m,n]$. On obtient donc $d_3(\mathrm{E}) = \frac{1}{\sqrt[3]{4}}d(m,n) = \frac{1}{\sqrt[3]{4}}\mathrm{long}(\mathrm{E})$, où $\mathrm{long}(\mathrm{E})$ est la longueur du segment E. ◁

La notion de diamètre transfini est reliée aux polynômes. Identifions $\mathbb{R}^2$ avec le plan complexe et notons $\mathcal{U}_n$ l'ensemble des polynômes unitaires de degré n de $\mathbb{C}[\mathrm{X}]$. Soit $\mathrm{E} \subset \mathbb{C}$ bornée et infinie. Pour $\mathrm{P} \in \mathbb{C}[\mathrm{X}]$ on pose $\|\mathrm{P}\|_{\mathrm{E}} = \sup_{z\in\mathrm{E}} |\mathrm{P}(z)|$ (c'est une norme) et $\mu_n(\mathrm{E}) = \inf_{\mathrm{P}\in\mathcal{U}_n} \|\mathrm{P}\|_{\mathrm{E}}$. On peut montrer que la suite $(\mu_n(\mathrm{E}))_{n\geqslant 0}$ vérifie $\mu_{n+m}(\mathrm{E}) \leqslant \mu_n(\mathrm{E})\mu_m(\mathrm{E})$ pour tout couple (n,m) ce qui implique la convergence de la suite $\mu_n(\mathrm{E})^{1/n}$ (voir l'exercice 2.3 du tome 1 d'analyse). On montre ensuite, à l'aide du déterminant de Vandermonde, que sa limite $\mu(\mathrm{E})$ n'est autre que le diamètre transfini de la partie E. A l'aide de cela on peut approfondir la question 3 de l'exercice et montrer (en faisant intervenir les polynômes de Tchebychev) que le diamètre transfini d'un segment de longueur a est égal à $a/4$. Le lecteur pourra aussi s'amuser à calculer (directement) le diamètre transfini d'un cercle de rayon 1 (il vaut 1). Indiquons que la seconde épreuve du concours commun des Mines de 1993 (option M) contient l'essentiel des résultats de cette note.

Les quatre exercices suivants étudient des problèmes d'extrema.

4.30. Deux problèmes d'extrema

Soit ABC un triangle du plan et M un point à l'intérieur du triangle (au sens large, en comptant les trois segments).

1. Déterminer les points M tels que le produit des trois distances de M aux côtés du triangle soit extrémal?

2. Même question avec la somme des distances aux côtés du triangle.

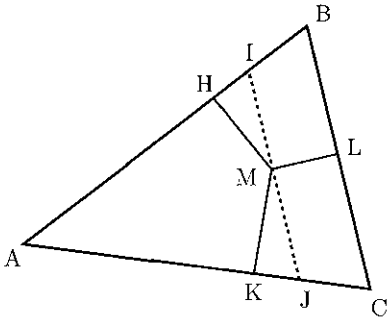
(**École polytechnique**)

▷ **Solution.**

1. L'intérieur du triangle est un compact. En effet, c"est l'image par l'application continue $(\lambda, \mu) \longmapsto \lambda A + \mu B + (1 - \lambda - \mu)C$ du compact $\{(\lambda, \mu) \in \mathbb{R}_+^2,\ \lambda + \mu \leqslant 1\}$. La fonction f qui à M associe le produit des trois distances est continue. Elle possède donc un minimum et un maximum sur ce compact.

Son minimum est clairement égal à 0 et est atteint en tout point situé sur l'un des côtés du triangle.

Appelons H, K, L les projetés orthogonaux de M sur (AB), (CA), (BC). On a donc $f(M) = MH \cdot MK \cdot ML$. Supposons que ML soit fixé, c'est-à-dire que M varie sur un segment [IJ] parallèle à (BC) ($I \in [AB]$, $J \in [AC]$).



Les angles $\widehat{HMI}$ et $\widehat{KMJ}$ sont constants quand M décrit [IJ]. On a donc $f(M) = \cos\widehat{HMI} \cdot MI \cdot \cos\widehat{KMJ} \cdot MJ \cdot ML$ qui est proportionnel au produit $MI \cdot MJ$. Comme $MI + MJ = IJ$, ce produit est maximum quand $MI = MJ$, c'est-à-dire quand M est le milieu de [IJ]. La droite (AM) est alors une médiane du triangle ABC. Si le maximum de f est atteint en M, on montre par le même raisonnement que M appartient aux trois médianes du triangle. C'est donc le centre de gravité G du triangle.

Le raisonnement précédent montre que f ne possède pas d'autre extremum local que ses extrema globaux. Si M_0 est un point situé strictement à l'intérieur du triangle et différent de G, il n'appartient pas aux trois médianes du triangle. Si par exemple M_0 n'appartient pas à la médiane relative au point A, on fait varier M sur un segment parallèle au côté (BC). Dans un voisinage de M_0, f est monotone, donc f ne possède pas d'extremum local en M_0.

2. On reprend les notations de la question précédente et on pose $g(M) = MH + MK + ML$. Quand ML est fixé, on obtient

$$g(M) = \cos\widehat{HMI} \cdot MJ + \cos\widehat{KMJ} \cdot MJ + ML.$$

Comme MI + MJ = IJ, cette somme est une fonction affine de MI. Une fonction affine est monotone donc les extrema de $g(\mathrm{M})$ pour $\mathrm{M} \in [\mathrm{IJ}]$ sont atteints en I et en J donc sur l'un des côtés du triangle. Mais le même raisonnement montre que quand M décrit un côté du triangle, un extremum est atteint en un sommet du triangle. Comme $g(\mathrm{A})$, $g(\mathrm{B})$ et $g(\mathrm{C})$ sont égaux respectivement aux trois hauteurs du triangle, c'est-à-dire à $\frac{2\mathrm{S}}{\mathrm{BC}}$, $\frac{2\mathrm{S}}{\mathrm{CA}}$ et $\frac{2\mathrm{S}}{\mathrm{AB}}$, où S est l'aire du triangle, le maximum (resp. le minimum) de g est obtenu au point opposé au plus petit (resp. plus grand) côté du triangle.

Si le triangle n'est pas isocèle, le maximum et le minimum sont obtenu en un unique point.

Supposons que le triangle soit isocèle en A, mais pas équilatéral. Avec les notations précédentes, on a

$$g(\mathrm{M}) = \cos\widehat{\mathrm{HMI}} \cdot \mathrm{MI} + \cos\widehat{\mathrm{KMJ}} \cdot \mathrm{MJ} + \mathrm{ML} = \cos\widehat{\mathrm{HMI}} \cdot \mathrm{IJ} + \mathrm{ML}$$

et g est constante sur chaque parallèle à (BC). On en déduit que les extrema de g sont obtenus en A d'une part et en tout point du segment [BC] d'autre part.

Si le triangle est équilatéral. la fonction g ne varie pas si on se déplace sur une parallèle à un quelconque des côtés. Elle est donc constante.

Le raisonnement précédent montre que g ne possède pas d'autres extrema locaux que ses extrema globaux. ◁

Voici une autre rédaction de la première question qui donne en outre la valeur du maximum de f. Avec les notations de la figure ci-dessus, on a $\mathrm{ML} \cdot \mathrm{BC} + \mathrm{MK} \cdot \mathrm{AC} + \mathrm{MH} \cdot \mathrm{AB} = 2\mathrm{S}$ *où* S *désigne l'aire du triangle* ABC. *L'inégalité arithmético-géométrique permet de dire que*

$$\mathrm{AB} \cdot \mathrm{AC} \cdot \mathrm{BC} \cdot f(\mathrm{M}) \leqslant \left(\frac{2\mathrm{S}}{3}\right)^3$$

donc que $f(\mathrm{M}) \leqslant \dfrac{8\mathrm{S}^3}{27abc}$ *avec* $a = \mathrm{BC}$, $b = \mathrm{AC}$ *et* $c = \mathrm{AB}$. *On a égalité lorsque* $a\mathrm{ML} = b\mathrm{MK} = c\mathrm{MH}$ *et il est facile de voir que cela défini le centre de gravité du triangle.*

Dans l'exercice 4.34 sera étudiée la fonction qui à M *associe la somme des distances* AM + BM + CM. *En attendant, voici un autre problème d'extremum où on cherche à maximiser un rapport d'aire.*

4.31. Problème d'extremum

1. Soit $\mathrm{E} = \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n,\ \sum_{i=1}^{n} x_i = 1\}$. Déterminer

$$\sup_{(x_1,\ldots,x_n)\in \mathrm{E}} \frac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} (1 - x_i)}.$$

2. Application : dans le plan euclidien, on considère un triangle ABC et des points $\mathrm{A}' \in [\mathrm{B}, \mathrm{C}]$, $\mathrm{B}' \in [\mathrm{C}, \mathrm{A}]$, $\mathrm{C}' \in [\mathrm{A}, \mathrm{B}]$ tels que les droites (AA'), (BB') et (CC') soient concourantes en O. Comment faut-il choisir O pour que le rapport des aires des triangles A'B'C' et ABC soit maximal.

(École polytechnique)

▷ **Solution.**

1. La fonction $f : (x_1, \ldots, x_n) \longmapsto \dfrac{\prod_{i=1}^{n} x_i}{\prod_{i=1}^{n} (1 - x_i)}$ n'est pas définie sur E tout entier, mais uniquement sur l'ensemble F des éléments $(x_1, \ldots, x_n)$ de E tels que $x_i \neq 1$, pour tout $i \in [\![1, n]\!]$. Pour $n = 1$, $\mathrm{F} = \emptyset$; pour $n = 2$, la fonction f est constante et vaut 1. Pour $n \geqslant 3$, il est tentant, étant donnée la symétrie du problème, de penser que le maximum est obtenu quand $x_1 = \cdots = x_n = \dfrac{1}{n}$. Posons $m = f\left(\dfrac{1}{n}, \ldots, \dfrac{1}{n}\right) = \dfrac{1}{(n-1)^n}$.

L'ensemble F n'étant pas compact, on ne peut pas affirmer directement que f est majorée sur F. On va se ramener au cas d'un compact en montrant que f prend des petites valeurs si l'un des x_i est proche de 1. Soit $\varepsilon \in]0, 1[$ et $x = (x_1, \ldots, x_n) \in \mathrm{F}$. Supposons qu'il existe $k \in [\![1, n]\!]$ tel que $x_k > 1 - \varepsilon$. Pour $i \neq k$, on a $x_i \leqslant 1 - x_k < \varepsilon$ et $1 - x_i \geqslant 1 - \varepsilon$. On en déduit que

$$f(x) \leqslant \frac{x_k(1 - x_k)^{n-1}}{(1 - x_k)(1 - \varepsilon)^{n-1}} \leqslant \frac{\varepsilon^{n-2}}{(1 - \varepsilon)^{n-1}}.$$

Comme cette quantité tend vers 0 avec ε, on peut choisir ε tel que

$$\frac{\varepsilon^{n-2}}{(1 - \varepsilon)^{n-1}} \leqslant \frac{m}{2}.$$

Considérons $\mathrm{G} = \{(x_1, \ldots, x_n) \in \mathrm{F},\ \forall i \in [\![1, n]\!],\ x_i \leqslant 1 - \varepsilon\}$. L'ensemble G est clairement fermé et borné. Il est donc compact et f possède un

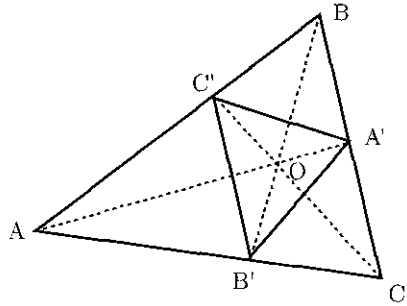
maximum sur G. Comme $f(x) \leqslant \dfrac{m}{2}$ pour tout $x \in \mathrm{F} \setminus \mathrm{G}$ et $m \in f(\mathrm{G})$, on a $\sup\limits_{\mathrm{G}} f = \sup\limits_{\mathrm{F}} f$. Ainsi f possède une borne supérieure sur F qui est atteinte en un point de G.

Soit $a = (a_1, \ldots, a_n) \in \mathrm{G}$ tel que $\sup\limits_{\mathrm{G}} f = f(a)$. Montrons que l'on a $a_1 = \cdots = a_n$ en raisonnant par l'absurde. Supposons qu'il existe deux indices i et j tels que $a_i \neq a_j$. Considérons $x = (x_1, \ldots, x_n) \in \mathrm{G}$ tel que $x_k = a_k$ si $k \neq i, j$. Alors x_i et x_j varient avec $x_i + x_j = a_i + a_j = s$, où $s \in]0, 1[$. On a alors

$$\begin{aligned}\frac{f(x)}{f(a)} &= \frac{x_i x_j}{(1-x_i)(1-x_j)} : \frac{a_i a_j}{(1-a_i)(1-a_j)} \\ &= \frac{x_i(s-x_i)}{(1-x_i)(1-s+x_i)} : \frac{a_i(s-a_i)}{(1-a_i)(1-s+a_i)} = \frac{g(x_i)}{g(a_i)},\end{aligned}$$

où g est définie sur $[0, s]$ par $g(z) = \dfrac{z(s-z)}{(1-z)(1-s+z)}$. Le calcul de g' montre que g atteint son maximum en $\dfrac{s}{2}$. Ainsi $g\left(\dfrac{s}{2}\right) > g(a_i)$ et on peut trouver $x \in \mathrm{G}$ tel que $f(x) > f(a)$, ce qui contredit la définition de a. Ainsi $a_1 = \cdots = a_n = \dfrac{1}{n}$ et $\sup\limits_{\mathrm{E}} f = \dfrac{1}{(n-1)^n}$.

2. On sait que si on considère le déterminant dans une base orthonormale, l'aire du triangle ABC est égale à $\dfrac{1}{2}|\det(\overrightarrow{\mathrm{AB}}, \overrightarrow{\mathrm{AC}})|$. On a le même résultat pour l'aire de A'B'C' et il faut donc exprimer $\det(\overrightarrow{\mathrm{A'B'}}, \overrightarrow{\mathrm{A'C'}})$ en fonction de $\det(\overrightarrow{\mathrm{AB}}, \overrightarrow{\mathrm{AC}})$.



Le point O est à l'intérieur du triangle, strictement (si O est sur l'un des côtés du triangle, les points A', B' et C' sont alignés et A'B'C' n'est pas un triangle). On note (α, β, γ) ses coordonnées barycentriques relativement à (A, B, C). Ce sont des réels strictement positifs tels que $\alpha + \beta + \gamma = 1$.

Montrons qu'alors A' est le barycentre de $((\mathrm{B}, \beta), (\mathrm{C}, \gamma))$. En effet si on note A'' ce point, par associativité du barycentre, O est le barycentre de $(\mathrm{A}, \alpha), (\mathrm{A''}, \beta+\gamma))$ et A'' appartient à (BC) et (OA). C'est A'. On montre

de même que B′ et C′ sont les barycentres respectifs de $((A,\alpha),(C,\gamma))$ et $((A,\alpha),(B,\beta))$. On en déduit que

$$\overrightarrow{AA'} = \frac{1}{\beta+\gamma}(\beta\overrightarrow{AB} + \gamma\overrightarrow{AC}),\quad \overrightarrow{AB'} = \frac{\gamma}{\alpha+\gamma}\overrightarrow{AC} \text{ et } \overrightarrow{AC'} = \frac{\beta}{\alpha+\beta}\overrightarrow{AB},$$

puis

$$\begin{aligned}\overrightarrow{A'B'} &= -\frac{\beta}{\beta+\gamma}\overrightarrow{AB} + \left(\frac{\gamma}{\alpha+\gamma} - \frac{\gamma}{\beta+\gamma}\right)\overrightarrow{AC}\\ &= -\frac{\beta}{\beta+\gamma}\overrightarrow{AB} + \frac{\gamma(\beta-\alpha)}{(\alpha+\gamma)(\beta+\gamma)}\overrightarrow{AC}\end{aligned}$$

et de même

$$\overrightarrow{A'C'} = \frac{\beta(\gamma-\alpha)}{(\alpha+\beta)(\beta+\gamma)}\overrightarrow{AB} - \frac{\gamma}{\beta+\gamma}\overrightarrow{AC}.$$

On a donc

$$\begin{aligned}\frac{\det(\overrightarrow{A'B'},\overrightarrow{A'C'})}{\det(\overrightarrow{AB},\overrightarrow{AC})} &= \begin{vmatrix} -\dfrac{\beta}{\beta+\gamma} & \dfrac{\beta(\gamma-\alpha)}{(\alpha+\beta)(\beta+\gamma)} \\ \dfrac{\gamma(\beta-\alpha)}{(\alpha+\gamma)(\beta+\gamma)} & -\dfrac{\gamma}{\beta+\gamma}\end{vmatrix}\\ &= \frac{2\alpha\beta\gamma}{(\beta+\gamma)(\alpha+\gamma)(\alpha+\beta)}.\end{aligned}$$

Le rapport entre les aires des triangles A′B′C′ et ABC est donc $\frac{2\alpha\beta\gamma}{(\beta+\gamma)(\alpha+\gamma)(\alpha+\beta)} = \frac{2\alpha\beta\gamma}{(1-\alpha)(1-\beta)(1-\gamma)}$. D'après la question 1, il est maximal lorsque $\alpha=\beta=\gamma=\frac{1}{3}$, c'est-à-dire lorsque O est le centre de gravité du triangle. Le rapport vaut alors $2\left(\frac{1}{3-1}\right)^3 = \frac{1}{4}$. ◁

Le question suivante a été posée par Gianfranco Di Fagnano en 1775 et l'idée essentielle de la solution semble due à Fejér vers 1900.

4.32. Triangle de périmètre minimal inscrit dans un triangle

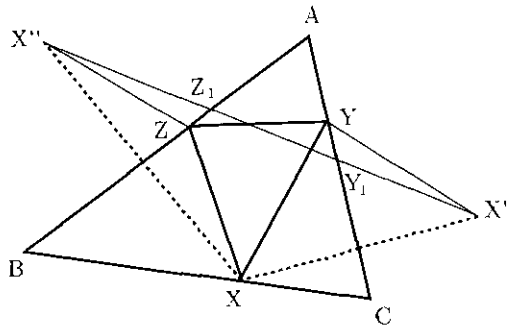
Soit ABC un triangle dont tous les trois angles sont aigus. Soit $X \in [A,B]$, $Y \in [B,C]$ et $Z \in [C,A]$. Montrer qu'il existe une configuration telle que le périmètre du triangle XYZ soit minimal et la caractériser.

(École polytechnique, École normale supérieure)

▷ **Solution.**

L'ensemble $E = [B,C]\times[C,A]\times[A,B]$ étant compact et la fonction $f : (X,Y,Z) \longmapsto XY+YZ+ZX$ étant continue, elle possède un minimum

sur E. Soit $(X, Y, Z) \in E$, X' et X'' les symétriques respectifs de X par rapport aux droites (AC) et (AB).



On a

$$f(X, Y, Z) = X'Y + YZ + ZX'' \geqslant X'X'',$$

avec égalité si, et seulement si, X', Y, Z et X'' sont alignés dans cet ordre. Fixons le point X. Le point X'' est l'image de X' dans la rotation de centre A et d'angle $2(\overrightarrow{AC}, \overrightarrow{AB})$. On a donc $\widehat{X'AX''} = 2\widehat{CAB} < \pi$, car l'angle $\widehat{CAB}$ est aigu. Le segment $[X'X'']$ coupe donc les demi-droites [AC) et [AB) en Y_1 et Z_1 respectivement, avec X', Y_1, Z_1 et X'' alignés dans cet ordre. Comme $\widehat{XBX''} = 2\widehat{XBA} < \pi$ et $\widehat{XCX'} = 2\widehat{XCA} < \pi$, les points X' et X'' appartiennent au demi-plan limité par (BC) contenant A. Ce demi-plan étant convexe, il contient Y_1 et Z_1 qui appartiennent respectivement aux demi-droites [C, A) et [B, A). Finalement, on obtient $Y_1 \in [A, C]$ et $Z_1 \in [AB]$. De ce qui précède on peut conclure

$$f(X, Y, Z) \geqslant f(X, Y_1, Z_1) = X'X''.$$

On a $AX' = AX = AX''$. Le triangle $AX'X''$ étant isocèle, on en déduit

$$X'X'' = 2AX \cdot \cos\widehat{AX'X''} = 2AX \cdot \cos\frac{\pi - \widehat{X'AX''}}{2} = 2AX \cdot \sin\widehat{CAB}.$$

Ainsi $f(X, Y_1, Z_1) = X'X''$ est minimal si AX est minimal, c'est-à-dire si X est le projeté orthogonal de A sur (BC) *i.e* le pied de la hauteur issue de A. Ce projeté appartient au segment [B, C] car les angles du triangle sont aigus.

Le minimum de f est donc atteint en un seul triplet (X_0, Y_0, Z_0) : X_0 est le projeté orthogonal de A sur (BC), Y_0 et Z_0 les points d'intersection de $X'X''$ avec (AC) et (AB) respectivement. Si au lieu de fixer X, on avait fixé Y (resp. Z), on aurait trouvé qu'au triplet (X_0, Y_0, Z_0) où f est minimum. Y_0 (resp. Z_0) est le projeté orthogonal de B sur (AC) (resp. de C sur (AB)). Le minimum étant atteint en un point unique,

on en déduit qu'il est atteint pour le triplet (X_0, Y_0, Z_0) formé des pieds des trois hauteurs. Le triangle $X_0Y_0Z_0$ est appelé *le triangle orthique* du triangle ABC. ◁

Montrons que les droites (AX_0), (BY_0) *et* (CZ_0) *sont les bissectrices intérieures du triangle* $X_0Y_0Z_0$. *Les quadrilatères* AY_0X_0B *et* AZ_0X_0C *sont inscriptibles dans des cercles de diamètre* [AB] *et* [AC] *respectivement. On a donc*

$$(X_0Y_0, X_0A) \equiv (BY_0, BA) \equiv \frac{\pi}{2} + (CA, BA) \pmod{\pi}$$

$$(X_0A, X_0Z_0) \equiv (CA, CZ_0) \equiv (CA, BA) + \frac{\pi}{2} \pmod{\pi}.$$

Le angles de droites (X_0Y_0, X_0A) *et* (X_0A, X_0Z_0) *sont égaux. Comme les demi-droites* $[X_0Y_0)$ *et* $[X_0Z_0)$ *sont dans le même demi-plan limité par la droite* (BC), *on a* $(\overrightarrow{X_0Y_0}, \overrightarrow{X_0A}) = (\overrightarrow{X_0A}, \overrightarrow{X_0Z_0})$ *et* (X_0A) *est bissectrice intérieure de l'angle* X_0 *du triangle* $X_0Y_0Z_0$. *La démonstration est la même pour les autres angles.*

Pour cette raison $X_0Y_0Z_0$ *est appelé une trajectoire de lumière. En effet, considérons trois miroirs* M_1, M_2 *et* M_3 *situés dans des plans perpendiculaires au plan* (ABC) *et coupant le plan* (ABC) *suivant les segments* [BC], [CA] *et* [AB]. *Un rayon lumineux issus de* Z_0 *et se dirigeant vers* X_0 *sera réfléchi sur* M_1, *selon la loi de réflexion de Descartes, dans la direction* (X_0Y_0) *; il touchera le miroir* M_2 *en* Y_0, *sera réfléchi dans la direction* (Y_0Z_0), *touchera le miroir* M_3 *en* Z_0, *sera réfléchi dans la direction* (Z_0X_0). *On obtient ainsi une trajectoire fermée.*

Si le triangle est rectangle ou obtus en A, *le triangle de périmètre minimal dégénère en un segment :* $Y_0 = Z_0 = A$ *et* X_0 *projeté orthogonal de* A *sur* (BC).

Même si sa solution utilise un peu d'analyse, nous avons choisi de placer ici l'énoncé suivant qui concerne aussi les trajectoires de lumière.

4.33. Billard convexe compact

Soit K un compact convexe de $\mathbb{R}^2$, dont la frontière $\delta(K)$ est une courbe fermée de classe C^∞. Montrer que, pour tout $n \geqslant 2$ on peut, en jouant au billard dans K, faire n rebonds et revenir au même point.

(École normale supérieure)

▷ **Solution.**

Quand la bille rebondit en un point M de la frontière, l'angle d'incidence est égal à l'angle de réflexion, c'est-à-dire que la normale à $\delta(K)$ en M est la bissectrice de l'angle formé par les trajectoires incidentes

et réfléchies : elle obéit à la loi de Descartes. C'est pourquoi une telle trajectoire est appelée une trajectoire de lumière.

Nous allons démontrer qu'un polygone à n côtés inscrits dans $\delta(\mathrm{K})$ de périmètre minimal est une trajectoire de lumière et donc répond à la question. L'existence d'un tel polygone est simple à établir. On paramètre la frontière $\delta(\mathrm{K})$ par $\mathrm{F} : \mathbb{R} \longmapsto \mathbb{R}^2$, $\mathcal{C}^\infty$, L-périodique (où L est la longueur de la courbe), injective sur $[0, \mathrm{L}[$ et vérifiant $\|\mathrm{F}'(t)\| = 1$ pour tout réel t. Soit $\mathcal{C} = \{(t_1, t_2, \ldots, t_n) \in \mathbb{R}^n, 0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n \leqslant \mathrm{L}\}$. L'ensemble $\mathcal{C}$ est compact. L'application Φ définie sur $\mathcal{C}$ par

$$\Phi(t_1, \ldots, t_n) = \sum_{k=1}^{n} \|\mathrm{F}(t_k) - \mathrm{F}(t_{k+1})\|,$$

où $t_{n+1} = t_1 + \mathrm{L}$ est continue sur $\mathcal{C}$ donc elle possède un maximum. Notons $(u_1, \ldots, u_n)$ un point où il est atteint et pour tout $k \in [\![1, n]\!]$, $\mathrm{M}_k = \mathrm{F}(u_k)$.

On peut supposer que les points M_k sont distincts. En effet, ils ne sont pas tous égaux et si $t_k < t_{k+1} = t_{k+2}$ $(1 \leqslant k \leqslant n-1)$, en prenant $t \in]t_k, t_{k+2}[$, on obtient

$$\begin{aligned}\|\mathrm{F}(t_k) - \mathrm{F}(t)\| + \|\mathrm{F}(t) - \mathrm{F}(t_{k+2})\| &\geqslant \|\mathrm{F}(t_k) - \mathrm{F}(t_{k+2})\| \\ &= \|\mathrm{F}(t_k) - \mathrm{F}(t_{k+1})\| + \|\mathrm{F}(t_{k+1}) - \mathrm{F}(t_{k+2})\|,\end{aligned}$$

donc on peut remplacer t_{k+1} par t. Par définition, pour tout $k \in [\![1, n-1]\!]$, la fonction f_k définie sur $[t_k, t_{k+2}]$ par

$$f_k(t) = \|\mathrm{F}(t_k) - \mathrm{F}(t)\| + \|\mathrm{F}(t) - \mathrm{F}(t_{k+2})\|$$

possède un maximum en t_{k+1}.

Montrons que la fonction f_k est dérivable sur $]t_k, t_{k+1}[$. Pour tout réel t_0, la fonction $f : t \longmapsto \|\mathrm{F}(t_0) - \mathrm{F}(t)\| = \sqrt{\|\mathrm{F}(t_0) - \mathrm{F}(t)\|^2}$ est dérivable en tout t non congru à t_0 modulo L et

$$f'(t) = \frac{\langle -\mathrm{F}'(t), \mathrm{F}(t_0) - \mathrm{F}(t)\rangle}{\|\mathrm{F}(t_0) - \mathrm{F}(t)\|}.$$

Pour $t \in]t_k, t_{k+2}[$, on obtient donc

$$\begin{aligned}f_k'(t_{k+1}) &= \frac{\langle -\mathrm{F}'(t_{k+1}), \mathrm{F}(t_k) - \mathrm{F}(t_{k+1})\rangle}{\|\mathrm{F}(t_k) - \mathrm{F}(t_{k+1})\|} + \frac{\langle -\mathrm{F}'(t_{k+1}), \mathrm{F}(t_{k+2}) - \mathrm{F}(t_{k+1})\rangle}{\|\mathrm{F}(t_{k+2}) - \mathrm{F}(t_{k+1})\|} \\ &= \left\langle -\mathrm{F}'(t_{k+1}), \frac{\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_k}}{\mathrm{M}_{k+1}\mathrm{M}_k} + \frac{\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_{k+2}}}{\mathrm{M}_{k+1}\mathrm{M}_{k+2}} \right\rangle = 0.\end{aligned}$$

Les vecteurs $u_k = \dfrac{\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_k}}{\mathrm{M}_{k+1}\mathrm{M}_k}$ et $v_k = \dfrac{\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_{k+2}}}{\mathrm{M}_{k+1}\mathrm{M}_{k+2}}$ sont unitaires et colinéaires à $\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_k}$ et $\overrightarrow{\mathrm{M}_{k+1}\mathrm{M}_{k+2}}$ respectivement.

Le vecteur $u_k + v_k$ est un vecteur directeur de la bissectrice de $(\overrightarrow{M_{k+1}M_k}, \overrightarrow{M_{k+1}M_{k+2}})$. Il est orthogonal à $F'(t_{k+1})$ donc dirige la normale à la courbe en M_{k+1}. On en déduit qu'une trajectoire allant de M_k à M_{k+1} repart après rebond de M_{k+1} vers M_{k+2}. La trajectoire $M_1M_2 \dots M_nM_1$ a les propriétés voulues. ◁

L'exercice précédent montre le lien entre trajectoire de lumière et périmètre minimal. La même démarche peut être utilisée pour résoudre l'exercice 4.32 dont nous reprenons les notations. Supposons que le minimum soit obtenu en (X_0, Y_0, Z_0) *et qu'aucun de ces points ne soit un sommet du triangle* ABC.

Considérons l'application $s \longmapsto X(s) = X_0 + s\overrightarrow{BC}$. *Pour* $|s|$ *assez petit,* $X(s)$ *appartient à* $]B, C[$. *La fonction* $\varphi : s \longmapsto X(s)Y_0 + X(s)Z_0 + Y_0Z_0$ *possède un minimum en* 0. *Comme le gradient de la fonction* $M \longmapsto Y_0M$ *(resp.* $M \longmapsto Z_0M$*) en* X_0, *point distinct de* Y_0 *(resp.* Z_0*) est* $u = \dfrac{\overrightarrow{Y_0X_0}}{Y_0X_0}$ *(resp.* $v = \dfrac{\overrightarrow{Z_0X_0}}{Z_0X_0}$*), on obtient*

$$\varphi'(0) = (u + v, X'(0)) = \langle u + v, \overrightarrow{BC}\rangle = 0.$$

Le vecteur $u + v$ *dirige la bissectrice intérieure de l'angle* $\widehat{Y_0X_0Z_0}$ *donc* (BC) *est bissectrice extérieure de l'angle* $\widehat{Y_0X_0Z_0}$. *On en déduit que l'on a* $\widehat{BX_0Z_0} = \widehat{Y_0X_0C}$. *On peut faire le même raisonnement pour les trois sommets ce qui montre que* $X_0Y_0Z_0$ *est une trajectoire de lumière. À partir de là, on retrouve les résultats obtenus géométriquement. Les droites* (AB), (BC) *et* (CA) *sont les bissectrices extérieures du triangle* $X_0Y_0Z_0$, *donc* A, B *et* C *sont les centres des cercles exinscrits. On en déduit que* (AX_0), (BY_0) *et* (CZ_0) *sont les bissectrices intérieures du triangle* $X_0Y_0Z_0$. *On a enfin* $(AX_0) \perp (BC)$ *et de même* $(BY_0) \perp (CA)$ *et* $(CZ_0) \perp (AB)$ *: le triangle* $X_0Y_0Z_0$ *est le triangle orthique du triangle* ABC.

Pour être complet, il suffirait de démontrer que dans le cas où les angles du triangle sont aigus, tout triangle XYZ *contenant un sommet donne un périmètre plus grand.*

La question de l'exercice suivant a été posée par Fermat et résolue par Torricelli.

4.34. Problème de Fermat

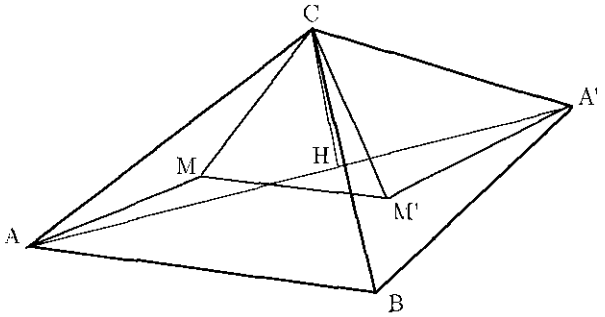
Soit ABC un triangle et M un point à l'intérieur du triangle (au sens large, en comptant les trois segments). Déterminer les points M tels que la somme MA + MB + MC soit minimale.

(École polytechnique)

▷ **Solution.**

L'intérieur du triangle ABC étant un compact, la fonction continue $f : M \longmapsto MA + MB + MC$ y possède un minimum. Il y a de nombreuses manières d'aborder ce problème classique (utilisation des nombres complexes ou du calcul différentiel par exemple) et nous avons choisi de présenter ici une solution assez géométrique.

On oriente le plan de telle manière que le triangle ABC soit direct et on considère le point A′ tel que le triangle A′BC soit équilatéral et extérieur au triangle ABC ; il est de sens indirect.



Considérons la rotation r de centre C et d'angle $\frac{\pi}{3}$. On a $r(B) = A'$. Pour M donné on note $M' = r(M)$. Comme r est une isométrie, on a $MB = M'A'$ et comme le triangle CMM′ est équilatéral, $MC = MM'$. On peut donc écrire

$$f(M) = AM + MM' + M'A' \geqslant AA'.$$

par l'inégalité triangulaire, avec égalité si les points A, M, M′ et A′ sont alignés dans cet ordre. Pour cela il faut prendre M et M′ sur le segment [AA′] de telle façon que CMM′ soit équilatéral direct. Si H est le projeté orthogonal de C sur (AA′), M et M′ sont symétriques par rapport à (CH) et les angles $(\overrightarrow{CM}, \overrightarrow{CH})$ et $(\overrightarrow{CH}, \overrightarrow{CM'})$ ont pour mesure $\frac{\pi}{6}$.

Notons T et T′ les uniques points M et M′ vérifiant cela (voir la figure suivante plus bas). Le point T est appelé *point de Torricelli* du triangle ABC. Il reste à voir si les points A, T, T′, A′ sont alignés dans cet ordre et si T est à l'intérieur du triangle. Pour que le segment [AA′] coupe [BC], il faut que les angles $(\overrightarrow{CA}, \overrightarrow{CA'})$ et $(\overrightarrow{BA'}, \overrightarrow{BA})$ aient une mesure comprise entre 0 et π. Comme les angles du triangle BA′C ont pour mesure $\frac{\pi}{3}$, il faut donc que les angles $\widehat{BCA}$ et $\widehat{CBA}$ soient inférieurs à $\frac{2\pi}{3}$. On suppose cette condition réalisée.

Comme CTT′ est équilatéral, on a

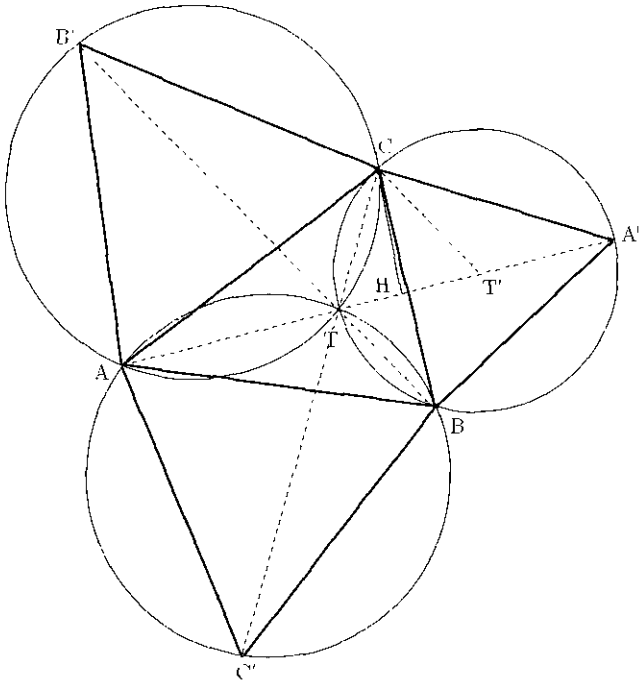
$$(\widehat{\overrightarrow{TA'}, \overrightarrow{TC}}) \equiv (\widehat{\overrightarrow{TT'}, \overrightarrow{TC}}) \equiv \frac{\pi}{3} \equiv (\widehat{\overrightarrow{BA'}, \overrightarrow{BC}}) \pmod{\pi}$$

donc les points B, T, C, A′ sont cocycliques. De plus T et A′ ne sont pas sur le même arc BC donc $\widehat{BTC} = \frac{2\pi}{3}$. Le point T est à l'intérieur du triangle ABC si, et seulement si, $\widehat{BAC} \leqslant \frac{2\pi}{3}$.

Si ces conditions sont réalisées, on a alors

$$\widehat{ACT'} = \widehat{ACT} + \frac{\pi}{3} \leqslant \widehat{ACB} + \frac{\pi}{3} \leqslant \widehat{ACA'}$$

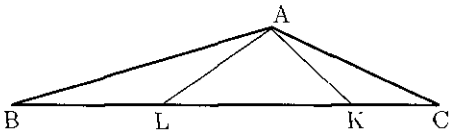
donc $T' \in [AA']$. Les points A, T, T′ et A′ sont alignés dans cet ordre. *donc la fonction f est minimale en T.*



Conclusion. Si les trois angles du triangle sont inférieurs ou égaux à $\frac{2\pi}{3}$, le minimum de f est atteint en T et vaut AA′. Pour des raisons de symétrie, si on construit les points B′ et C′ extérieurs au triangle tels que ACB′ et BC′A soient équilatéraux, on trouve que T appartient aux segments [BB′] et [CC′] et que le minimum de f vaut BB′ et CC′. Ainsi les droites (AA′), (BB′) et (CC′) sont concourantes en T et on a, de plus, AA′ = BB′ = CC′. Le point T est aussi le point d'intersection des cercles circonscrits aux triangles BCA′, ABC′ et ACB′.

Notons que dans le cas où $\widehat{BAC} = \dfrac{2\pi}{3}$, alors T = A.

Il reste à étudier le cas où l'un des angles du triangle est strictement supérieur à $\dfrac{2\pi}{3}$. Supposons que ce soit au point A. Montrons qu'alors le minimum de f est encore atteint en A. Pour cela, considérons les points K et L du segment [BC] tels que $\widehat{BAK} = \widehat{LAC} = \dfrac{2\pi}{3}$.



Soit M un point distinct de A à l'intérieur du triangle ABC.

• Si M est à l'intérieur du triangle ABK, l'étude du cas où l'angle $\widehat{BAC}$ vaut $\dfrac{2\pi}{3}$, appliquée au triangle ABK montre que

$$MA + MB + MK > AB + AK.$$

On en déduit que

$$f(M) = MA + MB + MC > AB + AK + MC - MK.$$

On veut montrer que $f(M) > AB + AC$. Il suffit de vérifier pour cela que

$$MC + AK \geqslant AC + MK,$$

ce qui est évident, car dans le quadrilatère convexe MKCA la somme des longueurs des diagonales est strictement supérieure à la somme des longueurs de deux côtés opposés (introduire le point d'intersection des diagonales pour le voir).

• Si M est à l'intérieur du triangle ALC, on fait le même raisonnement en échangeant les rôles de K et L et de B et C.

Conclusion. Lorsqu'un des angles du triangle est supérieur à $\dfrac{2\pi}{3}$, le minimum de f est obtenu au sommet correspondant. ◁

Comme pour le recherche du triangle de périmètre minimal, on peut donner une solution faisant appel au calcul différentiel. L'application $f : M \longmapsto MA + MB + MC$ *est différentiable sur* $\mathbb{R}^2 \setminus \{A, B, C\}$ *et son gradient vaut* $\dfrac{\overrightarrow{AM}}{AM} + \dfrac{\overrightarrow{BM}}{BM} + \dfrac{\overrightarrow{CM}}{CM}$. *Si donc le minimum de* f *n'est pas obtenu en un des sommets du triangle, il est obtenu en un point* M *qui vérifie* $\dfrac{\overrightarrow{AM}}{AM} + \dfrac{\overrightarrow{BM}}{BM} + \dfrac{\overrightarrow{CM}}{CM} = 0$. *Considérons des points* O, I, J *et* K *tels que* $\overrightarrow{OI} = \dfrac{\overrightarrow{AM}}{AM}$, $\overrightarrow{OJ} = \dfrac{\overrightarrow{BM}}{BM}$ *et* $\overrightarrow{OK} = \dfrac{\overrightarrow{CM}}{CM}$. *On a alors* $OI = OJ = OK = 1$ *et* $\overrightarrow{OI} + \overrightarrow{OJ} + \overrightarrow{OK} = 0$ *: O est le centre de gravité et le centre du cercle*

circonscrit du triangle IJK, *donc ce triangle est équilatéral. On a donc* $\widehat{\mathrm{IOJ}} = \widehat{\mathrm{JOK}} = \widehat{\mathrm{KOI}} = \dfrac{2\pi}{3}$. *Les droites* (MA), (MB) *et* (MC) *font donc nécessairement entre elles des angles de* $\dfrac{2\pi}{3}$. *On peut en déduire que* M *est le point de Torricelli du triangle* ABC. *Le minimum est donc obtenu soit en* T *soit en un sommet du triangle* ABC.

Le prochain thème abordé est celui des applications affines. Rappelons rapidement l'essentiel sur cette notion dans le cadre vectoriel. Si E *est un espace vectoriel, une application* $f : \mathrm{E} \longrightarrow \mathrm{E}$ *est dite affine si elle est la composée* $f = t_u \circ g$ *d'une translation* $t_u : x \longmapsto x+u$ *et d'une application linéaire* g. *On a alors* $u = f(0)$ *de sorte que cette décomposition est unique. L'application* g *est appelée la partie linéaire de* f *et sera notée* $\vec{f}$. *Il est très simple de vérifier qu'une application affine conserve les barycentres (la réciproque est vraie), transforme un sous-espace affine en un sous-espace affine, etc. Par ailleurs* f *est injective (resp. surjective, bijective) si, et seulement si, sa partie linéaire* $\vec{f}$ *est injective (resp. surjective, bijective).*

Le premier exercice ci-après fait intervenir des symétries affines.

4.35. Polygone dont les milieux des côtés sont donnés

On considère n points $\mathrm{A}_1, \ldots, \mathrm{A}_n$ de $\mathbb{R}^p$. À quelle condition peut-on construire, une ligne brisée fermée $\mathrm{M}_1, \ldots, \mathrm{M}_n$, $\mathrm{M}_{n+1} = \mathrm{M}_1$ telle que le point A_i soit le milieu de $[\mathrm{M}_i, \mathrm{M}_{i+1}]$ pour tout $i \in [\![1, n]\!]$? Comment en construire une ? Comment traduire la condition d'existence trouvée dans le cas n pair ?

(École polytechnique)

▷ **Solution.**

On note S_A la symétrie par rapport au point A. Le point M_1 étant donné, on a nécessairement

$$\mathrm{M}_2 = s_{\mathrm{A}_1}(\mathrm{M}_1),\ \mathrm{M}_3 = s_{\mathrm{A}_2} \circ s_{\mathrm{A}_1}(\mathrm{M}_1), \ldots,\ \mathrm{M}_n = s_{\mathrm{A}_{n-1}} \circ \cdots \circ s_{\mathrm{A}_1}(\mathrm{M}_1)$$

et les points $\mathrm{M}_1, \ldots, \mathrm{M}_n$ répondent au problème posé si, et seulement si,

$$\mathrm{M}_1 = s_{\mathrm{A}_n}(\mathrm{M}_n) = s_{\mathrm{A}_n} \circ s_{\mathrm{A}_{n-1}} \circ \cdots \circ s_{\mathrm{A}_1}(\mathrm{M}_1).$$

Notons $f = s_{\mathrm{A}_n} \circ s_{\mathrm{A}_{n-1}} \circ \cdots \circ s_{\mathrm{A}_1}$.

• Si n est impair, f est une symétrie centrale, qui possède un unique point invariant, son centre. On doit prendre ce centre comme point M_1. Le problème posé possède une solution unique, quels que soient les points $\mathrm{A}_1, \mathrm{A}_2, \ldots, \mathrm{A}_n$.

• Si n est pair, f est une translation. Notons u le vecteur de cette translation. Si u n'est pas nul, l'application f ne possède pas de point invariant et le problème posé n'a pas de solution. Si $u = 0$, f est l'identité. La condition sur M_1 est toujours réalisée ; on peut choisir M_1 quelconque et le problème posé possède une infinité de solutions.

Déterminons le vecteur u. Étant donnés deux points A et B de l'espace affine, on montre, en cherchant l'image de A, que $s_B \circ s_A$ est la translation de vecteur $2\overrightarrow{AB}$. On en déduit que

$$u = 2(\overrightarrow{A_1A_2} + \overrightarrow{A_3A_4} + \cdots + \overrightarrow{A_{n-1}A_n}).$$

La condition d'existence d'une solution dans le cas n pair s'écrit

$$\overrightarrow{A_1A_2} + \overrightarrow{A_3A_4} + \cdots + \overrightarrow{A_{n-1}A_n} = 0.$$

Elle signifie géométriquement que l'isobarycentre des points A_i d'indice pair est égal à l'isobarycentre des points d'indice impair. Par exemple, en dimension 2, quatre points donnés A_1, A_2, A_3, A_4 sont les milieux des côtés d'un quadrilatère si, et seulement si, $[A_1, A_3]$ et $[A_2, A_4]$ ont le même milieu, autrement dit si, et seulement si, (A_1, A_2, A_3, A_4) est un parallélogramme. ◁

Voici une autre approche. Il s'agit de déterminer des éléments $M_1, \ldots, M_n$ *de* $\mathbb{R}^p$ *tels que, pour tout* $i \in [\![1, n-1]\!]$, $M_i + M_{i+1} = 2A_i$ *et* $M_n + M_1 = 2A_n$. *On peut voir l'exercice comme la résolution d'un système linéaire. Notons* S_n *la matrice du système. Il est clair que* $\operatorname{rg}(S_n) \geqslant n-1$ *et en développant par rapport à la dernière ligne, on trouve* $\det S_n = 1 + (-1)^{n+1}$. *Si* n *est impair,* $\det S_n = 2$ *et le système a une solution unique. Si* n *est pair,* $\det S_n = 0$, *la matrice est de rang* $n-1$. *En additionnant les lignes du système avec des signes alternés, on trouve* $\sum_{i=1}^{n} (-1)^i A_i = 0$. *C'est la condition de compatibilité du système. Si elle est réalisée, on peut choisir* M_1 *arbitrairement.*

4.36. Triangles à côtés parallèles

Soient E un plan vectoriel réel, (A, B, C) et (A′, B′, C′) deux triangles tels que (AB) est parallèle à (A′B′), (BC) est parallèle à (B′C′), (AC) est parallèle à (A′C′). Montrer que les droites (AA′), (BB′) et (CC′) sont parallèles ou concourantes.

(École polytechnique)

▷ **Solution.**

Comme (A, B, C) est une base affine du plan, il existe une unique

application affine f de (ABC) dans (A'B'C') transformant A, B, C en A', B', C'. L'application linéaire associée $\overrightarrow{f}$ transforme $\overrightarrow{AB}$ en $\overrightarrow{A'B'}$, $\overrightarrow{AC}$ en $\overrightarrow{A'C'}$ et $\overrightarrow{CB}$ en $\overrightarrow{C'B'}$. Par hypothèse, il existe $(\lambda, \mu, \nu) \in \mathbb{R}^3$ tel que $\overrightarrow{f}\left(\overrightarrow{AB}\right) = \lambda\overrightarrow{AB}$, $\overrightarrow{f}\left(\overrightarrow{AC}\right) = \mu\overrightarrow{AC}$ et $\overrightarrow{f}\left(\overrightarrow{CB}\right) = \nu\overrightarrow{CB}$. Par linéarité de $\overrightarrow{f}$, on obtient

$$\lambda\overrightarrow{AC} + \lambda\overrightarrow{CB} = \lambda\overrightarrow{AB} = \overrightarrow{f}\left(\overrightarrow{AB}\right) = f(\overrightarrow{AC} + \overrightarrow{CB}) = \mu\overrightarrow{AC} + \nu\overrightarrow{CB}.$$

On en déduit que $\lambda = \mu = \nu$ et $\overrightarrow{f}$ est une homothétie vectorielle de rapport λ. Ainsi f est soit une translation, soit une homothétie affine selon que λ est égal ou non à 1. Dans le premier cas les droites (AA'), (BB') et (CC') sont parallèles. Dans le second cas, elles sont concourantes au O centre de l'homothétie f. ◁

Depuis Felix Klein on perçoit les géométries comme l'étude de propriétés invariantes sous l'action d'un groupe. Regardons l'exemple simple du triangle dans un plan pour illustrer cette idée profonde. Un sous-groupe G *du groupe affine de* $\mathbb{R}^2$ *opère naturellement sur l'ensemble des triangles : deux triangles* ABC *et* A'B'C' *sont dans la même orbite s'il existe* $f \in$ G *tel que* f(A) = A', f(B) = B' *et* f(C) = C'. *Classifier les triangles dans diverses géométries, c'est-à-dire pour divers groupes* G, *revient à donner des invariants qui caractérisent les classes d'équivalence. Par exemple si* G *est le groupe affine tout entier il n'y a qu'une seule orbite : en géométrie affine tous les triangles se valent. Si* G *est le sous-groupe des similitudes, on vérifie aisément que deux triangles sont semblables si, et seulement si, leurs trois côtés sont proportionnels. Si* G *est le sous-groupe des isométries, on prouve que deux triangles sont dans la même orbite si, et seulement si, leurs trois côtés sont égaux ou encore si, et seulement si, ils ont un angle égal entre deux côtés égaux. Enfin si* G *est le sous-groupe affine unimodulaire (*i.e. *le groupe des applications affines dont la partie linéaire est de déterminant* 1*) on voit que deux triangles sont dans la même classe si et seulement si ils ont la même aire (algébrique).*

L'exercice suivant, dans le cadre de la géométrie affine, commence l'étude pour les petits degrés de la classification des graphes des polynômes.

4.37. Équivalence entre parties de $\mathbb{R}^2$

Si C et C′ sont deux parties de $\mathbb{R}^2$, on dit que C et C′ sont équivalentes si, et seulement si, il existe un automorphisme affine f de $\mathbb{R}^2$ tel que $f(\mathrm{C}) = \mathrm{C}'$. Soient P, Q deux polynômes réels de degrés respectifs m et n et C (resp. C′) le graphe de P (resp. Q). Étudier l'équivalence de C et C′ lorsque (m, n) vaut $(2,2)$, $(2,3)$, $(3,3)$ et $(4,4)$.

(École normale supérieure)

▷ **Solution.**

L'équivalence ainsi définie entre des parties de $\mathbb{R}^2$ est évidemment un relation d'équivalence.

• Supposons que $m = 2$ et posons $\mathrm{P} = a\mathrm{X}^2 + b\mathrm{X} + c$ $(a \neq 0)$. Le graphe C de P est la parabole d'équation $y = a\left(x + \dfrac{b}{2a}\right)^2 - \dfrac{\Delta}{4a}$, soit

$$\frac{y}{a} + \frac{\Delta}{4a^2} = \left(x + \frac{b}{2a}\right)^2.$$

Considérons l'automorphisme affine $g : (x, y) \longmapsto \left(x + \dfrac{b}{2a}, \dfrac{y}{a} + \dfrac{\Delta}{4a^2}\right)$. L'image de C par g est la parabole $\Gamma : y = x^2$. Donc C est équivalente à Γ. Ainsi, si $(m, n) = (2,2)$, C et C′ sont équivalentes à Γ donc sont équivalentes par transitivité.

• Supposons que $m = 3$ et posons $\mathrm{P} = a\mathrm{X}^3 + b\mathrm{X}^2 + c\mathrm{X} + d$ $(a \neq 0)$. Il existe des réels e et f tels que $\mathrm{P} = a\left(\left(\mathrm{X} + \dfrac{b}{3a}\right)^3 + e\mathrm{X} + f\right)$. Le graphe C de P a pour équation $y = a\left(\left(x + \dfrac{b}{3a}\right)^3 + ex + f\right)$, soit

$$\frac{1}{a}y - ex - f = \left(x + \frac{b}{3a}\right)^3.$$

Considérons l'automorphisme affine

$$h : (x, y) \longmapsto \left(x + \frac{b}{3a}, -ex + \frac{1}{a}y - f\right).$$

L'image de C par h est la courbe Γ' d'équation $y = x^3$. Donc C est équivalente à Γ'. Si $(m, n) = (3,3)$, C et C′ sont équivalentes à Γ' donc sont équivalentes.

• Supposons que $(m, n) = (2,3)$. D'après ce qui précède, il suffit d'examiner le cas où $\mathrm{P} = \mathrm{X}^2$ et $\mathrm{Q} = \mathrm{X}^3$. Supposons qu'il existe un automorphisme affine f tel que $f(\Gamma) = \Gamma'$. Considérons la droite D d'équation

$y = x$. Elle coupe Γ' en trois points $(0,0)$, $(1,1)$ et $(-1,-1)$. On a $\Gamma' \cap \mathrm{D} = f(\Gamma) \cap \mathrm{D} = f(\Gamma \cap f^{-1}(\mathrm{D}))$. Comme f est bijective, $\Gamma \cap f^{-1}(\mathrm{D}))$ est de cardinal 3. Mais il est impossible que la droite affine $f^{-1}(\mathrm{D})$ coupe la parabole Γ en plus de deux points. On a une contradiction : C et C' ne sont pas équivalentes si $(m,n) = (2,3)$.

• Supposons que $m = 4$ et posons $\mathrm{P} = a\mathrm{X}^4 + b\mathrm{X}^3 + c\mathrm{X}^2 + d\mathrm{X} + e$ $(a \neq 0)$. Il existe des réels α, β et γ telles que

$$\mathrm{P} = a\left(\left(\mathrm{X} + \frac{b}{4a}\right)^4 + \alpha\left(\mathrm{X} + \frac{b}{4a}\right)^2 + \beta\mathrm{X} + \gamma\right).$$

Le graphe de P a pour équation

$$\frac{1}{a}y - \beta x - \gamma = \left(x + \frac{b}{4a}\right)^4 + \alpha\left(x + \frac{b}{4a}\right)^2.$$

On considère l'automorphisme affine

$$k : (x,y) \longmapsto \left(x + \frac{b}{4a}, \frac{1}{a}y - \beta x - \gamma\right).$$

L'image de C par k est la courbe d'équation $y = x^4 + \alpha x^2$.

Si $\alpha = 0$, C est équivalente à Γ_0 courbe d'équation $y = x^4$.

Si $\alpha > 0$, $y = x^4 + \alpha x^2$ équivaut à $\frac{1}{\alpha^2}y = \left(\frac{x}{\sqrt{\alpha}}\right)^4 + \left(\frac{x}{\sqrt{\alpha}}\right)^2$. Si ℓ est l'automorphisme affine

$$(x,y) \longmapsto \left(\frac{1}{\alpha^2}y, \frac{1}{\sqrt{\alpha}}x\right),$$

l'image de $k(\mathrm{C})$ par ℓ est la courbe Γ_1 d'équation $y = x^4 + x^2$, donc C est équivalente à Γ_1.

Si $\alpha < 0$, $y = x^4 + \alpha x^2$ équivaut à $\frac{1}{\alpha^2}y = \left(\frac{x}{\sqrt{-\alpha}}\right)^4 - \left(\frac{x}{\sqrt{-\alpha}}\right)^2$. Si ℓ' est l'automorphisme affine

$$(x,y) \longmapsto \left(\frac{1}{\alpha^2}y, \frac{1}{\sqrt{-\alpha}}x\right),$$

l'image de $k(\mathrm{C})$ par ℓ' est la courbe Γ_{-1} d'équation $y = x^4 - x^2$, donc C est équivalente à Γ_{-1}.

Montrons que les courbes Γ_0, Γ_1 et Γ_{-1} sont non équivalentes deux à deux. Soit $f : (x,y) \longmapsto (ax + by + c, a'x + b'y + c')$. L'image de Γ_0 par f^{-1} a pour équation

$$a'x + b'y + c' = (ax + by + c)^4.$$

Si $f^{-1}(\Gamma_0) = \Gamma_1$ ou Γ_2, on a nécessairement $c = 0$, $b = 0$ et $a' = 0$, sinon on obtient des termes en y^4, x^3 ou x. Mais alors on n'a pas de terme en x^2. Il est donc impossible d'avoir $f^{-1}(\Gamma_0) = \Gamma_1$ ou Γ_2 : Γ_0 n'est équivalente ni à Γ_1 ni à Γ_2. On démontre de même que Γ_1 et Γ_2 ne sont pas équivalentes.

Si $m = n = 4$, C et C′ ne sont pas nécessairement équivalentes. Dans les graphes des polynômes P de degré 4, il y a trois classes d'équivalence correspondant à $P = X^4$, $P = X^4 + X^2$ et $P = X^4 - X^2$. ◁

4.38. Étude d'une suite d'applications affines

Soit E un espace préhilbertien réel. Pour toute droite affine $\mathcal{D}$ de E, on note $p_{\mathcal{D}}$ la projection orthogonale sur $\mathcal{D}$ et $\Psi_{\mathcal{D}}$ l'application qui à $x \in E$ associe le milieu du segment $[x, p_{\mathcal{D}}(x)]$. Soit $n \geqslant 2$, $\mathcal{D}_1, \ldots, \mathcal{D}_n$, n droites affines de E non coplanaires deux à deux. Soit $(\omega_k)_{k\in\mathbb{N}}$ une suite d'éléments de $[\![1, n]\!]$, non stationnaire. Soit enfin la suite $(f_k)_{k\in\mathbb{N}}$ définie par $\begin{cases} f_0 = \mathrm{Id}_E \\ \forall k \geqslant 1 \quad f_k = \psi_{\mathcal{D}_{\omega_k}} \circ f_{k-1}. \end{cases}$ On suppose que la suite $(f_k(0))$ tend vers ℓ. Étudier la suite (f_k).

(**École normale supérieure**)

▷ **Solution.**

Pour tout $x \in E$, on étudie la limite de la suite $\big(f_k(x)\big)$. Une projection orthogonale est continue, car 1-lipschitzienne. On en déduit que les applications $\Psi_{\mathcal{D}_i}$, puis les applications f_k sont continues. On a, pour tout $x \in E$, $\|f_k(x) - f_k(0)\| = \|\overrightarrow{f_k}(x)\| \leqslant \|\overrightarrow{f_k}\| \cdot \|x\|$. Nous allons montrer que la suite $\left(\|\overrightarrow{f_k}\|\right)$ tend vers 0, ce qui montrera que, pour tout $x \in E$, $\big(f_k(x)\big)$ converge vers ℓ.

Pour toute droite affine $\mathcal{D}$ de E, l'application $\Psi_{\mathcal{D}}$ est affine et sa partie linéaire est $\overrightarrow{\Psi_{\mathcal{D}}} = \dfrac{1}{2}\left(\mathrm{Id}_E + \overrightarrow{p_{\mathcal{D}}}\right)$. Pour tout $i \in [\![1, n]\!]$, on a donc

$$\|\overrightarrow{\Psi_{\mathcal{D}_i}}\| \leqslant \frac{1}{2}\left(1 + \|\overrightarrow{p_{\mathcal{D}_i}}\|\right) \leqslant 1,$$

car la norme d'une projection orthogonale non nulle est égale à 1. Pour tout $i \in [\![1, n]\!]$, notons u_i un vecteur unitaire dirigeant $\mathcal{D}_i$. Si $i \neq j$ et $u \in E$, il existe $\lambda \in \mathbb{R}$ tel que $\overrightarrow{p_{\mathcal{D}_i}}(u) = \lambda u_i$. On en déduit que

$$\overrightarrow{p_{\mathcal{D}_j}} \circ \overrightarrow{p_{\mathcal{D}_i}}(u) = \lambda \overrightarrow{p_{\mathcal{D}_j}}(u_i) = \lambda(u_i, u_j)u_j,$$

puis

$$\|\overrightarrow{p_{\mathcal{D}_j}} \circ \overrightarrow{p_{\mathcal{D}_i}}(u)\| = |\lambda|\,|(u_i, u_j)| = \|\overrightarrow{p_{\mathcal{D}_i}}(u)\|\,|(u_i, u_j)| \leqslant \|u\|\,|(u_i, u_j)|.$$

Cela montre que $\|\overrightarrow{p_{\mathcal{D}_j}} \circ \overrightarrow{p_{\mathcal{D}_i}}\| \leqslant |\langle u_i, u_j\rangle| < 1$, car les vecteurs u_i et u_j sont de norme 1 et ne sont pas colinéaires. On obtient

$$\|\overrightarrow{\Psi_{\mathcal{D}_j}} \circ \overrightarrow{\Psi_{\mathcal{D}_i}}\| = \frac{1}{4}\|\mathrm{Id}_E + \overrightarrow{p_{\mathcal{D}_i}} + \overrightarrow{p_{\mathcal{D}_j}} + \overrightarrow{p_{\mathcal{D}_j}} \circ \overrightarrow{p_{\mathcal{D}_i}}\| \leqslant \frac{1}{4}(3 + |\langle u_i, u_j\rangle|) < 1.$$

Notons $M = \max_{i \neq j} \frac{1}{4}(3 + |\langle u_i, u_j\rangle|)$ et pour tout entier $k \in \mathbb{N}^*$, n_k le plus grand entier p pour lequel il existe des entiers $i_1, i_2, \ldots, i_p$ de $[\![1, k]\!]$ tels que, pour $j \in [\![1, p-1]\!]$, $\omega_{i_j} \neq \omega_{i_j+1}$ et $i_{j+1} > i_j + 1$. Comme la norme triple est une norme d'algèbre et que $\|\Psi_{\mathcal{D}_i}\| \leqslant 1$, pour tout $i \in [\![1, n]\!]$, on obtient, pour tout $k \in \mathbb{N}^*$,

$$\|\overrightarrow{f_k}\| \leqslant \prod_{j=1}^{n_k} \|\Psi_{\mathcal{D}_{\omega_{i_j+1}}} \circ \Psi_{\mathcal{D}_{\omega_{i_j}}}\| \leqslant M^{n_k}.$$

Quand k tend vers $+\infty$, n_k tend vers $+\infty$, car la suite (ω_k) n'est pas stationnaire, et comme $0 \leqslant M < 1$, la suite $\left(\|\overrightarrow{f_k}\|\right)$ tend vers 0.

Conclusion. Pour tout $x \in E$, $\left(f_k(x)\right)$ tend vers ℓ. ◁

L'énoncé suivant est difficile.

4.39. Applications du plan conservant l'orthogonalité

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ une application conservant l'orthogonalité (si abc est un triangle rectangle en b, alors $f(a)f(b)f(c)$ est un triangle rectangle en $f(b)$). Que peut-on dire de f ?

(**École normale supérieure**)

▷ **Solution.**

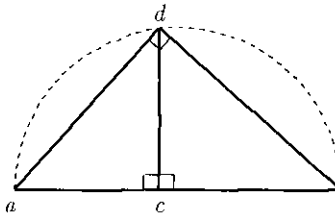
L'espace vectoriel $\mathbb{R}^2$ est muni de sa structure d'espace affine naturelle. Il est clair que toutes les similitudes affines conservent l'orthogonalité. Nous allons montrer que ce sont les seules applications ayant cette propriété. La difficulté de l'exercice tient à ce que f n'a aucune propriété *a priori* : ce serait évidemment beaucoup plus facile si on supposait f affine au départ.

• Énonçons quelques conséquences de l'hypothèse.

(i) f est injective : en effet si a et b sont distincts, on prend un point c tel que le triangle abc soit rectangle. Alors $f(a)f(b)f(c)$ est un triangle rectangle et donc $f(a) \neq f(b)$.

(ii) f conserve l'alignement : si a, b, c sont trois points alignés distincts, soit d tel que les droites (ac) et (cd) soient orthogonales. Les triangles acd

et bcd sont rectangles en c. Les triangles $f(a)f(c)f(d)$ et $f(b)f(c)f(d)$ sont donc rectangles en $f(c)$. On a alors $(f(c)f(a)) \perp (f(c)f(d))$ et $(f(c)f(b)) \perp (f(c)f(d))$. Les points $f(a), f(b), f(c)$ sont donc alignés.



On peut ajouter que si $c \in [ab]$ alors $f(c) \in [f(a)f(b)]$. En effet, on peut choisir d tel que (ac) et (cd) soient orthogonales et que, de plus, abd soit rectangle en d (on prend d sur le cercle de diamètre $[ab]$ et sur la perpendiculaire à (ab) en c : voir figure ci-dessus). Alors le triangle $f(a)f(d)f(b)$ est rectangle en $f(d)$ et le point $f(c)$, projeté orthogonal de $f(d)$ sur $(f(a)f(b))$ appartient à $[f(a)f(b)]$.

(iii) f conserve le parallélisme : soient a, b, c, d quatre points tels que les droites (ab) et (cd) soient parallèles. On les suppose distinctes, sinon les images des quatre points sont alignées d'après ce qui précède, et il n'y a rien à démontrer.

Considérons la droite orthogonale à (ab) contenant a. Elle coupe la droite (cd) en e. Les triangles abe et aec sont rectangles en a et e, respectivement. Les triangles $f(a)f(b)f(c)$ et $f(a)f(c)f(e)$ sont rectangles en $f(a)$ et $f(e)$ respectivement. On a donc $(f(a)f(e)) \perp (f(a)f(b))$ et $(f(a)f(e)) \perp (f(c)f(e))$; on en déduit $(f(a)f(b))$ parallèle à $(f(c)f(e))$. Les points c, d, e étant alignés, les points $f(c), f(d), f(e)$ le sont également. On a donc $(f(a)f(b))$ parallèle à $(f(c)f(d))$. (Tout ceci suppose $c \neq e$; sinon, on remplace c par d).

On en déduit que f transforme un parallélogramme en un parallélogramme.

• Montrons maintenant que f est une application affine. Il s'agit de démontrer que l'application

$$\varphi : a \in \mathbb{R}^2 \longmapsto f(a) - f(0) \in \mathbb{R}^2$$

est linéaire (0 désigne l'élément $(0,0)$ de $\mathbb{R}^2$).

Les démonstrations qui suivent résultent du fait que les points qui interviennent dans les démonstrations peuvent être construits à l'aide de parallèles.

(i) Montrons que f est un morphisme de $(\mathbb{R}^2, +)$.

Soit a et b deux éléments de $\mathbb{R}^2$, $c = a + b$. Si a et b ne sont pas colinéaires, alors $0acb$ est un parallélogramme, donc $f(0)f(a)f(c)f(b)$ en

est un également. On a donc $f(c) - f(a) = f(b) - f(0)$, ce qui s'écrit $f(c) - f(0) = f(a) - f(0) + f(b) - f(0)$, c'est-à-dire $\varphi(c) = \varphi(a) + \varphi(b)$.

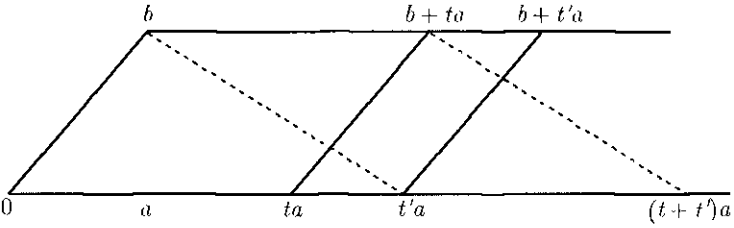
Si a et b sont colinéaires, on peut supposer $a \neq 0$. Il existe $\lambda \in \mathbb{R}$ tel que $b = \lambda a$. Il faut démontrer que $\varphi((1+\lambda)a) = \varphi(a) + \varphi(\lambda a)$. Cela résultera du point suivant.

(ii) Montrons que, pour $a \in \mathbb{R}^2$ et $t \in \mathbb{R}$, $\varphi(ta) = t\varphi(a)$.

C'est évident pour $a = 0$. Fixons donc $a \neq 0$. Pour tout $t \in \mathbb{R}$, le point ta appartient à la droite $(0a)$. Le point $f(ta)$ appartient donc à la droite $(f(0)f(a))$. Il existe donc $t' \in \mathbb{R}$ tel que $f(ta) - f(0) = t'(f(a) - f(0))$, c'est-à-dire $\varphi(ta) = t'\varphi(a)$. Le réel t' est unique, car $a \neq 0$ implique $f(a) \neq f(0)$.

Considérons l'application $\sigma : t \longmapsto t'$. Il s'agit de montrer que $\sigma = \mathrm{Id}_{\mathbb{R}}$. On remarque que, par construction, $\sigma(1) = 1$ et $\sigma(0) = 0$. Nous allons montrer que σ est un morphisme de corps de $\mathbb{R}$.

Soit $(t, t') \in \mathbb{R}^{*2}$ (si t ou t' sont nuls, il n'y a rien à démontrer). Soit $b \in \mathbb{R}^2$ tels que a et b ne soient pas colinéaires.



Les points $0, ta, b+ta$ et b forment un parallélogramme. Il en de même de leurs images par f. On a donc $f(b+ta) - f(b) = f(ta) - f(0)$. On montre de même que $f(t'a), f((t+t')a), f(b+ta)$ et $f(b)$ forment un parallélogramme. On a donc $f(b+ta) - f(b) = f((t+t')a) - f(t'a)$.

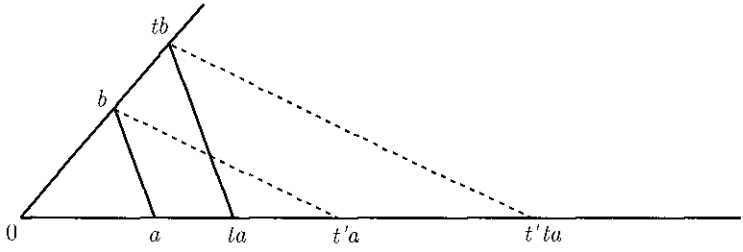
On en déduit que $f((t+t')a) - f(t'a) = f(ta) - f(0)$ et donc

$$f((t+t')a) - f(0) = f(ta) - f(0) + f(t'a) - f(0),$$

c'est-à-dire $\varphi((t+t')a) = \varphi(ta) + \varphi(t'a)$. Ceci montre que

$$\boxed{\sigma(t+t') = \sigma(t) + \sigma(t').}$$

Les droites (ab) et $(tatb)$ sont parallèles. Il en est donc de même des droites $(f(a)f(b))$ et $(f(ta)f(tb))$. On a $f(ta) - f(0) = \sigma(t)(f(a) - f(0))$ par définition de σ. Du théorème de Thalès on déduit que l'on a aussi $f(tb) - f(0) = \sigma(t)(f(b) - f(0))$.



Les droites $(t'ab)$ et $(tt'atb)$ aussi sont parallèles. Il en est de même de $(f(t'a)f(b))$ et $(f(tt'a)f(tb)$. Comme $f(tb) - f(0) = \sigma(t)(f(b) - f(0))$, on déduit du théorème de Thalès que

$$f(tt'a) - f(0) = \sigma(t)(f(t'a) - f(0)).$$

On obtient donc $\varphi(tt'a) = \sigma(t)\varphi(t'a) = \sigma(t)\sigma(t')\varphi(a)$ et donc

$$\boxed{\sigma(tt') = \sigma(t)\sigma(t')}.$$

Donc σ est un morphisme de corps de $\mathbb{R}$ et il est connu que le seul morphisme de corps de $\mathbb{R}$ est l'identité. Donc σ est égal à $\mathrm{Id}_{\mathbb{R}}$. Ceci achève de démontrer que φ est linéaire et donc que f est affine.

• Montrons pour conclure que f est une similitude affine. L'application φ, comme f, est injective donc bijective car c'est un endomorphisme d'un espace de dimension finie. L'application f est affine et conserve l'orthogonalité. On a donc, pour $(a, b) \in (\mathbb{R}^2)^2$,

$$a \perp b \Longrightarrow \varphi(a) \perp \varphi(b).$$

Soit (e_1, e_2) une base orthonormale de $\mathbb{R}^2$. Alors $(\varphi(e_1), \varphi(e_2))$ est une base de $\mathbb{R}^2$, orthogonale. Les vecteurs $e_1 + e_2$ et $e_1 - e_2$ sont orthogonaux donc $\varphi(e_1) + \varphi(e_2)$ et $\varphi(e_1) - \varphi(e_2)$ également. On en déduit que $\varphi(e_1)$ et $\varphi(e_2)$ ont même norme, que l'on note k $(k > 0)$.

La famille $\left(\frac{1}{k}\varphi(e_1), \frac{1}{k}\varphi(e_2)\right)$ est une base orthonormale de $\mathbb{R}^2$, donc φ est la composée d'une homothétie vectorielle de rapport k et d'un endomorphisme orthogonal. C'est une similitude vectorielle et f est donc une similitude affine.

Conclusion. Les applications de $\mathbb{R}^2$ dans $\mathbb{R}^2$ qui conservent l'orthogonalité sont les similitudes affines. ◁

La démonstration du fait que f est affine reprend les arguments de la démonstration du théorème fondamental de la géométrie affine qui dans le cas réel s'énonce ainsi : soit $\mathcal{E}$ et $\mathcal{E}'$ deux espaces affines réels de même dimension $\geqslant 2$ et f une bijection ensembliste de $\mathcal{E}$ sur $\mathcal{E}'$, conservant l'alignement. Alors f est une bijection affine de $\mathcal{E}$ sur $\mathcal{E}'$.

Voici quelques exercices sur le thème des isométries.

4.40. Sous-groupes finis du groupe affine réel

Montrer qu'un sous-groupe fini du groupe des bijections affines d'un plan vectoriel réel E est isomorphe à un sous-groupe du groupe $O_2(\mathbb{R})$.

(École normale supérieure)

▷ **Solution.**

Notons GA(E) le groupe des bijections affines de E et soit G un sous-groupe fini de GA(E). Si $f \in G$, sa partie linéaire $\vec{f}$ est dans le groupe GL(E). Nous savons de plus que l'application $\psi : f \longmapsto \vec{f}$ est un morphisme surjectif de groupes de GA(E) dans GL(E) dont le noyau est le sous-groupe des translations. Comme G est fini, il ne contient aucune translation non triviale car celles-ci sont d'ordre infini. Par suite, la restriction de ψ à G établit un isomorphisme entre G et $\psi(G)$ qui est un sous-groupe fini de GL(E).

Il nous suffit donc de traiter le cas où G est un sous-groupe fini de GL(E). Pour cela, il suffit de trouver un produit scalaire sur E conservé par tous les éléments de G. L'argument est classique. On choisit un produit scalaire $(\ ,\ )$ quelconque et on pose pour $(x, y) \in E^2$,

$$\langle x, y\rangle_G = \frac{1}{|G|} \sum_{g\in G} \langle g(x), g(y)\rangle$$

On voit aisément que $\langle\ ,\ \rangle_G$ est un nouveau produit scalaire sur E et que tous les éléments de G sont orthogonaux pour $\langle\ ,\ \rangle_G$. En choisissant une base orthonormée B de E pour ce nouveau produit scalaire, l'application qui à $g \in G$ associe sa matrice dans B, établit alors un isomorphisme entre G et un sous-groupe de $O_2(\mathbb{R})$. ◁

Notons que le résultat reste vrai avec la même preuve en toute dimension finie. L'argument de moyenne utilisé dans cet exercice est très important (cf. aussi la note qui suit l'exercice 6.22 dans le tome 1).

4.41. Une partie bornée de $\mathbb{R}^2$ n'est pas dédoublable

Le plan $\mathbb{R}^2$ est muni de sa structure euclidienne canonique. Soit B une partie bornée de $\mathbb{R}^2$ ayant au moins deux points.

1. Montrer qu'il existe un et un seul disque fermé de rayon minimal contenant B.

2. Montrer qu'il n'existe pas de partition de B en deux parties B_1 et B_2 toutes deux isométriques à B (on dit que B n'est pas *dédoublable*).

(École polytechnique)

▷ **Solution.**

1. Le lecteur trouvera la solution de cette question classique dans l'exercice 4.45. On note D le disque de rayon minimal, a son centre et r son rayon.

2. On raisonne par l'absurde en supposant qu'il existe deux parties B_1 et B_2 disjointes de $\mathbb{R}^2$ telles que $B = B_1 \cup B_2$ et deux isométries f_1 et f_2 de $\mathbb{R}^2$ telles que $B_1 = f_1(B)$ et $B_2 = f_2(B)$.

Pour $i \in \{1, 2\}$, on a $f_i(B) = B_i \subset B \subset D$. On en déduit $B \subset f_i^{-1}(D)$. Comme $f_i^{-1}(D)$ est aussi un disque de rayon r, on obtient, par unicité de D, $f_i^{-1}(D) = D$ et donc $f_i(D) = D$. Comme $f_i(D)$ a pour centre $f_i(a)$, on en déduit $f_i(a) = a$.

L'isométrie f_i est donc soit une rotation de centre a, soit une réflexion d'axe contenant a. Si f_i est une réflexion, alors on a, pour tout $x \in B$, $f_i(x) \in B_i$ et donc $f_i(x) \in B$. On en déduit $x = f_i(f_i(x)) \in f_i(B) = B_i$ et donc $B_i = B$. C'est faux. On a donc montré que f_1 et f_2 sont des rotations de centre a.

Les rotations f_1 et f_2 ont même centre donc elles commutent. On en déduit $f_1(f_2(B)) \subset f_1(B) \subset B_1$ et $f_1(f_2(B)) = f_2(f_1(B)) \subset f_2(B) \subset B_2$. On obtient $f_1(f_2(B)) \subset B_1 \cap B_2 = \emptyset$. C'est impossible car $B \neq \emptyset$. On aboutit à la contradiction voulue. ◁

La preuve qui précède est due à H. HADWIGER *et* H. DEBRUNNER *(1964). Aucune partie bornée de $\mathbb{R}^2$ n'est dédoublable, mais il existe des parties non bornées de $\mathbb{R}^2$ dédoublables (c'est le paradoxe de* SIERPINSKI-MAZURKIEWICZ *(1914)). Par contre, aucune partie de $\mathbb{R}$ n'est dédoublable (résultat dû à* SIERPINSKI*). Toutes ces questions ont fait l'objet du problème de l'Agrégation Interne de 2004.*

L'exercice suivant est assez difficile. Il montre qu'une application du plan dans lui-même est une isométrie affine dès qu'elle conserve la distance 1.

4.42. Application conservant la distance unité

Soit $f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ une application conservant la distance 1. Montrer que f est une isométrie affine.

(École normale supérieure)

▷ **Solution.**

Notons F l'ensemble de toutes les applications de $\mathbb{R}^2$ dans $\mathbb{R}^2$ qui conservent la distance 1. Il est clair que F contient les isométries. On veut montrer l'autre inclusion. Si $f \in F$, notons D_f l'ensemble des distances conservées par f. C'est un ensemble non vide, contenant 1 et 0.

Posons $\mathrm{D} = \bigcap_{f \in \mathrm{F}} \mathrm{D}_f$. C'est l'ensemble (non vide) de toutes les distances conservées par *toutes* les applications f de F. Le problème revient à montrer que $\mathrm{D} = \mathbb{R}_+$. On procède en trois étapes.

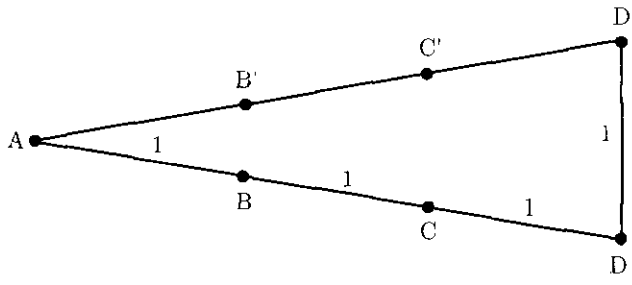
• Montrons que l'ensemble D est stable pour le produit.

Soient α, β deux réels dans D et $f \in \mathrm{F}$. On peut supposer α et β non nuls. Soit h une homothétie de rapport α. Posons $g = h^{-1} \circ f \circ h$. Si A, B sont deux points distants de 1, $h(\mathrm{A})h(\mathrm{B}) = \alpha$ et comme f conserve la distance α, $g(\mathrm{A})g(\mathrm{B}) = 1$. Il en résulte que $g \in \mathrm{F}$. Ainsi, g conserve la distance β et comme $f = h \circ g \circ h^{-1}$ on voit aisément que f conserve la distance $\alpha\beta$. Cela vaut pour tout f, donc $\alpha\beta \in \mathrm{D}$. On voit tout l'intérêt qu'il y a à considérer l'ensemble F tout entier plutôt que seulement l'un de ses éléments.

• On va maintenant trouver divers éléments de D en dehors de 0 et 1. Montrons pour commencer que $\sqrt{3} \in \mathrm{D}$.

Soit $f \in \mathrm{F}$. Il apparaît que l'image d'un triangle équilatéral de côté 1 par f est un triangle équilatéral de côté 1. Soient A, B distants de $\sqrt{3}$ et C, D tels que (ACD) et (BCD) soient équilatéraux de côtés 1. En considérant les images par f de ces deux triangles, on voit que soit $f(\mathrm{A}) = f(\mathrm{B})$, soit $f(\mathrm{A})f(\mathrm{B}) = \sqrt{3}$. Notons Γ le cercle de centre A et de rayon $\sqrt{3}$. Le résultat obtenu pour le couple (A, B) vaut pour tout couple (A, B′) avec $\mathrm{B}' \in \Gamma$. Si l'on choisit B′ sur Γ tel que $\mathrm{BB}' = 1$, on doit avoir $f(\mathrm{B})f(\mathrm{B}') = 1$ et comme $f(\mathrm{A})f(\mathrm{B}') \in \{0, \sqrt{3}\}$ on ne peut avoir $f(\mathrm{B}) = f(\mathrm{A})$. Conclusion : f conserve la distance $\sqrt{3}$. Comme ceci vaut pour tout f, $\sqrt{3} \in \mathrm{D}$. Il en résulte que $3 \in \mathrm{D}$.

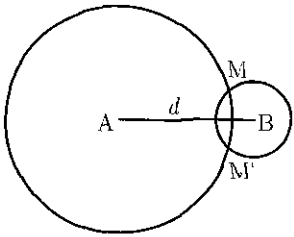
Montrons maintenant que $\frac{1}{3}$ et $\frac{2}{3}$ sont aussi dans D. Pour cela, on considère la figure suivante :



Comme f conserve les distances 1 et 3, cette figure est envoyée par f sur une figure isométrique. En particulier, $f(\mathrm{B})f(\mathrm{B}') = \mathrm{BB}' = \frac{1}{3}$ et $f(\mathrm{C})f(\mathrm{C}') = \mathrm{CC}' = \frac{2}{3}$ (par le théorème de Thalès).

On en déduit que l'ensemble $H = \{2^n\sqrt{3}^p,\ n \in \mathbb{N},\ p \in \mathbb{Z}\}$ est inclus dans D, grâce à la stabilité par produit. Par ailleurs, H est dense dans $\mathbb{R}_+$. En effet H est l'image par l'exponentielle de $\left\{n \ln 2 + \frac{p}{2} \ln 3.\ n \in \mathbb{N}, p \in \mathbb{Z}\right\}$ qui est dense dans $\mathbb{R}$, car $\frac{\ln 2}{\ln 3} \notin \mathbb{Q}$ (attention toutefois, ce n'est pas un sous-groupe additif de $\mathbb{R}$). L'exponentielle étant un homéomorphisme de $\mathbb{R}$ dans $\mathbb{R}_+^*$ le résultat suit.

• On peut enfin conclure que $D = \mathbb{R}_+$. Soit $d > 0$ un réel quelconque, $f \in F$ et A, B deux points de $\mathbb{R}^2$ distants de d. Soit $\varepsilon > 0$ dans D, $\varepsilon < \frac{d}{2}$. Par densité de D on peut trouver $d' \in D$ tel que $d - \varepsilon < d' < d$. Le cercle de centre B et de rayon ε coupe le cercle de centre A et de rayon d' en deux points M, M'.



On a $f(A)f(M) = f(A)f(M') = d'$ et $f(B)f(M) = f(B)f(M') = \varepsilon$. Il en résulte que le cercle de centre $f(A)$ et de rayon d' coupe le cercle de centre $f(B)$ et de rayon ε en $f(M)$ et $f(M')$. Pour cela il est nécessaire que $d' - \varepsilon \leqslant f(A)f(B) \leqslant d' + \varepsilon$ et donc que $d - 2\varepsilon \leqslant f(A)f(B) \leqslant d + \varepsilon$. Comme il y a dans D des éléments arbitrairement petits, on en déduit en faisant tendre ε vers 0 que $f(A)f(B) = d$.

Conclusion. $D = \mathbb{R}_+$ et toute application de $\mathbb{R}^2$ conservant la distance unité est une isométrie. ◁

Le résultat n'est pas vrai pour $\mathbb{R}$ (prendre f qui fixe les points de $\mathbb{R} \setminus \mathbb{Z}$ et qui envoie chaque entier n sur $n + 1$) mais se généralise à $\mathbb{R}^n$ pour $n \geqslant 3$.

Si X est une partie de $\mathbb{R}^n$, on peut s'intéresser au groupe des isométries de X. Deux approches sont possibles : on peut regarder le sous-groupe de $\mathcal{S}(X)$ formé des bijections isométriques de X dans X, c'est-à-dire des bijections qui conservent la distance euclidienne. Mais on peut aussi regarder le sous-groupe des isométries f de $\mathbb{R}^n$ telles que $f(X) = X$. La restriction de f à X est alors une permutation isométrique de X. En fait, il est facile de voir que dans le cas où X engendre affinement $\mathbb{R}^n$, le morphisme de restriction qu'on vient de définir est un isomorphisme. Par exemple si ABC est un triangle équilatéral de $\mathbb{R}^2$, le groupe des isométries contient 6 éléments : l'identité, les 3 transpositions (qui correspondent au

symétries orthogonales selon les hauteurs-médianes-médiatrices), et les 3-cycles qui correspondent aux rotations de centre G *et d'angle* $\pm\frac{2\pi}{3}$. *Autrement dit, toutes les permutations des points* A, B, C *sont isométriques. Il n'en est plus de même pour un carré* ABCD *: la permutation qui échange* A *et* B *par exemple n'est pas isométrique. On montre que le groupe des isométries d'un polygone régulier à* n *côtés de* $\mathbb{R}^2$ *contient* $2n$ *éléments : ce groupe, noté* D_n *est le groupe diédral d'ordre* n. *L'exercice suivant complète l'exercice 1.20 avec l'étude du groupe des isométries du simplexe régulier. Le lecteur pourra trouver dans l'exercice 2.21 du tome Algèbre 1 un exemple montrant l'intérêt qu'il peut y avoir à réaliser un groupe comme le groupe des isométries d'un objet géométrique.*

4.43. Groupe des isométries du simplexe régulier

Pour tout $p \in \mathbb{N}^*$, on identifie le vecteur $(x_1, \dots, x_p) \in \mathbb{R}^p$ à $(x_1, \dots, x_p, 0) \in \mathbb{R}^{p+1}$ ce qui permet d'identifier $\mathbb{R}^p$ à une partie de $\mathbb{R}^{p+1}$.

Pour $r > 0$, on définit un simplexe t_n de $\mathbb{R}^n$ par récurrence en posant $t_1 = (A_0, A_1)$, où $A_0 \in \mathbb{R}$, $A_1 \in \mathbb{R}$ et $d(A_0, A_1) = r$, $t_2 = (A_0, A_1, A_2)$, où $A_2 \in \mathbb{R}^2$ et $d(A_2, A_0) = d(A_1, A_0) = r$, ..., $t_n = (A_0, \dots, A_n)$, où $A_n \in \mathbb{R}^n$ et $d(A_n, A_i) = r$ pour $i \in [\![0, n-1]\!]$.

1. Justifier l'existence d'un tel simplexe.

2. Montrer que la famille t_n est affinement libre.

3. Déterminer l'isobarycentre des points de t_n.

4. Montrer que l'ensemble des isométries directes de $\mathbb{R}^n$ laissant t_n globalement invariant est isomorphe au groupe alterné $\mathcal{A}_{n+1}$.

(**École polytechnique**)

▷ **Solution.**

On note $(e_1, \dots, e_n)$ la base canonique de $\mathbb{R}^n$.

1. On raisonne par récurrence : on montre qu'on peut construire t_n et qu'il vérifie de plus $A_n \in \mathbb{R}^n \setminus \mathbb{R}^{n-1}$. L'existence de t_1 est évidente. Il s'agit de choisir deux réels quelconques, distants de r. Le simplexe t_{n-1} étant construit ($n \geqslant 1$), on cherche $A_n \in \mathbb{R}^n$ tel que $d(A_n, A_i) = r$ pour tout $i \in [\![0, n-1]\!]$. On pose $A_n = B_n + \lambda e_n$, où $B_n \in \mathbb{R}^{n-1}$. On a alors, pour tout $i \in [\![0, n-1]\!]$, $d(A_n, A_i) = \sqrt{d(B_n, A_i)^2 + \lambda^2}$. On cherche $B_n \in \mathbb{R}^{n-1}$ tel que la distance $d(B_n, A_i)$ soit indépendante de $i \in [\![0, n-1]\!]$ et inférieure à r.

Montrons que le point Ω_{n-1}, isobarycentre de $(A_0, \dots, A_{n-1})$, convient. On a $\Omega_{n-1} = \frac{1}{n}\sum_{k=0}^{n-1} A_k$. En posant $C_i = A_i - \Omega_{n-1}$, on obtient

$\sum_{k=0}^{n-1} C_k = 0$ et $\|C_i - C_j\| = r$, soit $r^2 = \|C_i\|^2 + \|C_j\|^2 - 2\langle C_i, C_j\rangle$ pour tout $i \neq j$. En sommant cette égalité sur j, il vient

$$\begin{aligned}(n-1)r^2 &= (n-1)\|C_i\|^2 + \sum_{j\neq i}\|C_j\|^2 - 2\langle C_i, \sum_{j\neq i} C_j\rangle \\ &= (n-1)\|C_i\|^2 + \sum_{j=0}^{n-1}\|C_j\|^2 - \|C_i\|^2 - 2\langle C_i, -C_i\rangle \\ &= n\|C_i\|^2 + \sum_{j=0}^{n-1}\|C_j\|^2.\end{aligned}$$

Ceci montre que $\|C_i\|$ est indépendant de i, puis que $\|C_i\| = \sqrt{\dfrac{n-1}{2n}}r$. En prenant $B_n = \Omega_{n-1}$, on obtient $d(B_n, A_i\rangle = \|C_i\|$, indépendant de i. Si $\lambda = \pm\sqrt{\dfrac{n+1}{2n}}\,r$, le point A_n et le simplexe t_n ont les propriétés voulues. Comme $\lambda \neq 0$, on a bien $A_n \in \mathbb{R}^n \setminus \mathbb{R}^{n-1}$. Cela termine la démonstration par récurrence.

Remarquons que si B_n est un point de $\mathbb{R}^{n-1}$, équidistant de tous les A_i, pour $i \in [\![0, n-1]\!]$, on a, pour tout $i \neq j$, $\overrightarrow{\Omega_{n-1}B_n} \perp \overrightarrow{A_iA_j}$. Ainsi $\overrightarrow{\Omega_{n-1}B_n}$ appartient à $\mathrm{Vect}(\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_{n-1}})^\perp$. Or la matrice de la famille $(\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_{n-1}})$ dans la base canonique de $\mathbb{R}^{n-1}$ est triangulaire supérieure avec des termes non nuls sur la diagonale car $\overrightarrow{A_0A_i} \in \mathbb{R}^i \setminus \mathbb{R}^{i-1}$ par hypothèse de récurrence. Ainsi $\mathrm{Vect}(\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_{n-1}}) = \mathbb{R}^{n-1}$ et $\overrightarrow{\Omega_{n-1}B_n} = 0$, donc $B_n = \Omega_{n-1}$. Donc B_n est unique et, t_{n-1} étant construit, on a deux points A_n possibles. Le point A_0 étant fixé, il y a donc 2^n simplexes t_n possibles.

2. En reprenant la démonstration de la question précédente, on voit que, pour tout $n \in \mathbb{N}^*$, la famille $(\overrightarrow{A_0A_1}, \ldots, \overrightarrow{A_0A_n})$ est libre. La famille $(A_0, A_1, \ldots, A_n)$ est donc affinement libre.

3. En notant, pour tout $n \in \mathbb{N}^*$, Ω_n l'isobarycentre de la famille $(A_0, A_1, \ldots, A_n)$, il résulte de la question 1 que $A_n = \Omega_{n-1} \pm \sqrt{\dfrac{n+1}{2n}}\,re_n$, puis que

$$\begin{aligned}\Omega_n &= \frac{1}{n+1}(A_0 + \cdots + A_n) \\ &= \frac{1}{n+1}(n\Omega_{n-1} + A_n) = \frac{1}{n+1}\left((n+1)\Omega_{n-1} \pm \sqrt{\frac{n+1}{2n}}re_n\right) \\ &= \Omega_{n-1} \pm \frac{r}{\sqrt{2n(n+1)}}e_n.\end{aligned}$$

On obtient, pour tout $n \in \mathbb{N}^*$,

$$\Omega_n = \mathrm{A}_0 + \sum_{k=1}^{n} \frac{\pm r}{\sqrt{2k(k+1)}} e_k.$$

4. Si f est une isométrie de $\mathbb{R}^n$ laissant t_n invariant, elle induit une permutation de $\{\mathrm{A}_0, \mathrm{A}_1, \ldots, \mathrm{A}_n\}$. Réciproquement, soit $n \in \mathbb{N}^*$ et σ une permutation de $[\![0, n]\!]$. La famille $(\mathrm{A}_0, \mathrm{A}_1, \ldots \mathrm{A}_n)$ étant une base affine, il existe une unique application affine f de $\mathbb{R}^n$ telle que, pour tout $i \in [\![1, n]\!]$, $f(\mathrm{A}_i) = \mathrm{A}_{\sigma(i)}$. Montrons que $\vec{f}$ est un endomorphisme orthogonal.

Soit $(i, j, k) \in [\![1, n]\!]^3$, distincts. On a $\|\mathrm{A}_i\mathrm{A}_j\|^2 = r^2$ et

$$(\overrightarrow{\mathrm{A}_i\mathrm{A}_j}, \overrightarrow{\mathrm{A}_i\mathrm{A}_k}) = \frac{1}{2}(\|\mathrm{A}_i\mathrm{A}_j\|^2 + \|\mathrm{A}_i\mathrm{A}_k\|^2 - \|\mathrm{A}_j\mathrm{A}_k\|^2) = \frac{r^2}{2}.$$

On en déduit que, pour tout $(i, j) \in [\![1, n]\!]^2$,

$$\langle \vec{f}(\overrightarrow{\mathrm{A}_0\mathrm{A}_i}), \vec{f}(\overrightarrow{\mathrm{A}_0\mathrm{A}_j}) \rangle = \langle \overrightarrow{\mathrm{A}_0\mathrm{A}_i}, \overrightarrow{\mathrm{A}_0\mathrm{A}_j} \rangle.$$

L'application linéaire $\vec{f}$ conserve le produit scalaire sur les vecteurs de la base $(\overrightarrow{\mathrm{A}_0\mathrm{A}_1}, \ldots, \overrightarrow{\mathrm{A}_0\mathrm{A}_n})$ de $\mathbb{R}^n$, donc c'est un endomorphisme orthogonal.

L'application F qui à σ associe f est une bijection de $\mathcal{S}_n$ sur l'ensemble $\mathcal{I}_n$ des isométries de $\mathbb{R}^n$ qui laissent invariant t_n. Il est clair que F est un isomorphisme de groupes de $(\mathcal{S}_n, \circ)$ sur $(\mathcal{I}_n, \circ)$.

Si σ est la transposition $\tau_{i,j}$, qui échange i et j on obtient pour f la réflexion par rapport à l'hyperplan médiateur du segment $[\mathrm{A}_i, \mathrm{A}_j]$ et dans ce cas $\det(\vec{f}) = -1$. Il en résulte que pour toute permutation σ le déterminant de $\overrightarrow{\mathrm{F}(\sigma)}$ est égal à la signature de σ. Ainsi, $\mathrm{F}(\sigma)$ est une isométrie positive si et et seulement si σ appartient au groupe alterné d'ordre $n+1$. L'application F réalise un isomorphisme de groupes de $\mathcal{A}_{n+1}$ sur l'ensemble des isométries positives qui laissent t_n invariant. ◁

En particulier, l'ensemble des isométries positives de $\mathbb{R}^3$ laissant invariant un tétraèdre régulier est isomorphe à $\mathcal{A}_4$. Il comporte donc $\frac{4!}{2} = 12$ éléments.

Nous regroupons maintenant quelques exercices ayant pour dénominateur commun la notion de convexité. Les résultats de convexité sont souvent intuitifs mais difficiles et mêlent des considérations géométriques, topologiques et algébriques. Ils sont par conséquent difficiles à classer. Par choix nous avons placé la grande majorité des énoncés dans le tome 3 d'analyse et n'avons gardé ici que les exercices où les arguments topologiques sont réduits au minimum.

Le fait évident qu'une intersection de parties convexes est convexe permet de définir l'enveloppe convexe d'une partie X quelconque (plus petit convexe contenant X). Les points de cette enveloppe convexe sont exactement les barycentres à coefficients positifs des points de X. Le résultat de Carathéodory qui sont montre que si on est dans un espace de dimension n on peut se limiter aux barycentres des familles de $n+1$ points pour obtenir l'enveloppe convexe d'une partie. La seconde question est l'application essentielle de ce résultat.

4.44. Théorème de Carathéodory (1907)

Soit X une partie non vide de $\mathbb{R}^n$ et C(X) son enveloppe convexe.

1. Montrer que tout point de C(X) est barycentre à coefficients positifs d'une famille de $n+1$ points de X.

2. En déduire que si X est compacte, alors C(X) est compacte.

(École polytechnique)

↪ **Solution.**

1. Soit $x \in \mathrm{C}(\mathrm{X})$. On sait que l'on peut écrire x comme barycentre à coefficients positifs de points de X : $x = \sum_{k=1}^{p} \lambda_k x_k$ avec $\sum_{k=1}^{p} \lambda_k = 1$ et $\lambda_k \geqslant 0$ pour tout k, les x_k étant des points de X. On va supposer que $p > n+1$ et montrer que l'on peut écrire x comme barycentre de $p-1$ points. Par itération, cela prouve bien le résultat demandé.

On va chercher d'autres combinaisons barycentriques de $x_1, \ldots, x_p$ qui valent x et montrer que parmi elles, il y en a une où l'un des points x_k a un coefficient nul. Comme par hypothèse $p-1 \geqslant n$, la famille $(x_2 - x_1, x_3 - x_1, \ldots, x_p - x_1)$ est liée. Il existe donc des réels non tous nuls $\alpha_2, \alpha_3, \ldots, \alpha_p$ vérifiant $\sum_{i=2}^{p} \alpha_i (x_i - x_1) = 0$. Posons $\alpha_1 = -\sum_{i=2}^{p} \alpha_i$. On a alors $\sum_{i=1}^{p} \alpha_i x_i = 0$, d'où l'on déduit que pour tout réel t,

$$x = \sum_{i=1}^{p} (\lambda_i + t\alpha_i) x_i.$$

On va chercher à choisir t de manière à annuler l'un des coefficients (au moins) mais en gardant tous les autres positifs (la somme vaut toujours 1 car $\sum_{i=1}^{p} \alpha_i = 0$). Comme les α_i ne sont pas tous nuls et de somme nulle, il existe nécessairement $i \in [\![1, p]\!]$ tel que $\alpha_i < 0$.

Soit $\tau = \min\left\{-\frac{\lambda_i}{\alpha_i}, \alpha_i < 0\right\} > 0$. Posons $\mu_i = \lambda_i + \tau\alpha_i$, pour tout $i \in [\![1, p]\!]$. Par construction, les μ_i sont positifs ou nul, de somme égale à 1 et il existe (au moins) un indice j tel que $\mu_j = 0$ (tout indice qui correspond au maximum). On obtient $x = \sum_{i \neq j} \mu_i x_i$ et on a écrit x comme combinaison convexe de $p - 1$ points de X.

2. Posons $\Lambda = \left\{(\lambda_1, \ldots, \lambda_{n+1}) \in (\mathbb{R}_+)^{n+1}, \sum_{k=1}^{n+1} \lambda_k = 1\right\}$. Il s'agit clairement d'une partie compacte de $\mathbb{R}^{n+1}$. La question qui précède montre que $C(X)$ est l'image de $\Lambda \times X^{n+1}$ par l'application f qui à $(\lambda_1, \ldots, \lambda_{n+1}, x_1, \ldots, x_{n+1}) \in \Lambda \times X^{n+1}$ associe $\sum_{k=1}^{n+1} \lambda_k x_k$. Cette application étant clairement continue, et l'ensemble source étant compact en tant que produit de compacts, $C(X)$ est compact. ◁

On sera amené à utiliser le théorème de Carathéodory dans les deux dernières questions de l'exercice suivant. La première question est un exercice très classique et très souvent posé à l'oral de tous les concours.

4.45. Théorème de Jung (1901)

Soit K une partie bornée de $\mathbb{R}^n$ euclidien.

1. Montrer qu'il existe une unique boule fermée $B(b_K, r_K)$ de rayon minimal contenant K.

2. Soit G le groupe formé par les isométries affines f de $\mathbb{R}^n$ telles que $f(K) = K$. Montrer qu'il existe un point de $\mathbb{R}^n$ fixe par tous les éléments de G.

3. Montrer que si K est compact, alors b_K est dans l'enveloppe convexe de $K \cap S(b_K, r_K)$, où $S(b_K, r_K)$ est la sphère de centre b_K et de rayon r_K.

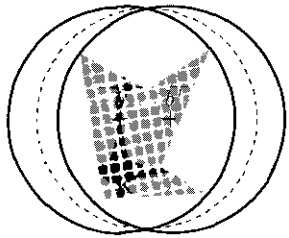
4. On prend $n = 2$. Montrer que $r_K \leqslant \frac{\delta(K)}{\sqrt{3}}$ où $\delta(K)$ désigne le diamètre de K.

(École normale supérieure)

▷ **Solution.**

1. Posons $A = \{r \in \mathbb{R}_+,\ \exists x \in \mathbb{R}^n, K \subset B(x,r)\}$: c'est l'ensemble des rayons des boules fermées qui contiennent K. Comme K est supposée bornée, A n'est pas vide et comme il est minoré par 0, il admet une borne inférieure $r_K \geqslant 0$. Soit $C = \{x \in \mathbb{R}^n,\ K \subset B(x, r_K)\}$. La question posée consiste à prouver que C est un singleton. Par définition de r_K, il existe pour tout $k \in \mathbb{N}^*$ un point $x_k \in \mathbb{R}^n$ tel que $K \subset B\left(x_k, r_K + \frac{1}{k}\right)$. Montrons que la suite $(x_k)_{k\geqslant 1}$ est bornée. Soit $a \in K$. On a pour tout k, $\|x_k - a\| \leqslant r_K + \frac{1}{k}$ et donc $\|x_k\| \leqslant \|a\| + r_K + \frac{1}{k} \leqslant \|a\| + r_K + 1$. La suite $(x_k)_{k\geqslant 1}$ est bornée et on peut donc en extraire une sous-suite convergente $(x_{\varphi(k)})_{k\geqslant 1}$ ($\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ strictement croissante), d'après le théorème de Bolzano-Weierstrass. Notons b sa limite.

Pour tout $x \in K$, et tout $k \geqslant 1$, on a $\|x - x_{\varphi(k)}\| \leqslant r_K + \frac{1}{\varphi(k)}$ donc, en passant à la limite, on en déduit que $\|x - b\| \leqslant r_K$. On a donc montré que $K \subset B(b, r_K)$. Par conséquent $b \in C$ et C n'est pas vide. Il est assez clair géométriquement que C ne peut pas contenir deux points.



Supposons en effet que C contienne un autre point b' distinct de b. D'après l'égalité de la médiane, on a, pour tout $x \in K$,

$$\begin{aligned}\left\|x - \frac{b+b'}{2}\right\|^2 &= \frac{1}{2}\|x-b\|^2 + \frac{1}{2}\|x-b'\|^2 - \frac{1}{4}\|b-b'\|^2 \\ &\leqslant r_K^2 - \frac{1}{4}\|b-b'\|^2 < r_K^2\end{aligned}$$

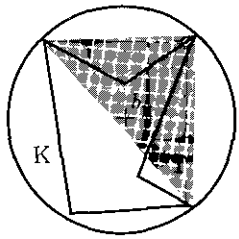
et K est inclus dans une boule fermée de centre $\frac{b+b'}{2}$ et de rayon strictement inférieur à r_K ; c'est contraire à la définition de r_K.

Conclusion. L'ensemble C est un singleton dont on notera b_K l'unique élément. On a bien démontré que K est inclus dans une unique boule fermée de rayon minimal.

2. Cette question offre une première application du résultat précédent. Soit f un élément de G. On a $K \subset B(b_K, r_K)$ et donc $f(K) \subset f(B(b_K, r_K))$. Par hypothèse, $f(K) = K$ et, f étant une isométrie affine, $f(B(b_K, r_K))$ est la boule de centre $f(b_K)$ et de rayon r_K. Par l'unicité

établie dans la question précédente, on a $f(b_K) = b_K$. Le point b_K est donc invariant par tous les éléments de G.

3. Comme K est supposé compact, l'ensemble $K \cap S(b_K, r_K)$ est aussi compact. Il est clairement non vide, sans quoi on pourrait trouver une boule de centre b_K et de rayon strictement inférieur à r_K qui contienne toujours K (car la fonction continue $x \longmapsto \|x - b_K\|$ atteint son maximum sur le compact K). D'après, la seconde question de l'exercice 4.44, l'enveloppe convexe de $K \cap S(b_K, r_K)$, que l'on notera T est également compacte. Voici une figure dans le cas du plan (K est un polygone et T le triangle grisé).



Supposons par l'absurde que b_K n'appartient pas à T. Par compacité de T, il existe $y \in T$ tel que $\|b_K - y\| = d(b_K, T)$. Pour tout $t \in [0, 1]$ et $a \in T$, $ta + (1-t)y \in T$ par convexité de T. On a donc

$$\|b_K - (ta + (1-t)y)\| = \|b_K - y + t(y-a)\| \geqslant \|b_K - y\|,$$

c'est-à-dire, en élevant au carré et en développant,

$$t^2\|y-a\|^2 + 2t\langle b_K - y, y - a\rangle \geqslant 0.$$

En divisant par t puis en faisant tendre t vers 0, on obtient, pour tout $a \in T$, $\langle b_K - y, y - a\rangle \geqslant 0$.

Cette inégalité signifie que T *est inclus dans le demi-espace ne contenant pas* b_K, *limité par l'hyperplan* H *orthogonal à la droite* $(b_K y)$ *passant par* y. *On appelle* H *un hyperplan d'appui de* T *en* y.

Montrons que cette inégalité est contradictoire avec la définition de b_K. Posons $u = y - b_K \neq 0$. Pour tout $n \in \mathbb{N}^*$, la boule de centre $b_K + \frac{1}{n}u$ et de rayon r_K ne contient pas K. Il existe donc $a_n \in K$ tel que

$$d(b_K + \frac{1}{n}u, a_n) > r_K \geqslant d(b_K, a_n).$$

Cette inégalité s'écrit $\|b_K - a_n + \frac{1}{n}u\| \geqslant \|b_K - a_n\|$ et implique donc $2\langle b_K - a_n, u\rangle + \frac{1}{n}\|u\|^2 \geqslant 0$. La suite $(a_n)_{n\geqslant 1}$ du compact K possède une

sous-suite $(a_{\varphi(n)})_{n\in\mathbb{N}^*}$ qui converge vers $a \in \mathrm{K}$. On a alors, pour tout $n \in \mathbb{N}^*$,

$$\begin{cases} d(b_\mathrm{K} + \dfrac{1}{\varphi(n)}u, a_{\varphi(n)}) > r_\mathrm{K} \geqslant d(x, a_{\varphi(n)}) \\ 2\langle b_\mathrm{K} - a_{\varphi(n)}, u\rangle + \dfrac{1}{\varphi(n)}\|u\|^2 \geqslant 0 \end{cases}.$$

En faisant tendre n vers $+\infty$, on obtient $d(b_\mathrm{K}, a) = r_\mathrm{K}$ et $\langle b_\mathrm{K} - a, u\rangle \geqslant 0$. Ainsi a appartient à $\mathrm{S}(x, r_\mathrm{K}) \cap \mathrm{K}$ et donc à T. Mais on a

$$\langle b_\mathrm{K} - y, y - a\rangle = \langle -u, u + b_\mathrm{K} - a\rangle = -\|u\|^2 - \langle u, b_\mathrm{K} - a\rangle < 0,$$

puisque $\|u\| = \|y - b_\mathrm{K}\| > 0$ et $\langle u, b_\mathrm{K} - a\rangle \geqslant 0$. Cela est contradictoire avec l'inégalité démontrée plus haut. On peut conclure : b_K appartient à l'enveloppe convexe de $\mathrm{S}(b_\mathrm{K}, r_\mathrm{K}) \cap \mathrm{K}$.

4. De la question précédente et du théorème de Carathéodory, il résulte qu'il existe $(x_1, x_2, x_3) \in \mathrm{S}(b_\mathrm{K}, r_\mathrm{K}) \cap \mathrm{K}$ tel que b_K soit combinaison convexe de (x_1, x_2, x_3). Le point b_K est alors à l'intérieur du triangle $x_1x_2x_3$. La somme des trois angles géométriques $\widehat{x_1b_\mathrm{K}x_2}$, $\widehat{x_2b_\mathrm{K}x_3}$ et $\widehat{x_3b_\mathrm{K}x_1}$ est donc égale à 2π et l'un de ces angles est donc supérieur à $\dfrac{2\pi}{3}$. Supposons par exemple que c'est $\widehat{x_1b_\mathrm{K}x_2}$. On a alors

$$d(x_1, x_2) = 2r_\mathrm{K}\sin\left(\frac{\widehat{x_1b_\mathrm{K}x_2}}{2}\right) \geqslant 2r_\mathrm{K}\sin\frac{\pi}{3} \geqslant r_\mathrm{K}\sqrt{3}.$$

Puisque x_1 et x_2 appartiennent à K, on a *a fortiori* $\delta(\mathrm{K}) \geqslant r_\mathrm{K}\sqrt{3}$.

Conclusion. On a la majoration $r_\mathrm{K} \leqslant \dfrac{\delta(\mathrm{K})}{\sqrt{3}}$.

Notons que la constante $\dfrac{1}{\sqrt{3}}$ *est la meilleure possible car, si* K *est un triangle équilatéral de côté* a*, on a* $\delta(\mathrm{K}) = a$ *et* r_K *est le rayon du cercle circonscrit, à savoir* $r_\mathrm{K} = \dfrac{2}{3}\cdot\dfrac{a\sqrt{3}}{2} = \dfrac{\delta(\mathrm{K})}{\sqrt{3}}$*. Dans* $\mathbb{R}^n$*, on peut montrer que* $r_\mathrm{K} \leqslant \delta(\mathrm{K})\sqrt{\dfrac{n}{2(n+1)}}$*, l'égalité étant obtenue pour un simplexe régulier.*

4.46. Points extrémaux d'un ensemble convexe

Si E est un $\mathbb{R}$-espace vectoriel et C une partie convexe non vide de E, on note Ext(C) l'ensemble des x de C tels que $\mathrm{C} \setminus \{x\}$ soit encore convexe (un tel point de C est dit *extrémal*). Pour $p \in \mathbb{N}^*$ et $x = (x_1, \ldots, x_p) \in \mathbb{R}^p$, on écrit $x \geqslant 0$ si $x_i \geqslant 0$ pour tout $i \in [\![1.p]\!]$.

On fixe $(m, n) \in (\mathbb{N}^*)^2$, $\mathrm{A} \in \mathcal{M}_{m,n}(\mathbb{R})$ et $\delta \in \mathbb{R}^m$.

1. Soient $C = \{x \in \mathbb{R}^n,\ x \geqslant 0 \text{ et } Ax = \delta\}$ qu'on suppose non vide, $c = (c_1, \ldots c_n) \in C$ et $J = \{j \in \{1, \ldots, n\},\ c_j \neq 0\}$. Si $1 \leqslant j \leqslant n$, on note A_j la j-ième colonne de A.

Montrer l'équivalence entre :

(i) $c \in \text{Ext}(C)$;

(ii) la famille $(A_j)_{j\in J}$ est libre.

2. On considère $D_1 = \{x \in \mathbb{R}^n,\ x \geqslant 0 \text{ et } \delta - Ax \geqslant 0\}$ et $D_2 = \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^m,\ x \geqslant 0,\ y \geqslant 0 \text{ et } y + Ax = \delta\}$. Soit $x \in D_1$.

Montrer l'équivalence entre :

(i) $x \in \text{Ext}(D_1)$;

(ii) il existe $y \in \mathbb{R}^m$ tel que $(x, y) \in \text{Ext}(D_2)$.

(**École normale supérieure**)

▷ **Solution.**

Un point c de C n'est pas extrémal si, et seulement si, $C \setminus \{x\}$ n'est pas convexe, c'est-à-dire s'il existe $(y, z) \in C \setminus \{x\}$ et $\lambda \in]0, 1[$ tels que $\lambda y + (1 - \lambda)z = x$. A contrario x est extrémal si, et seulement si, pour tout $(y, z) \in C^2$ et $\lambda \in]0, 1[$, $\lambda y + (1 - \lambda)z = x$ implique $y = z = x$.

1. Notons que C est clairement convexe comme intersection de parties convexes. On remarque que si $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, $Ax = \sum_{j=1}^{n} x_j A_j$.

• Supposons que $c \in \text{Ext}(C)$. Montrons que la famille $(A_j)_{j\in J}$ est libre, en raisonnant par l'absurde. Soit $(\lambda_j)_{j\in J}$ une famille non nulle telle que $\sum_{j\in J} \lambda_j A_j = 0$. Considérons $u = (u_1, \ldots, u_n) \in \mathbb{R}^n$ défini par $u_j = \lambda_j$ si $j \in J$ et $u_j = 0$ sinon. Par construction $Au = \sum_{j=1}^{n} u_j A_j = \sum_{j\in J} \lambda_j A_j = 0$. Pour $t > 0$, posons $x = c - tu$ et $y = c + tu$. Montrons que pour t assez petit, x et y appartiennent à C.

On sait que $Ax = Ac - tAu = Ac = \delta$ et de même $Ay = \delta$. Il suffit de vérifier que x et y sont positifs. Si $j \notin J$ alors $u_j = c_j = 0$ donc $x_j = y_j = 0$. Si $j \in J$, alors $c_j > 0$ et x_j et y_j sont positifs si $t|\lambda_j| \leqslant c_j$. Il suffit de prendre $t \max_{j\in J} |\lambda_j| \leqslant \min_{j\in J}(c_j)$. Par construction $\max_{j\in J} |\lambda_j| > 0$ et $\min_{j\in J}(c_j) > 0$, donc il suffit de prendre $0 < t < t_0$ où $t_0 = \dfrac{\min_{j\in J}(c_j)}{\max_{j\in J} |\lambda_j|}$.

Pour un tel t, on a x et y dans $C \setminus \{c\}$, car $tu \neq 0$, et $c = \dfrac{1}{2}(x + y)$. Cela contredit le fait que c est extrémal. La famille $(A_j)_{j\in J}$ est donc libre.

• Supposons réciproquement que la famille $(A_j)_{j\in J}$ est libre. Considérons $\lambda \in]0, 1[$ et $(x, y) \in C^2$ tels que $c = \lambda x + (1 - \lambda)y$.

Pour tout $j \notin J$, on a $c_j = 0 = \lambda x_j + (1-\lambda)y_j$. Comme $x_j \geqslant 0$ et $y_j \geqslant 0$, on en déduit $\lambda x_j = (1-\lambda)y_j = 0$ et donc $x_j = y_j = 0$. L'égalité $Ax = Ay = \delta$ implique donc $\sum_{j\in J} x_j A_j = \sum_{j\in J} y_j A_j$ et, puisque la famille $(A_j)_{j\in J}$ est libre, $x_j = y_j$ pour tout $j \in J$. On a donc $x = y$ et $c = \lambda x + (1-\lambda)y = x = y$. Le point c est extrémal.

2. • Supposons que $x \in \mathrm{Ext}(D_1)$. S'il existe $y \in \mathbb{R}^m$ tel que l'on ait $(x, y) \in \mathrm{Ext}(D_2)$, alors nécessairement $y = \delta - Ax$. Comme $x \in D_1$, $x \geqslant 0$ et $Ax \leqslant \delta$, donc $y \geqslant 0$ et $(x, y) \in D_2$. Montrons que $(x, y) \in \mathrm{Ext}(D_2)$.

Soit (z, t) et (u, v) deux éléments de D_2 et $\lambda \in]0, 1[$ tels que

$$(x, y) = \lambda(z, t) + (1-\lambda)(u, v) = (\lambda z + (1-\lambda)u, \lambda t + (1-\lambda)v).$$

Par hypothèse $z \geqslant 0$ et $u \geqslant 0$, $Az = \delta - t \leqslant \delta$ et $Au = \delta - v \leqslant \delta$, donc z et u appartiennent à D_1. Puisque x est un point extrémal de D_1, de $\lambda \in]0, 1[$ et $x = \lambda z + (1-\lambda)u$, on déduit $z = u = x$, puis $t = \delta - Az = \delta - Ax = y$ et de même $v = y$. Ainsi, $(z, t) = (u, v) = (x, y)$ et $(x, y) \in \mathrm{Ext}(D_2)$.

• Supposons réciproquement qu'il existe $y \in \mathbb{R}^m$ tel que l'on ait $(x, y) \in \mathrm{Ext}(D_2)$. Montrons que $x \in \mathrm{Ext}(D_1)$. Par hypothèse $y = \delta - Ax$. Soit $(z, u) \in D_1^2$ et $\lambda \in]0, 1[$ tel que $x = \lambda z + (1-\lambda)u$. Posons $t = \delta - Az$ et $v = \delta - Au$. On a alors (z, t) et (u, v) dans D_2. Comme

$$\begin{aligned}\lambda t + (1-\lambda)v &= \lambda(\delta - Az) + (1-\lambda)(\delta - Au)\\ &= \delta - A(\lambda z + (1-\lambda)v) = \delta - Ax = y,\end{aligned}$$

on a $(x, y) = \lambda(z, t) + (1-\lambda)(u, v)$ et comme (x, y) est un élément extrémal de D_2, on en déduit $(z, t) = (u, v) = (x, y)$ et donc $z = u = x$. Cela montre que $x \in \mathrm{Ext}(D_1)$. ◁

Un théorème de Krein-Milman affirme qu'un convexe compact de $\mathbb{R}^n$ est toujours l'enveloppe convexe de l'ensemble de ses points extrémaux. Un carré plein de $\mathbb{R}^2$ par exemple est l'enveloppe convexe de ses quatre sommets. Le lecteur trouvera ce résultat dans le tome 3 d'Analyse.

4.47. Rectangle circonscrit à un convexe compact

Soit K un compact convexe non vide de $\mathbb{R}^2$, δ le diamètre de K.

1. Montrer qu'il existe un rectangle ABCD contenant K dont chaque côté contient au moins un point de K et tel que $AB = \delta$. On pose $BC = h$.

2. Montrer que l'aire de K est supérieure ou égale à $\dfrac{\delta h}{2}$.

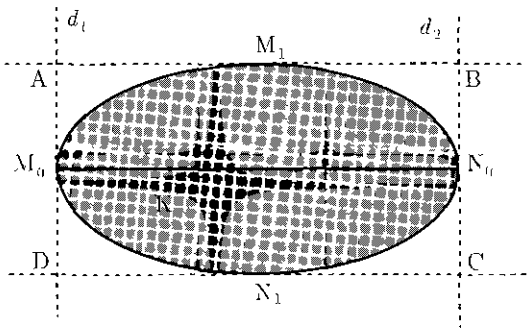
3. Montrer que, pour tout point P de $\mathbb{R}^2$, la valeur moyenne de la distance de P aux points de K est supérieure ou égale à $\frac{\delta}{12}$.

(École normale supérieure)

▷ **Solution.**

On supposera que $\delta > 0$, sinon K est réduit à un point et tous les résultats sont évidents.

1. L'application $(M, N) \longmapsto MN$ est continue sur K^2 qui est compact, donc elle atteint son maximum : il existe $(M_0, N_0) \in K^2$ tel que $M_0N_0 = \delta$. Considérons la projection orthogonale p sur la droite (M_0N_0). L'application p est continue donc l'image $p(K)$ est un compact de la droite (M_0N_0) ; comme p est affine, c'est aussi un convexe. C'est donc un segment, qui contient $[M_0N_0]$. Comme une projection orthogonale est de norme inférieure ou égale à 1, on a, pour tout $(M, N) \in K^2$, $p(M)p(N) \leqslant MN \leqslant \delta \leqslant M_0N_0$. Ainsi, le segment $p(K)$ est de longueur inférieure ou égale à δ donc $p(K) = [M_0N_0]$. Notons d_1 et d_2 les droites passant par M_0 et N_0 respectivement et orthogonales à (M_0N_0). L'ensemble K est contenu dans la partie du plan limitée par d_1 et d_2.



Si on note q la projection orthogonale sur d_1, on trouve comme précédemment que $q(K)$ est un segment que l'on note $[AD]$. Notons B et C les projetés orthogonaux de A et D sur d_2. Par construction, K est contenu dans la partie du plan limitée par (AB) et (CD) donc est à l'intérieur du rectangle ABCD. Les côtés (AD) et (BC) contiennent respectivement les points M_0 et N_0 de K. Comme A et D appartiennent à $q(K)$, il existe M_1 et N_1 dans K tels que $A = q(M_1)$ et $D = q(N_1)$. Les points M_1 et N_1 appartiennent respectivement à (AB) et (CD). Les quatre côtés du rectangle contiennent donc chacun au moins un point de K. Enfin $AB = M_0N_0 = \delta$.

2. Avec les notations de la première question, les points M_0, N_0, M_1 et N_1 appartiennent à K donc le quadrilatère convexe $M_0M_1N_0N_1$ est inclus dans K. En notant μ l'aire, on obtient

$$\begin{aligned}\mu(K) &\geqslant \mu(M_0M_1N_0N_1) \geqslant \mu(M_0M_1N_0) + \mu(M_0N_0N_1)\\ &\geqslant \frac{AM_0 \cdot M_0N_0}{2} + \frac{DM_0 \cdot M_0N_0}{2} \geqslant \frac{AD \cdot M_0N_0}{2} \geqslant \frac{\delta h}{2}.\end{aligned}$$

L'énoncé laisse supposer que l'aire de K *est définie, ce que nous avons admis. En toute rigueur, cela doit être démontré. On peut raisonner ainsi. On munit le plan d'un repère orthonormal d'origine* (M_0, i, j)*, où* i *et* j *sont colinéaires et de même sens que* $\overrightarrow{M_0N_0}$ *et* $\overrightarrow{M_0A}$*. Pour tout* $x \in [0, \delta]$*, on considère la droite passant par le point* M *de coordonnées* $(x, 0)$ *et orthogonale à* (AB)*. Elle coupe* K *selon un segment, ensemble des points d'abscisse* x *dont l'ordonnée varie entre* $f(x)$ *et* $g(x)$ *(*$f(x) \leqslant g(x)$*). On définit ainsi deux applications de* $[0, \delta]$ *dans* $\mathbb{R}$ *et* $K = \{(x, y) \in [0.\delta] \times \mathbb{R},\ f(x) \leqslant y \leqslant g(x)\}$*. Pour définir l'aire de* K*, il suffit de démontrer que les applications* f *et* g *sont continues, car alors* $\mu(\mathcal{C}) = \displaystyle\int_0^\delta (g(x) - f(x))\, dx$*.*

La fonction f *est convexe et la fonction* g *concave, car* K *étant convexe, pour* $(x, x') \in [0, \delta]^2$ *et* $\lambda \in [0, 1]$*, les points de coordonnées* $(\lambda x + (1-\lambda)x', \lambda f(x) + (1-\lambda)f(x'))$ *et* $(\lambda x + (1-\lambda)x', \lambda g(x) + (1-\lambda)g(x'))$ *appartiennent à* K *donc*

$$\begin{aligned}f(\lambda x + (1-\lambda)x') &\leqslant \lambda f(x) + (1-\lambda)f(x') \leqslant \lambda g(x) + (1-\lambda)g(x')\\ &\leqslant g(\lambda x + (1-\lambda)x').\end{aligned}$$

Les fonctions f *et* g *sont donc continues sur* $]0, \delta[$*.*

Notons que $f(0) = g(0) = 0$*, car sinon* [AD] *contiendrait un point* M *de* K *différent de* M_0 *et on aurait* $MN_0 > \delta$*. De même,* $f(\delta) = g(\delta) = 0$*. Soit* (x_n) *une suite de* $[0, \delta]$ *tendant vers* 0*. La suite* $(f(x_n))$ *est bornée. Si* y *est une valeur d'adhérence de cette suite,* $(0, y)$ *est une valeur d'adhérence de la suite* $((x_n. f(x_n))$ *donc* M *de coordonnées* $(0, y)$ *appartient à* K *car* K *est fermé, donc* $y = 0$*. La suite* $(f(x_n))$ *qui est bornée et possède une seule valeur d'adhérence* 0 *converge vers* 0*. Ainsi* f *est continue en* 0*. On montre de même que* g *est continue en* 0 *et que* f *et* g *sont continues en* δ*.*

3. La remarque qui terminait la question précédente justifie aussi l'existence de cette valeur moyenne $m(P)$. On a, pour tout point P du plan de coordonnées (x_0, y_0),

$$m(\mathrm{P}) = \frac{\iint_{\mathrm{K}} \mathrm{MP}\,\mathrm{d}x\,\mathrm{d}y}{\mu(\mathrm{K})} = \frac{\iint_{\mathrm{K}} \sqrt{(x-x_0)^2+(y-y_0)^2}\,\mathrm{d}x\,\mathrm{d}y}{\mu(\mathrm{K})}$$
$$\geqslant \frac{\iint_{\mathrm{K}} |x-x_0|\,\mathrm{d}x\,\mathrm{d}y}{\mu(\mathrm{K})}.$$

Comme K est inclus dans le rectangle ABCD, on a $\mu(\mathrm{K}) \leqslant h\delta$. D'autre part, K contient l'intérieur $\mathcal{D}$ du quadrilatère $\mathrm{M_0M_1N_0N_1}$. On a donc $m(\mathrm{P}) \geqslant \frac{1}{h\delta} \iint_{\mathcal{D}} |x-x_0|\,\mathrm{d}x\,\mathrm{d}y$.

Notons h_1 l'ordonnée de A et h_2 celle de D. On a donc $h_1 - h_2 = h$. On fixe $y \in [h_2, h_1]$; les points de $\mathcal{D}$ d'ordonnée y ont une abscisse qui varie entre a et b. Il faut minorer $\displaystyle\int_a^b |x-x_0|\,\mathrm{d}x$. Par un changement de variable, on obtient $\displaystyle\int_a^b |x-x_0|\,\mathrm{d}x = \int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} \left|x - x_0 + \frac{a+b}{2}\right| \mathrm{d}x$.

L'application $\varphi : z \longmapsto \displaystyle\int_{-\frac{b-a}{2}}^{\frac{b-a}{2}} |x-z|\,\mathrm{d}x$ est convexe et impaire sur $\mathbb{R}$. On a donc, pour tout $z \in \mathbb{R}$, $\varphi(z) = \dfrac{\varphi(z)+\varphi(-z)}{2} \geqslant \varphi(0)$. Comme $\varphi(0) = \dfrac{(b-a)^2}{4}$, on obtient $\displaystyle\int_a^b |x-x_0|\,\mathrm{d}x \geqslant \frac{(b-a)^2}{4}$.

Par le théorème de Thalès, on a pour y dans $[0, h_1]$, $\dfrac{b-a}{\delta} = \dfrac{h_1-y}{h_1}$ et donc $\dfrac{(b-a)^2}{4} = \dfrac{\delta^2(h_1-y)^2}{4h_1^2}$. On obtient de même, pour y dans $[h_2, 0]$, $\dfrac{(b-a)^2}{4} = \dfrac{\delta^2(h_2-y)^2}{4h_2^2}$. On en déduit que

$$\iint_{\mathcal{D}} |x-x_0|\,\mathrm{d}x\,\mathrm{d}y \geqslant \int_{h_2}^0 \frac{\delta^2(h_2-y)^2}{4h_2^2}\,\mathrm{d}x + \int_0^{h_1} \frac{\delta^2(h_1-y)^2}{4h_1^2}\,\mathrm{d}y$$
$$\geqslant -\frac{\delta^2 h_2}{12} + \frac{\delta^2 h_1}{12} \geqslant \frac{\delta^2 h}{12}.$$

On a donc $\boxed{m(\mathrm{P}) \geqslant \dfrac{\delta}{12}}$. ◁

Les exercices suivants, assez divers, peuvent être rangés dans la catégorie géométrie discrète, mettant en jeu des points variant dans des ensembles discrets et en particuliers des points à coordonnées entières.

4.48. Théorème de Pick (1899)

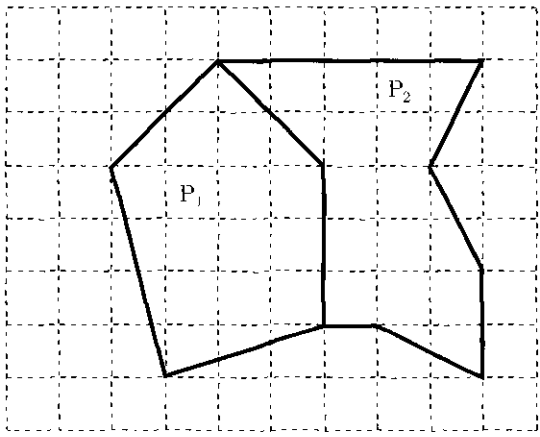
Soit P un polygone (plein) du plan dont les sommets sont à coordonnées entières. On note A l'aire de ce polygone, I le nombre de points à coordonnées entières qui sont strictement à l'intérieur de P et F le nombre de points à coordonnées entières qui sont sur la frontière de P. Montrer que $A = I + \frac{F}{2} - 1$.

(École normale supérieure)

▷ **Solution.**

Il est naturel de commencer par regarder quelques exemples simples pour se convaincre du résultat. Prenons pour P un carré horizontal de côté $n \geqslant 2$ dont les sommets sont $(0,0)$ $(n,0)$ $(0,n)$ et (n,n). On a $A = n^2$, $F = 4(n+1) - 4 = 4n$ (les sommets sont comptés deux fois) et $I = (n-1)^2$. On vérifie effectivement que $(n-1)^2 + 2n - 1 = n^2$.

Pour établir le résultat dans le cas général, l'idée consiste à découper le polygone P en polygones plus simples. On ne va pas prendre des carrés du type ci-dessus, mais des triangles. Il suffira alors d'établir la formule pour un triangle, car on observe que la formule est additive au sens suivant : supposons que P_1 et P_2 soient deux polygones ayant une ou plusieurs arêtes *consécutives* en commun mais dont les intérieurs sont disjoints.



On note P le polygone $P_1 \cup P_2$ et A_1, I_1, F_1 (resp. A_2, I_2, F_2) l'aire,

le nombre de points intérieurs et sur la frontière de P_1 (resp. P_2). On a bien entendu $A = A_1 + A_2$. *Notons x le nombre de points qui sont sur* la frontière commune à P_1 et P_2 et dans l'intérieur de $P_1 \cup P_2$. On a clairement

$$I = I_1 + I_2 + x \quad \text{et} \quad F = F_1 + F_2 - 2x - 2$$

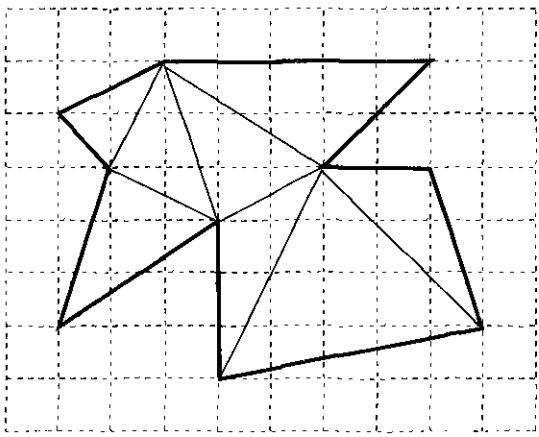
de sorte que

$$I + \frac{F}{2} - 1 = I_1 + I_2 + x + \frac{F_1 + F_2 - 2x - 2}{2} - 1 = I_1 + \frac{F_1}{2} - 1 + I_2 + \frac{F_2}{2} - 1.$$

Donc si la formule est vraie pour P_1 et P_2, elle est aussi vraie pour P.

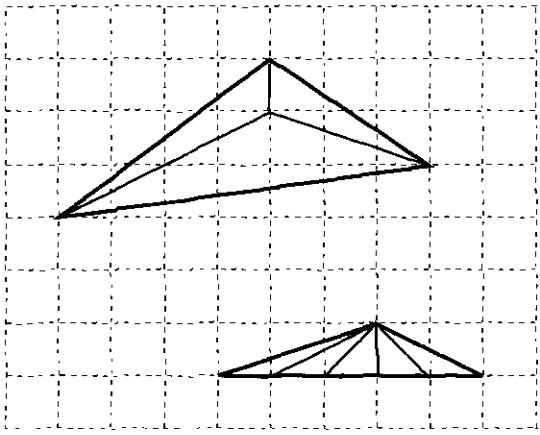
Montrons maintenant que tout polygone P peut être décomposé en triangles à sommets dans $\mathbb{Z}^2$, et même en triangles élémentaires, c'est-à-dire qui n'ont aucun point à coordonnées entières dans leur intérieur ou au bord hormis les sommets. Pour cela, on procède de la manière suivante.

On commence par trianguler le polygone P n'importe comment (il est aisé de montrer par récurrence sur le nombre de sommets que tout polygone admet une triangulation) :



Si un des triangles contient un point à coordonnées entières dans son intérieur, on le relie aux trois sommets. On continue cela tant qu'il reste des points intérieurs à coordonnées entières. Enfin, pour un triangle sans point intérieur à coordonnées entières, on regarde s'il y a des points à coordonnées entières sur un des bords en dehors des sommets. Si tel est le cas, on relie le sommet opposé à tous les points entiers du côté en question.

Bref, tout polygone P admet une triangulation en triangles élémentaires. Comme P est obtenu en assemblant ces triangles élémentaires avec 1 ou 2 arêtes en commun, il suffit, d'après l'additivité démontrée



ci-dessus, de vérifier que la formule est vraie pour un triangle élémentaire T. Comme pour un tel triangle on a I = 0 et F = 3, on est ramené à montrer que l'aire de T est égale à $\frac{1}{2}$. Quitte à effectuer une translation, on peut supposer que l'un des sommets de T est l'origine $(0,0)$. Notons (a,b) et (c,d) les deux autres sommets. Le fait que T soit élémentaire implique que $u=(a,b)$ et $v=(c,d)$ forment une base du réseau $\mathbb{Z}^2$. En effet, si $(x,y)\in\mathbb{Z}^2$, on peut écrire $(x,y)=\alpha u+\beta v$ avec $(\alpha,\beta)\in\mathbb{R}^2$ car (u,v) est une base de $\mathbb{R}^2$. On va voir que α et β sont entiers. On peut écrire $\alpha=\mathrm{E}(\alpha)+r$ et $\beta=\mathrm{E}(\beta)+s$ avec r,s dans $[0,1[$. Le point $(x,y)-\mathrm{E}(\alpha)u-\mathrm{E}(\beta)v=ru+rs$ est dans $\mathbb{Z}^2$ et se trouve dans le parallélogramme de base (u,v). Comme T est élémentaire, ce parallélogramme ne contient aucun point à coordonnées entières en dehors de ses sommets. Ainsi, $r=s=0$. Il en résulte que la matrice $\begin{pmatrix} a & c \\ b & d \end{pmatrix}$ est dans $\mathrm{GL}_2(\mathbb{Z})$. Son déterminant est donc égal à ± 1. Or ce déterminant représente l'aire algébrique du parallélogramme défini par u et v. L'aire de T est donc égale à $\frac{1}{2}$. ◁

La formule ne convient plus lorsque le polygone a des trous : au lieu de 1 *il faut retrancher en fait la caractéristique d'Euler du polygone.*

4.49. Droites à une distance strictement positive de $\mathbb{Z}^2$

Déterminer l'ensemble des droites affines D de $\mathbb{R}^2$ telles que l'ensemble des distances des points de $\mathbb{Z}^2$ à D admette une borne inférieure non nulle.

(École normale supérieure)

▷ **Solution.**

Soit D une droite affine du plan $\mathbb{P}^2$.

• Si D a pour équation $x = x_0$, elle convient si et seulement si $x_0 \notin \mathbb{Z}$ et la borne inférieure de la distance des points de D à $\mathbb{Z}^2$ est la distance de x_0 à $\mathbb{Z}$. De même si D a pour équation $y = y_0$, il faut et il suffit que $y_0 \notin \mathbb{Z}$.

• Supposons dorénavant que D a pour équation $y = ax + b$, avec $(a, b) \in \mathbb{R}^* \times \mathbb{R}$. Nous savons que la distance d'un point (m, n) à D est égale à $\dfrac{|am + b - n|}{\sqrt{1 + a^2}}$. On cherche donc les couples (a, b) pour lesquels il existe $c > 0$ tel que

$$\forall (m, n) \in \mathbb{Z}^2, \quad (am + b - n)^2 \geqslant c(1 + a^2).$$

Nous savons que, α et β étant deux réels non nuls, $\alpha\mathbb{Z} + \beta\mathbb{Z}$ est un sous-groupe de $\mathbb{R}$ qui est dense dans $\mathbb{R}$ si, et seulement si, $\dfrac{\beta}{\alpha}$ est irrationnel.

Si $a \notin \mathbb{Q}$, alors $a\mathbb{Z} + \mathbb{Z}$ est dense dans $\mathbb{R}$. En particulier, pour tout $c > 0$, il existe $(m, n) \in \mathbb{Z}^2$ tel que $(b - (-am + n))^2 < c(1 + a^2)$. La droite D ne convient pas.

Si $a \in \mathbb{Q}$, on écrit $a = \dfrac{p}{q}$ avec p et q premiers entre eux et on a alors $a\mathbb{Z} + \mathbb{Z} = \dfrac{1}{q}(p\mathbb{Z} + q\mathbb{Z}) = \dfrac{1}{q}\mathbb{Z}$. On veut donc avoir, pour tout $m \in \mathbb{Z}$, $\left(\dfrac{m}{q} + b\right)^2 \geqslant c(1 + a^2)$, *i.e.* $(m + bq)^2 \geqslant c(1 + a^2)q^2$. Le réel $c > 0$ existe si, et seulement si, bq n'appartient pas à $\mathbb{Z}$. Alors $|m + bq|$ est minoré par la distance de bq à $\mathbb{Z}$.

Conclusion. La distance de la droite affine D aux points de $\mathbb{Z}^2$ est minorée par une constante strictement positive si D possède une équation de la forme $x = x_0$ avec $x_0 \notin \mathbb{Z}$, $y = y_0$ avec $y_0 \notin \mathbb{Z}$ ou $y = ax + b$ avec $a \in \mathbb{Q}^*$, $a = \dfrac{p}{q}$ (p et q premiers entre eux) et $bq \notin \mathbb{Z}$. ◁

4.50. Polygone ayant trois sommets à coordonnées entières

Soit P un polygone convexe régulier de $\mathbb{R}^2$. On suppose que deux sommets consécutifs de P et un troisième sont à coordonnées entières. Montrer que P est un carré.

(**École normale supérieure**)

▷ **Solution.**

On suppose que P est un polygone régulier à n côtés ($n \geqslant 3$). On note $M_1, \ldots, M_n$ les sommets du polygone P et I son centre. Il existe $(k, \ell) \in [\![1, n]\!]^2$, avec ℓ distinct de k et $k+1$ tel que M_k, M_{k+1} et M_ℓ aient

des coordonnées entières. Les coordonnées du milieu de $[M_k, M_{k+1}]$ et du vecteur $\overrightarrow{M_k M_{k+1}}$ sont rationnelles. La médiatrice de $[M_k, M_{k+1}]$ a donc une équation à coefficients rationnels. Il en est de même de la médiatrice de $[M_k, M_\ell]$. Le point I qui est l'intersection de ces deux droites a donc des coordonnées rationnelles. On en déduit que les coordonnées des vecteurs $\overrightarrow{IM_k}$ et $\overrightarrow{IM_{k+1}}$ sont rationnelles. Notons z_k et z_{k+1} les affixes de ces vecteurs. Leur partie réelle et leur partie imaginaire sont rationnelles. On sait que $z_{k+1} = e^{i\frac{2\pi}{n}} z_k$. On en déduit que $e^{i\frac{2\pi}{n}} = \dfrac{z_{k+1}}{z_k} = \dfrac{z_{k+1}\overline{z}_k}{|z_k|^2}$. Cela montre que $\cos\dfrac{2\pi}{n}$ et $\sin\dfrac{2\pi}{n}$ sont des nombres rationnels.

On veut en déduire que $n = 4$, c'est-à-dire que $\cos\dfrac{2\pi}{n} = 0$. Il s'agit d'un exercice relativement classique dont une version plus générale consiste à chercher tous les rationnels r tels que $\cos r\pi$ soit rationnel. Il est clair que $n = 3$ ne convient pas car $\sin\dfrac{2\pi}{3} \notin \mathbb{Q}$. Supposons par l'absurde que $n \geqslant 5$ et posons $\cos\dfrac{2\pi}{n} = \dfrac{p}{q}$, où $p < q$ sont des entiers naturels premiers entre eux. L'idée est d'utiliser les polynômes de Tchebychev. Pour tout entier $m \in \mathbb{N}$ nous savons qu'il existe un unique polynôme T_m de degré m, à coefficients entiers tel que $\cos m\theta = T_m(\cos\theta)$ pour tout réel θ ; de plus T_m a la parité de m. On a donc en particulier,

$$1 = \cos 2\pi = T_n\left(\cos\frac{2\pi}{n}\right) = T_n\left(\frac{p}{q}\right).$$

• Si n est impair, le polynôme T_n n'a pas de terme constant et la relation $q^n = q^n T_n\left(\dfrac{p}{q}\right)$ montre que p divise q^n. Comme p et q sont premiers entre eux, cela impose $p = 1$. On a donc $\cos\dfrac{2\pi}{n} = \dfrac{1}{q} \leqslant \dfrac{1}{2}$. Mais alors $\sin\dfrac{2\pi}{n} = \dfrac{\sqrt{q^2-1}}{q}$ ne peut appartenir à $\mathbb{Q}$ (si $\dfrac{\sqrt{q^2-1}}{q} = \dfrac{p'}{q'}$, avec p' et q' dans $\mathbb{N}^*$, premiers entre eux, on a $\dfrac{q^2-1}{q^2} = \dfrac{p'^2}{q'^2}$ et comme $(q^2-1) \wedge q^2 = 1$, on en déduit $q'^2 = q^2$ et $q^2 - 1 = p'^2$, c'est-à-dire $1 = (q-p')(q+p')$, ce qui impossible car $q + p' \geqslant 3$). On en déduit que n est pair et on pose dans la suite $n = 2n'$ (avec $n' \geqslant 3$).

• Si n' est impair, on écrit de même $-1 = \cos\pi = T_{n'}\left(\dfrac{p}{q}\right)$ et on en déduit comme dans le cas précédent que $p = 1$ et une impossibilité.

• Si n' est divisible par 4, on écrit $n' = 4k$ et l'égalité $\cos\dfrac{\pi}{4} = T_k\left(\dfrac{p}{q}\right)$ montre que $\cos\dfrac{\pi}{4}$ est rationnel, ce qui n'est pas vrai.

• Si enfin $n' = 2k$ avec k impair, $k \geqslant 3$, on a $\cos\dfrac{\pi}{k} = 2\cos^2\dfrac{\pi}{n'} - 1 \in \mathbb{Q}$ et on est ramené à l'étude du cas n' impair ($2k$ remplace $2n'$).

Toutes les valeurs $n \geqslant 5$ conduisent à une impossibilité.

Conclusion. La seule valeur possible est $n = 4$ et le polygone P est un carré. ◁

Le raisonnement précédent permet de montrer que les seuls rationnels $r \in \left[0, \frac{1}{2}\right]$ *(on peut se limiter à cet intervalle par symétries) tels que* $\cos r\pi \in \mathbb{Q}$ *sont* 0, $\frac{1}{3}$ *et* $\frac{1}{2}$.

4.51. Recouvrement par des droites

On note $C_n = ([1, n] \times [1, n]) \cap \mathbb{N}^2$. Pour tout partie finie S de C_n, on note D(S) l'ensemble des points de C_n qui sont sur une droite passant par deux points de S. Enfin, on note μ_n le plus petit cardinal d'une partie S telle que D(S) = C_n. Montrer qu'il existe $\alpha > 0$ tel que pour tout $n \geqslant 1$, $\mu_n \geqslant \alpha n^{2/3}$.

(École normale supérieure)

▷ **Solution.**

• On a $\mu_2 = 4$, car pour $n = 2$ on voit que D(S) = S pour toute partie S (une droite passant par deux points de $[\![1, 2]\!]^2$ ne contient pas d'autre point du carré en dehors des deux points qui la définissent).

• On a $\mu_3 \leqslant 4$, car si on prend les 4 sommets du carré pour S on a D(S) = $[\![1, 3]\!]^2$ (le centre est sur les diagonales). On ne peut pas y arriver avec S de cardinal 3 : en effet, une droite contient au plus trois points du carré. Donc les 3 droites qu'on peut tracer avec 3 points contiennent au maximum 6 points et il y en a 9 dans le carré... On a donc $\mu_3 = 4$.

• Passons à l'étude du cas général. On voit que pour obtenir une minoration de μ_n il faut estimer le nombre maximum de points de C_n qu'on peut avoir sur une droite donnée passant par deux points du carré C_n. Voyons ce que donne une première estimation grossière. Il est clair qu'une droite possède au plus n points. Si S est de cardinal s, on a au plus $C_s^2 = \frac{s(s-1)}{2}$ droites. Ainsi, si D(S) = C_n alors $n^2 \leqslant n\frac{s(s-1)}{2}$. On en déduit que $2n \leqslant s(s-1) \leqslant s^2$ et cela donne $\sqrt{2n} \leqslant \mu_n$. C'est déjà pas mal mais l'énoncé demande mieux ! Pour cela il faut raffiner notre estimation du nombre de points sur une droite.

En fait, il n'y a pas tant de droites que cela qui contiennent n points, et certaines droites ont même seulement deux points (par exemple celle qui passe par $(1, 2)$ et $(2, 1)$). On va étudier cela en fonction de la pente. La pente d'une droite passant par deux points de C_n est soit nulle, soit infinie, soit de la forme $\pm\frac{a}{b}$ avec a, b entiers de $[\![1, n]\!]$. Dans les deux premiers cas la droite contient n points. Étudions ce qu'il en est pour

une droite de pente $\pm\frac{a}{b}$. On suppose a, b premiers entre eux. L'équation d'une telle droite est de la forme $by = \pm ax + c$ avec c entier (puisque la droite contient deux points à coordonnées entières). On connaît la structure des solutions d'une telle équation diophantienne : si (x_0, y_0) est une solution particulière, les autres solutions sont $(x_0 + kb, y_0 \pm ka)$ avec $k \in \mathbb{Z}$. Il résulte de cela que la droite contient forcément moins de $\min(\frac{n}{a}, \frac{n}{b})$ points du carré C_n.

Introduisons quelques notations. Si D est une droite comme ci-dessus, on note $p(D)$ l'entier $\max(a, b)$ (où $\frac{a}{b}$ est le représentant irréductible de la valeur absolue de la pente de la droite). On pose $p(D) = 1$ si D est horizontale ou verticale (pente nulle ou infinie). On vient donc de montrer que pour toute droite D le nombre de points de $D \cap C_n$ est majoré par $\frac{n}{p(D)}$.

Soit S une partie finie de C_n. Pour $k \in [\![1, n]\!]$ on note s_k le nombre de droites D définies par deux points de S telles que $p(D) = k$. Si $D(S) = C_n$, on a forcément $n^2 \leqslant \sum_{k=1}^{n} s_k \frac{n}{k}$ c'est-à-dire $n \leqslant \sum_{k=1}^{n} \frac{s_k}{k}$ (1).

Par ailleurs, on a $\sum_{k=1}^{n} s_k \leqslant C_s^2 = \frac{s(s-1)}{2} \leqslant \frac{s^2}{2}$ (2). On va maintenant estimer s_k. Par un point donné de S, il y a au plus $4k$ droites D telles que $p(D) = k$ qui passent par ce point (car la pente prend au plus $4k$ valeurs différentes) et cela est aussi valable pour $k = 1$. On a donc $s_k \leqslant 4ks$ (3) pour tout k.

On dispose alors de tous les éléments pour conclure. Imaginons que s_1 ne soit pas égal à la valeur maximale $4s$ autorisée par (3). On diminue alors s_2 (ou s_3,...) du nombre manquant et on rajoute cette quantité à s_1 : l'inégalité (2) reste vraie car la somme globale des s_i n'a pas changé et l'inégalité (1) aussi, car la valeur de $\sum_{k=1}^{n} \frac{s_k}{k}$ est augmentée par cette opération. On peut donc supposer que $s_1, s_2, \ldots, s_t$ prennent la valeur maximale autorisée par (3), que s_{t+1} prend le « reste » et que $s_k = 0$ pour $k > t + 1$. On a alors

$$\sum_{k=1}^{t} s_k = \sum_{k=1}^{t} 4ks = 4s\frac{t(t+1)}{2} \leqslant \frac{s^2}{2}$$

de sorte que $4t^2 \leqslant 4t(t+1) \leqslant s$ et donc $2t \leqslant \sqrt{s}$. Enfin,

$$n \leqslant \sum_{k=1}^{t+1} \frac{s_k}{k} \leqslant \sum_{k=1}^{t+1} 4s = 4(t+1)s \leqslant 4(2t)s \leqslant 4s^{3/2}$$

Et on en déduit que $s \geqslant \alpha n^{2/3}$ avec $\alpha = 4^{-2/3}$. Cela vaut pour toute partie S telle que $D(S) = C_n$ et on a donc bien prouvé que $\mu_n \geqslant \alpha n^{2/3}$. ◁

4.52. Rotations de $\mathbb{R}^3$ laissant un réseau invariant

Soit (u, v, w) une base de $\mathbb{R}^3$ et $\Lambda = \mathbb{Z}u + \mathbb{Z}v + \mathbb{Z}w$ le réseau qu'elle engendre. Que peut-on dire d'une rotation r de $\mathbb{R}^3$ telle que $r(\Lambda) = \Lambda$?

(École normale supérieure)

▷ **Solution.**

Soit M la matrice de la famille $(r(u), r(v), r(w))$ dans la base (u, v, w) c'est-à-dire la matrice de r dans la base (u, v, w).

Supposons $r(\Lambda) = \Lambda$, *i.e.* $\mathbb{Z}u + \mathbb{Z}v + \mathbb{Z}w = \mathbb{Z}r(u) + \mathbb{Z}r(v) + \mathbb{Z}r(w)$. Comme $(r(u), r(v), r(w))$ appartiennent à Λ, M appartient à $\mathcal{M}_3(\mathbb{Z})$. Comme $r(\Lambda)$ contient les trois vecteurs de la base (u, v, w), la famille $(r(u), r(v), r(w))$ est génératrice, donc M est inversible. Comme $r^{-1}(\Lambda) = \Lambda$, le même raisonnement s'applique à M^{-1} qui appartient donc à $\mathcal{M}_3(\mathbb{Z})$. La matrice M est donc inversible dans $\mathcal{M}_3(\mathbb{Z})$.

Si réciproquement $\mathrm{M} \in \mathrm{GL}_3(\mathbb{Z})$, on a $r(\Lambda) \subset \Lambda$, puisque M est à coefficients entiers et de même $r^{-1}(\Lambda) \subset \Lambda$, car M^{-1} est également à coefficients entiers. Ainsi, il faut et il suffit que $\mathrm{M} \in \mathrm{GL}_3(\mathbb{Z})$ pour que la rotation r laisse le réseau globalement invariant.

On rappelle que les éléments de $\mathrm{GL}_3(\mathbb{Z})$ sont les matrices à coefficients entiers de déterminant ± 1. La matrice M est la matrice de r dans la base (u, v, w). Si on appelle A la matrice de r dans la base canonique et P la matrice de passage de la base canonique à la base (u, v, w), on a donc $\mathrm{M} = \mathrm{P}^{-1}\mathrm{AP}$.

La matrice M est semblable à A. On en déduit

$$\mathrm{Tr}(\mathrm{M}) = \mathrm{Tr}(\mathrm{A}) = 1 + 2\cos\theta,$$

où θ est une mesure de l'angle de la rotation r. Comme $\mathrm{M} \in \mathrm{GL}_3(\mathbb{Z})$, sa trace est entière, donc $2\cos\theta \in \mathbb{Z} \cap [\![-2, 2]\!]$, *i.e.* $\cos\theta \in \{\pm 1, \pm\frac{1}{2}, 0\}$. On trouve $\theta \in \left\{0, \pi, \pm\frac{\pi}{2}, \pm\frac{\pi}{3}, \pm\frac{2\pi}{3}\right\}$. On en déduit que $6\theta \equiv 2\pi \pmod{2\pi}$ et donc $\mathrm{A}^6 = \mathrm{I}_3$ et $\mathrm{M}^6 = \mathrm{I}_3$.

On peut démontrer réciproquement que si $\theta \in \left\{0, \pi, \pm\frac{\pi}{2}, \pm\frac{\pi}{3}, \pm\frac{2\pi}{3}\right\}$, il existe un réseau Λ de $\mathbb{R}^3$ et une rotation d'angle θ qui laisse Λ invariant. On suppose que l'espace est orienté.

Une rotation d'angle nul laisse invariant tout réseau.

Considérons le réseau engendré par une base orthonormale directe (u, v, w). La rotation r d'axe $\mathrm{Vect}(u)$ et d'angle $\frac{\pi}{2}$ transforme u en u, v en w et w en $-v$. Elle laisse clairement invariant le réseau Λ. Les rotations r^2 et r^3 ont pour angles π et $-\frac{\pi}{2}$ et laissent invariant Λ.

Considérons un réseau engendré par (u, v, w), où les trois vecteurs sont de norme 1, $u \in \mathrm{Vect}(v,w)^\perp$ et (v,w) ont un angle de $\frac{\pi}{3}$ (le plan $\mathrm{Vect}(v,w)$ étant orienté par u). Soit r la rotation d'axe $\mathrm{Vect}(u)$ d'angle $\frac{\pi}{3}$. On a $r(u) = u$, $r(v) = w$ et $r(w) = w - v$. Avec les notations précédentes, $\mathrm{M} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 1 \end{pmatrix}$. Cette matrice appartient à $\mathcal{M}_3(\mathbb{Z})$ et son déterminant vaut 1 donc M appartient à $\mathrm{GL}_3(\mathbb{Z})$. Ainsi r laisse invariant le réseau. Il en est de même des rotations r^{-1}, r^2, r^{-2} qui ont pour angles $-\frac{\pi}{3}$, $\frac{2\pi}{3}$ et $-\frac{2\pi}{3}$. $\lhd$

4.53. Un problème de Pólya (1918)

1. *Lemme de Blichfeldt.* Soit X une partie de $\mathbb{R}^2$ d'aire strictement supérieure à 1. Montrer que X contient deux points distincts u et v tels que $v - u$ soit dans $\mathbb{Z}^2$.

2. *Théorème de Minkowski.* Soit C une partie convexe de $\mathbb{R}^2$ symétrique par rapport à 0 et d'aire strictement supérieure à 4. Montrer que C contient un point de $\mathbb{Z}^2$ distinct de l'origine.

3. *Le problème de Pólya.* Dans le plan $\mathbb{R}^2$, on considère une forêt obtenue en plantant un arbre circulaire de rayon $r > 0$ en chaque point à coordonnées entières hormis l'origine. Pour $\mathrm{N} \in \mathbb{N}^*$, on note ρ_N la borne supérieure de l'ensemble des rayons $r > 0$ tels que l'on puisse voir au moins un point situé à une distance supérieure ou égale à N. Montrer que $\dfrac{1}{\sqrt{1+\mathrm{N}^2}} \leqslant \rho_\mathrm{N} \leqslant \dfrac{1}{\mathrm{N}}$.

(**École normale supérieure**)

▷ **Solution.**

1. Supposons par l'absurde que pour tout couple (u, v) de points distincts de X, le vecteur $v - u$ n'appartienne pas à $\mathbb{Z}^2$. Dans ce cas, si a et b sont des points distincts de $\mathbb{Z}^2$, les parties $\mathrm{X} - a$ et $\mathrm{X} - b$ sont disjointes. Notons K le carré $[0,1[\times[0,1[$. En notant $\mu(\mathrm{A})$ l'aire de la partie A, on a

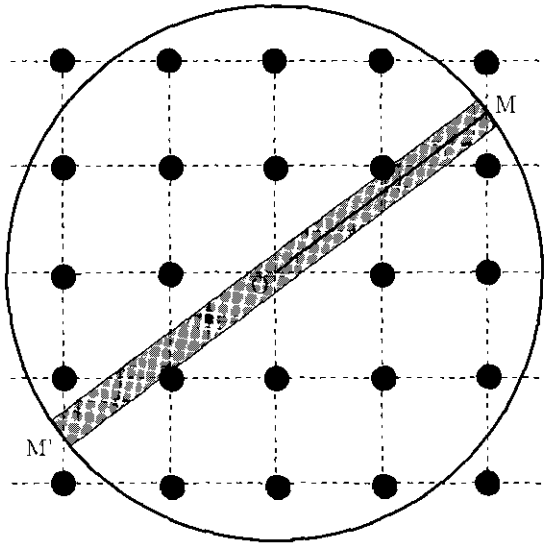
$$\mu(\mathrm{X}) = \sum_{a\in\mathbb{Z}^2} \mu(\mathrm{X} \cap (a + \mathrm{K})) = \sum_{a\in\mathbb{Z}^2} \mu((\mathrm{X} - a) \cap \mathrm{K}) \leqslant \mu(\mathrm{K}) = 1$$

car les translations conservent les aires et, comme on vient de le dire, les ensembles $\mathrm{X} - a$ sont deux à deux disjoints. Cela contredit l'hypothèse.

2. Notons C′ l'image de C par l'homothétie de centre O et de rapport $\frac{1}{2}$. L'aire de C′ est strictement supérieure à 1. Par le lemme de Blichfeldt on peut trouver deux points distincts u, v de C′ tels que $v - u \in \mathbb{Z}^2$.

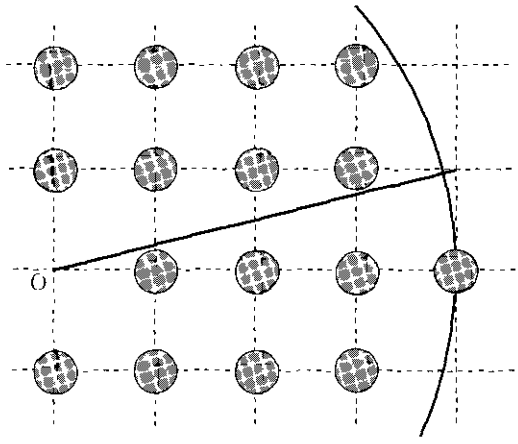
Par définition de C′, les points $2u$ et $2v$ sont dans C. Comme C est symétrique par rapport à l'origine, on a $-2v \in C$. Enfin comme C est convexe, $u - v = \frac{1}{2}(2u - 2v)$ est dans C et dans $\mathbb{Z}^2 \setminus \{(0,0)\}$. C'est le résultat demandé.

3. Soit $r > 0$ tel qu'il soit possible de voir au moins un point situé à une distance supérieure ou égale à N. En particulier, il existe un point M visible depuis l'origine et situé à une distance N. Le symétrique M′ de M par rapport à l'origine est également visible. Épaississons le segment [MM′] en un rectangle de largeur $2r$ et de longueur 2N comme le montre la figure ci-après.



Le fait que M est visible de O signifie que l'intérieur du rectangle ne contient aucun point de $\mathbb{Z}^2$ hormis $(0,0)$. Comme il s'agit d'un convexe symétrique par rapport à l'origine, la question précédente impose que son aire $4r\mathrm{N}$ soit inférieure ou égale à 4. Il en résulte que $r \leqslant \frac{1}{\mathrm{N}}$. En passant à la borne supérieure on obtient donc la majoration $\rho_{\mathrm{N}} \leqslant \frac{1}{\mathrm{N}}$.

Passons à la minoration. On va regarder quels sont les rayons r qui permettent de voir le point de coordonnées $(\mathrm{N}, 1)$ qui est à une distance $\sqrt{1 + \mathrm{N}^2} > \mathrm{N}$.



Par symétrie, il suffit clairement que r soit inférieur à la distance du point $(1,0)$ à la droite d'équation $y = \frac{x}{N}$. Cette distance vaut $\frac{1}{\sqrt{1+N^2}}$, car la distance d'un point (x_0, y_0) à une droite d'équation $ax+by+c=0$ est $\frac{|ax_0+b_0+c|}{\sqrt{a^2+b^2}}$. D'où le résultat. ◁

Le résultat de la question 2 de l'exercice précédent se généralise en dimension quelconque et c'est l'objet de l'exercice suivant. La démonstration est la même.

4.54. Théorème de Minkowski

Une base $(e_1, \ldots, e_p)$ de $\mathbb{R}^p$ définit le réseau

$$\mathcal{R}(e_1, \ldots, e_p) = \sum_{i=1}^{p} \mathbb{Z} e_i = \left\{ \sum_{i=1}^{p} x_i e_i,\ \forall i \in [\![1, p]\!],\ x_i \in \mathbb{Z} \right\}.$$

On appelle maille du réseau l'ensemble $\mathcal{M}(e_1, \ldots, e_p)$ des vecteurs $x = \sum_{i=1}^{p} x_i e_i$ tels que $x_i \in [0, 1[$, pour tout $i \in [\![1, p]\!]$.

1. Montrer que, pour tout réel $\rho > 0$, $\tau_\rho = \mathrm{B}(0, \rho) \cap \mathcal{R}(e_1 \ldots, e_p)$ est fini, où $\mathrm{B}(0, \rho)$ désigne la boule euclidienne de centre l'origine et de rayon ρ.

2. Montrer qu'il existe V_0 tel que tout convexe symétrique K de volume strictement supérieur à V_0 contienne un point du réseau autre que 0. On exprimera V_0 en fonction du volume de la maille.

(École polytechnique)

▷ **Solution.**

1. Le résultat est géométriquement évident : dans toute partie bornée de $\mathbb{R}^p$, il n'y a qu'un nombre fini de points du réseau. Mais justifions-le précisément. L'application $x = \sum_{i=1}^{p} x_i e_i \longmapsto \max_{1 \leqslant i \leqslant p} |x_i|$ est une norme sur $\mathbb{R}^p$. Toutes les normes sur $\mathbb{R}^p$ sont équivalentes donc il existe $\lambda > 0$ tel que, pour tout $x \in \mathbb{R}^p$, $\max_{1 \leqslant i \leqslant p} |x_i| \leqslant \lambda \|x\|$. Soit $x = \sum_{i=1}^{p} x_i e_i \in \tau_\rho$. On a alors, pour tout $i \in [\![1, p]\!]$,

$$|x_i| \leqslant \lambda \|x\| \leqslant \lambda \rho$$

et donc $x_i \in [\![-\mathrm{E}(\lambda\rho), \mathrm{E}(\lambda\rho)]\!]$, où E est la partie entière. Il y donc au plus $2\mathrm{E}(\lambda\rho) + 1$ valeurs de x_i et $(2\mathrm{E}(\lambda\rho) + 1)^p$ éléments dans τ_ρ.

2. On note μ le volume. Montrons que $\mathrm{V}_0 = 2^p \mu(\mathcal{M})$ convient, où $\mathcal{M}$ désigne la maille du réseau.

Soit K un convexe compact, symétrique de volume strictement supérieur à $2^p \mu(\mathcal{M})$. Notons K′ l'image de K par l'homothétie de centre 0 et de rapport $\frac{1}{2}$. Le volume de K′ est strictement supérieur à $\mu(\mathcal{M})$. Montrons que K′ contient deux points distincts u et v, tels que $u - v \in \mathcal{R}$.

Les ensembles $\mathcal{M} + a$, où a décrit $\mathcal{R}$ forment une partition de $\mathbb{R}^p$. En effet, pour tout $(x_1, \ldots, x_p) \in \mathbb{R}^p$, on a

$$x = \sum_{i=1}^{p} x_i e_i = \sum_{i=1}^{p} (x_i - \mathrm{E}(x_i)) e_i + \sum_{i=1}^{p} \mathrm{E}(x_i) e_i \in \mathcal{M} + a,$$

avec $a = \sum_{i=1}^{p} \mathrm{E}(x_i) e_i \in \mathcal{R}$. Réciproquement, si $x \in \mathcal{M} + a$, où $a = \sum_{i=1}^{p} a_i e_i$ est élément de $\mathcal{R}$, on a, pour tout $i \in [\![1, p]\!]$, $a_i = \mathrm{E}(x_i)$, ce qui montre l'unicité de a. On en déduit que

$$\mu(\mathrm{K}') = \sum_{a \in \mathcal{R}} \mu(\mathrm{K}' \cap (\mathcal{M} + a)) = \sum_{a \in \mathcal{R}} \mu((\mathrm{K}' - a) \cap \mathcal{M}),$$

car le volume est invariant par translation.

Supposons par l'absurde que pour tout couple (u, v) de points distincts de K′, le vecteur $u - v$ n'appartienne pas à $\mathcal{R}$. Dans ce cas, si a et b sont dans $\mathcal{R}$ et distincts, les parties $\mathrm{K}' - a$ et $\mathrm{K}' - b$ sont disjointes, puisque $\mathcal{R}$ est un groupe additif. On a alors

$$\mu(\mathrm{K}') = \mu\left(\left(\bigcup_{a \in \mathcal{R}} (\mathrm{K}' - a)\right) \cap \mathcal{M}\right) \leqslant \mu(\mathcal{M}),$$

ce qui contredit l'hypothèse.

Si u et v sont deux éléments distincts de K′ tels que $u - v \in \mathcal{R}$, on écrit $u - v = \frac{1}{2}(2u - 2v)$. Alors $2u$ et $2v$ appartiennent à K par définition de K′, donc $-2v$ appartient à K, car K est symétrique et donc $u - v$ est dans K car K est convexe. Ainsi $u - v$ est un élément de $\mathcal{R}$ non nul et qui appartient à K ◁

4.55. Formule d'Euler

Soit P un polyèdre convexe de $\mathbb{R}^3$. On note s le nombre de sommets, a le nombres d'arêtes et f le nombres de faces de P.

1. Montrer que l'on peut associer à P une figure plane comme suit : on choisit une face F de P et on représente à l'intérieur de F les arêtes et les sommets de P de façon à obtenir un pavage de F par des polygones convexes qui représentent les autres faces de P. Dans cette figure plane, deux arêtes ne se coupent pas et chaque sommet est commun à au moins trois arêtes. *Indication : utiliser une projection sur un plan à partir d'un point de l'espace bien choisi.*

2. Calculer de deux façons différentes la somme des angles de tous les polygones représentant les faces de P autres que F et en déduire la *formule d'Euler*

$$s - a + f = 2.$$

3. On suppose le polyèdre régulier ; chaque face possède α arêtes et chaque sommet est entouré de β faces. Montrer que

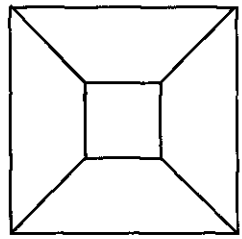
$$\frac{1}{\alpha} + \frac{1}{\beta} = \frac{1}{a} + \frac{1}{2}.$$

(**École normale supérieure**)

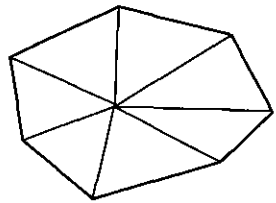
▷ **Solution.**

1. Le polyèdre P est l'intersection d'un nombre fini de demi-espaces $H_1, H_2, \ldots, H_f$ limités par les plans des faces, H_f étant limité par le plan Q contenant la face F. Considérons $P' = H_1 \cap \ldots \cap H_{f-1}$ et un point O appartenant à l'intérieur de $P' \setminus P$. Pour tout M de P, M et O sont de part et d'autre de Q. La droite (OM) coupe donc le plan Q en un unique point M′, qu'on note $p(M)$. Comme P′ est convexe, M′ appartient à P′. Puisque, de plus, M′ appartient à Q, M′ appartient à $P \cap Q = F$. L'application p réalise une bijection de $P \setminus \overset{\circ}{F}$ sur F. Les images des sommets de P sont des points. Les images des arêtes sont des segments. En effet l'image de l'arête [AB] est l'intersection du plan Q et de l'intérieur du triangle OAB. Chaque face est transformée en un

polygone convexe. On obtient ainsi un pavage de F par des polygones convexes. Voici par exemple ce qu'on obtient dans le cas du cube :



2. La somme des angles d'un polygone convexe à $n \geqslant 3$ côtés est égale à $(n-2)\pi$, comme on le voit par exemple en triangulant le polygone en n triangles ayant tous un sommet commun.



La somme S de tous les angles des polygones pavant F est donc égale à $(a_1 - 2f_1)\pi$, où a_1 est le nombre total d'arêtes des polygones et f_1 le nombre de polygones. On a $f_1 = f-1$, car chaque polygone correspond à une face de P distincte de F et $a_1 = 2a - a_2$, où a_2 est le nombre d'arêtes de F, car toute arête qui n'appartient pas à F appartient exactement à 2 polygones. On obtient $S = (2a - a_2 - 2f + 2)\pi$.

Pour calculer S, on peut aussi faire la somme des angles correspondant aux différents sommets des polygones. En appelant s_2 le nombre de sommets de F, on obtient $S = (s - s_2)2\pi + S_1$, où S_1 est la somme des angles de F et donc $S = (s - s_2)2\pi + (a_2 - 2)\pi = (2s - a_2 - 2)\pi$, puisque $s_2 = a_2$.

On en déduit $(2a - a_2 - 2f + 2)\pi = (2s - a_2 - 2)\pi$ et donc $s - a + f = 2$.

3. Chaque face possédant α arêtes, on obtient, en comptant le nombre total d'arêtes, $f\alpha = 2a$ car chaque arête appartient à deux faces. De même, chaque sommet est entouré de β faces. On en déduit que β arêtes aboutissent à chaque sommet, ce qui montre que $s\beta = 2a$, car chaque arête aboutit à deux sommets. En remplaçant dans la relation d'Euler, on obtient $\dfrac{2a}{\beta} - a + \dfrac{2a}{\alpha} = 2$, soit en divisant par $2a$, $\dfrac{1}{\alpha} + \dfrac{1}{\beta} = \dfrac{1}{a} + \dfrac{1}{2}$. ◁

Une face de P *a au moins trois côtés donc* $\alpha \geqslant 3$. *De même en un sommet, on assemble au moins trois faces donc* $\beta \geqslant 3$. *On en déduit*

$$\frac{1}{\alpha} > \frac{1}{2} - \frac{1}{\beta} \geqslant \frac{1}{2} - \frac{1}{3} \geqslant \frac{1}{6}$$

et donc $\alpha \leqslant 5$. On obtient ainsi $3 \leqslant \alpha \leqslant 5$ et de même $3 \leqslant \beta \leqslant 5$. L'un des deux entiers α, β doit être égal à 3, car sinon $\frac{1}{\alpha} + \frac{1}{\beta} \leqslant \frac{1}{2}$, ce qui est impossible car $a > 0$.

Chaque valeur du couple (α, β) vérifiant ces conditions correspond effectivement à un polyèdre régulier, qui est unique à une similitude près : le tétraèdre correspond à $(\alpha, \beta) = (3, 3)$, le cube à $(\alpha, \beta) = (4, 3)$, l'octaèdre à $(\alpha, \beta) = (3, 4)$, le dodécaèdre à $(\alpha, \beta) = (5, 3)$, l'icosaèdre à $(\alpha, \beta) = (3, 5)$.

Terminons par deux problèmes de pavage de rectangles. Le premier a été posé par De Bruijn en 1959 dans une revue hongroise.

4.56. Rectangles semi-entiers

On dit qu'un rectangle est semi-entier si l'un au moins de ses côtés a pour longueur un nombre entier. Montrer que tout rectangle que l'on peut paver par des rectangles semi-entiers est semi-entier.
(**École normale supérieure**)

▷ **Solution.**

Nous allons donner deux preuves, toutes les deux très jolies de ce résultat[2].

• Le plan est rapporté à un repère dont les axes sont parallèles aux côtés du grand rectangle. Il est sous-entendu que tous les rectangles du pavage ont leurs côtés parallèles aux axes. Étant donnés quatre réels a, b, c, d, considérons le rectangle R limité par les droites $x = a$, $x = b$, $y = c$ et $y = d$. La solution tient en quelques lignes après la considération d'une intégrale double judicieuse. On a en effet, par la formule de Fubini,

$$\begin{aligned}\iint_{\mathrm{R}} e^{2\pi i(x+y)}\mathrm{d}x\mathrm{d}y &= \int_a^b e^{2\pi ix}\mathrm{d}x \int_c^d e^{2\pi iy}\mathrm{d}y \\ &= -\frac{1}{4\pi^2}\left(e^{2\pi ib} - e^{2\pi ia}\right)\left(e^{2\pi id} - e^{2\pi ic}\right)\end{aligned}$$

et cette intégrale est nulle si, et seulement si, $b - a \in \mathbb{Z}$ ou $d - c \in \mathbb{Z}$, c'est-à-dire si, et seulement si, le rectangle R est semi-entier.

Par hypothèse, l'intégrale double de $e^{2\pi i(x+y)}$ sur chacun des rectangles du pavage est nulle. On en déduit que la somme de ces intégrales, qui est l'intégrale de $e^{2\pi i(x+y)}$ sur le grand rectangle, est nulle, donc que ce rectangle est semi-entier !

2. Il existe un grand nombre de preuves différentes mais il est toujours possible d'en trouver de nouvelles : voir par exemple celle publiée très récemment par François Xavier Vialard dans la RMS 116-3, p. 6-10.

• Voici une autre solution, complètement différente, et qui provient plutôt de la théorie des graphes même si on va éviter d'en parler. Elle illustre aussi la puissance des doubles sommations, mais dans une version discrète. On supposera sans perte de généralité que les sommets du rectangle sont $(0,0)$, $(a,0)$, $(0,b)$ et (a,b). Le but est de prouver que a ou b est entier. Notons $\mathcal{R} = \{\mathrm{R}_1, \ldots, \mathrm{R}_p\}$ l'ensemble des rectangles du pavage et $\mathcal{S}$ l'ensemble des points de $\mathbb{Z}^2$ qui sont dans le rectangle $[0,a] \times [0,b]$. Pour $\mathrm{R}_i \in \mathcal{R}$ et $p = (x,y) \in \mathcal{S}$ on pose $\varepsilon(\mathrm{R}_i, p) = 1$ si p est un sommet de R_i et $\varepsilon(\mathrm{R}_i, p) = 0$ sinon. On s'intéresse à la somme

$$\alpha = \sum_{i=1}^{p} \sum_{p \in \mathcal{S}} \varepsilon(\mathrm{R}_i, p) = \sum_{p \in \mathcal{S}} \sum_{i=1}^{p} \varepsilon(\mathrm{R}_i, p).$$

Comme les rectangles R_i sont semi-entiers, la somme $\sum_{p \in \mathcal{S}} \varepsilon(\mathrm{R}_i, p)$ prend les valeurs 0, 2 ou 4, car elle compte le nombre de sommets à coordonnées entières de R_i. On en déduit que α est un entier pair. Regardons maintenant la seconde expression de α. Pour p fixé, $\sum_{i=1}^{p} \varepsilon(\mathrm{R}_i, p)$ représente le nombre de rectangles du pavage dont p est un sommet. Ce nombre ne peut valoir que 0, 2 ou 4 sauf si p est l'un des 4 sommets du grand rectangle, auquel cas il vaut 1. Comme α est pair, il y a un nombre pair de ces sommets qui sont dans $\mathbb{Z}^2$. Étant donné qu'il y a au moins le sommet $(0,0)$, on a un autre coin dans $\mathbb{Z}^2$ et cela prouve le résultat. ◁

Le résultat suivant a été prouvé pour la première fois par Dehn en 1903.

4.57. Pavage d'un rectangle par des carrés

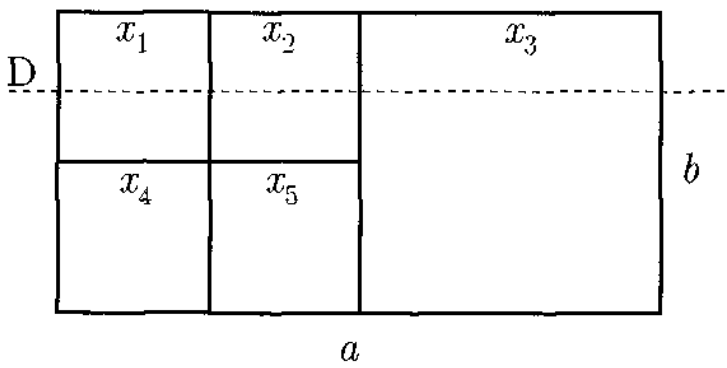
Soit R un rectangle de côtés a et b. On suppose qu'on peut faire une partition de R en n carrés de côtés $x_1, \ldots, x_n$. Montrer que les nombres $\dfrac{a}{b}$ et $\dfrac{x_i}{b}$ pour tout $i \in [\![1, n]\!]$ sont rationnels.

(École polytechnique)

▷ **Solution.**

Il est évidemment sous-entendu que les carrés du pavage ont leur côtés parallèles à ceux du rectangle R. On va traduire les hypothèses par un système linéaire à coefficients entiers vérifié par les x_i. Soit D une droite horizontale qui coupe R et telle que chaque carré du pavage traversé par D le soit dans son intérieur. Si $\mathrm{I_D}$ est l'ensemble des indices des carrés qui sont traversés par D on a l'équation $\sum_{i \in \mathrm{I_D}} x_i = a$. On considère le

système linéaire (S_1) obtenu en prenant toutes les équations différentes ainsi obtenues lorsque la droite D varie. On procède de même avec des droites Δ verticales ce qui conduit à un second système d'équations de la forme $\sum_{i\in J_\Delta} x_i = b$ où J_Δ est une partie de $[\![1,n]\!]$. Notons (S) le système linéaire obtenu en juxtaposant (S_1) et (S_2) et en regardant $a = x_0$ et $b = x_{n+1}$ aussi comme des inconnues. Si X est le vecteur colonne de taille $n+2$ contenant $x_0, x_1, \ldots, x_n, x_{n+1}$ on a donc une matrice $A \in \mathcal{M}_{m,n+2}(\mathbb{R})$ à coefficients dans $\mathbb{Z}$ (et même dans $\{-1,0,1\}$) telle que $AX = 0$ (m est le nombre total d'équations obtenues). Voici un exemple très simple de pavage pour illustrer ce qui précède :



On obtient dans ce cas les équations suivantes :

$$(S) = \begin{cases} x_1 + x_2 + x_3 = a \\ x_4 + x_5 + x_3 = a \\ x_1 + x_4 = b \\ x_2 + x_5 = b \\ x_3 = b \end{cases}$$

et la matrice A est $\begin{pmatrix} -1 & 1 & 1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix}$. L'idéal serait d'obtenir une matrice de rang $n+1$. En effet, dans ce cas $\operatorname{Ker} A$ est une droite vectorielle et, comme A est à coefficients rationnels, on peut en trouver un vecteur directeur dans $\mathbb{Q}^{n+2}$. Comme X est colinéaire à ce vecteur on a le résultat : tous les quotients $\frac{x_i}{x_j}$ sont rationnels. Autrement dit, R est l'image par une homothétie d'un rectangle à côtés entiers pavé par des carrés à côtés entiers.

Cependant, notre petit exemple montre que ce n'est pas forcément vrai : notre matrice de taille $(5,7)$ ne peut pas être de rang 6. En fait, on voit géométriquement qu'on a raté quelques équations importantes. On

a ainsi forcément $x_1 = x_2$ et $x_4 = x_5$ en appliquant la même démarche qu'avant mais dans les sous-rectangles formés des carrés 1 et 2 d'une part et des carrés 4 et 5 d'autre part. On peut vérifier qu'en ajoutant ces équations (une seule suffirait) on a bien un système de rang 6.

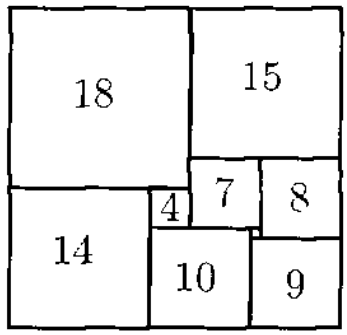
Dans le cas général, nous noterons désormais (S), non pas le système précédent, mais le système obtenu en écrivant toutes les égalités possibles entre deux sommes de x_i traduisant une égalité de distance entre deux segments verticaux ou horizontaux apparaissant dans le pavage. La solution particulière X vérifie la propriété suivante : $\sum_{i=1}^{n} x_i^2 = ab = x_0 x_{n+1}$. Celle-ci traduit le fait que l'aire du rectangle est la somme des aires des carrés. Nous allons commencer par prouver qu'il en est de même pour toute autre solution $Y = (y_0, \ldots, y_n)$ du système (S). Soit donc Y une telle solution. Pour tout réel t, le vecteur $X' = X + tY$ est encore solution du système. Nous choisissons t assez petit pour que les réels $x'_i = x_i + ty_i$ soient tous strictement positifs. Il est alors assez facile de se convaincre qu'il est possible de paver un rectangle de côté x'_0 et x'_{n+1} à l'aide des n carrés de côtés $x'_1, \ldots, x'_n$ en disposant ceux-ci de la même manière que dans le pavage d'origine de R. Cela provient de ce que les x'_i vérifient toutes les relations du système (S). On a donc, pour tout réel t assez petit,

$$\sum_{i=1}^{n} (x_i + ty_i)^2 = (x_0 + ty_0)(x_{n+1} + ty_{n+1}).$$

Cette égalité polynomiale valable sur un voisinage de 0 impose l'égalité des coefficients et en particulier le fait que $\sum_{i=1}^{n} y_i^2 = y_0 y_{n+1}$. C'est ce que nous voulions prouver.

On en déduit que le système (S') obtenu en ajoutant au système (S) : $AY = 0$ l'équation $y_{n+1} = -y_0$ est de rang $n+2$ puisque l'égalité qu'on vient de démontrer prouve qu'une solution de (S') est obligatoirement nulle. Il en résulte que (S) est au moins de rang $n+1$ et cela termine la preuve. $\lhd$

La réciproque est évidente : tout rectangle avec $\frac{a}{b} \in \mathbb{Q}$ *peut être pavé par des carrés. En fait, comme dit plus haut dans la solution, on se ramène par homothétie à un rectangle à côtés entiers pour lequel c'est trivial. Il a alors été naturel de se demander si on pouvait trouver des pavages à l'aide de carrés de côtés deux à deux distincts (appelés pavages parfaits). Voici le premier exemple qui a été trouvé, par Moroń, en 1925. Il s'agit d'un rectangle* 33×32 *pavé par* 9 *carrés :*



C'est le pavage parfait d'ordre minimal (l'ordre est le nombre de carrés utilisés, ici 9). Depuis, de nombreux résultats ont été obtenus sur ce sujet. On sait ainsi que l'ordre minimal d'un carré parfait est 21 et qu'il est unique à symétrie près. Il a été trouvé avec l'aide d'un ordinateur par Duijvestijn en 1978. En fait, Sprague a montré en 1940 que tout rectangle de côtés a, b avec $\frac{a}{b} \in \mathbb{Q}$ admet un pavage parfait.

Signalons pour conclure une autre approche du théorème de Dehn. Il est possible d'associer à un pavage un graphe planaire et le résultat découle alors assez rapidement du théorème de Kirchhoff sur les réseaux électriques[3].

3. Le lecteur intéressé pourra consulter l'article original dans lequel cette idée est apparue : R.L. BROOKS, C.A.B. SMITH, A.H. STONE, W.T. TUTTE, The dissection of rectangles into squares. *Duke. Math. J.*7 (1940).

Index

A

Abel (transformation d'), 149, 153

B

Bergström (inégalité de), 209
billard et lois de Descartes, 293
Blichfeldt (lemme de), 334

C

Carathéodory (théorème de), 316
Cauchy (suite de), 260
Cauchy (théorème d'entrelacement de), 144
Cauchy-Schwarz (inégalité de), 14, 32, 41, 84, 88, 103, 128, 137, 140, 154, 183, 185, 207–210, 220, 238
Choleski (décomposition de), 131, 133
convexité, 316–325

D

Darboux-Christoffel (formule de), 46
Davenport-Cassels (lemme de), 241
De Bruijn (problème de), 340
décomposition
  d'Iwasawa, 132
  de Choleski, 131, 133
  en valeurs singulières, 123
  polaire, 126–131, 177
  QR, 40, 131, 134
Dehn (problème de), 341

E

Euler (formule d'), 338

F

Farkas (lemme de), 80
Fermat (problème de), 295
Fischer (inégalité de), 134
Fischer-Cochran (théorème de), 200
frame, 99
Fréchet-Von Neumann-Jordan (théorème de), 12

G

Gauss (méthode de), 50
Gergonne (point de), 254
Gram (matrice de), 38, 53, 55–57, 122
Gram-Schmidt (orthonormalisation de), 36–48, 121, 131, 134, 182

H

Hadamard (inégalité d'), 40, 132, 133, 171
Hermite (polynômes d'), 48
Héron (formule de), 253

I

Iwasawa (décomposition d'), 132

J

John-Loewner (ellipsoïde de), 229
Jung (théorème de), 317

K

Kantorovich (inégalité de), 139
Kelly, 252
Ky-Fan (inégalités de), 150

L

Laguerre (polynômes de), 45
Leibniz (fonction scalaire de), 260
Liapounov (théorème de), 162

M

Maschke (théorème de), 62
Ménélaüs (théorème de), 248
minimax (théorème du), 142–152
Minkowski
  inégalité de, 223
  théorème de, 334, 336
Mirman (théorème de), 182

O

orthique (triangle), 293
orthoptique
  cercle, 276
  sphère, 278

P

Pappus (problème de), 274
Pappus (théorème de), 246
Perron-Frobenius (théorème de), 117, 118
Pfister (théorème de), 195
Pick (théorème de), 326
Pólya (problème de), 334
Ptolémée (théorème de), 257, 258
Pythagore (théorème de), 14, 19, 22, 28, 30, 48, 90

Q

quaternions, 74

R

rectangles semi-entiers, 340

S

Schur
  produit de, 175
  théorème de majoration de, 148, 152
  théorème de trigonalisation de, 182
simplexe régulier, 8, 37, 56, 313
Suleimanova (théorème de), 119
Sylvester
  caractérisation de, 114, 206
  loi d'inertie de, 213
Sylvester-Gallai (théorème de), 251

T

Torricelli (point de), 257, 296

W

Weyl (théorème de perturbation de), 147

IMPRIMÉ EN GRANDE-BRETAGNE
PAR CAMBRIDGE UNIVERSITY PRESS
DÉPÔT LÉGAL MARS 2008

Ce livre est le cinquième volume paru du recueil d'exercices résolus des oraux des Écoles normales supérieures et de l'École polytechnique, et le troisième et dernier tome d'algèbre.

Il est consacré à la théorie des espaces euclidiens et hermitiens, au groupe orthogonal, à la réduction des endomorphismes auto-adjoints, aux formes quadratiques, à la géométrie affine et à la géométrie euclidienne.

Il comporte 266 exercices, dont la solution est rédigée avec le soin et le souci d'exposer les idées et les démarches de raisonnement qui sont maintenant bien connus des lecteurs de la série. Comme dans les volumes précédents, la grande culture mathématique des auteurs leur permet de situer dans le contexte qui leur a donné naissance la plupart des énoncés, qui, à ce niveau sont presque toujours des résultats scientifiques dignes d'intérêt, et dont quelques-uns sont issus de recherches récentes.

Collection enseignement des mathématiques

ISBN 978-2-84225-092-8 Graphisme : Massin

exercices de mathématiques

oraux x-ens

analyse 1

Serge Francinou
Hervé Gianella
Serge Nicolas

CASSINI

EXERCICES DE MATHÉMATIQUES
ORAUX X-ENS

Enseignement des mathématiques

1. J.-Y. Ouvrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouvrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
9. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
14. H. Krivine, *Exercices de mathématiques pour physiciens*
15. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
16. M. Willem, *Analyse fonctionnelle élémentaire*
17. É. Amar, É. Matheron, *Analyse complexe*
18. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
19. D. Perrin, *Mathématiques d'école*
20. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
21. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
22. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*

SERGE FRANCINOU
HERVÉ GIANELLA
SERGE NICOLAS

Exercices de mathématiques des oraux de l'École polytechnique et des Écoles normales supérieures

Analyse. Tome I

Deuxième édition, revue et augmentée

CASSINI

SERGE FRANCINOU, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Marcelin Berthelot.

HERVÉ GIANELLA, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Saint-Louis.

SERGE NICOLAS, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Henri IV.

ISBN 978-2-84225-135-2
(1$^{\text{ère}}$ édition, 2003, ISBN 2-84225-032-X)

© Cassini, Paris, 2007.

Table des matières

Introduction 1

Chapitre 1. Nombres réels et complexes. Topologie de $\mathbb{R}$ et $\mathbb{C}$ 5

1.1. Développement décimal propre d'un réel 7
1.2. Caractérisation des rationnels 10
1.3. Non-dénombrabilité de $\mathbb{R}$ 12
1.4. Injection de $\mathbb{R}$ dans $\mathcal{P}(\mathrm{N})$ 13
1.5. Calcul d'une somme 14
1.6. Développement en série de Engel 16
1.7. Représentation des réels > 0 comme sommes de séries 18
1.8. Sur le nombre e 21
1.9. Irrationalité de π 22
1.10. Irrationalité de $(\arccos 1/3)/\pi$ 24
1.11. Nombres de Liouville 25
1.12. Nombres de Pisot 27
1.13. Sous-groupes additifs de $\mathbb{R}$ 29
1.14. Une inégalité 31
1.15. Inégalité du réordonnement 31
1.16. Inégalité de Tchebychev 33
1.17. Inégalité dans $\mathbb{C}$ (1) 34
1.18. Inégalité dans $\mathbb{C}$ (2) 36
1.19. Un lemme de confinement 39
1.20. Équation fonctionnelle de Shapiro 42
1.21. Homographies laissant invariant le disque unité 43
1.22. Suites de polygones 45
1.23. Mesure extérieure de Lebesgue 47
1.24. Recouvrements 49
1.25. Irrationnels denses dans une partie 51
1.26. Application ouverte 52
1.27. Partition dénombrable de $[0, 1]$ 53

Chapitre 2. Suites réelles et complexes 57

2.1. Une étude de monotonie 59
2.2. Permutation des termes d'une suite 60
2.3. Suites sous-additives 62
2.4. Suites monotones 63
2.5. Suites convexes bornées 64
2.6. Caractérisation de la convergence au sens de Cesàro 65

2.7. Suites convergentes au sens de Cesàro telles que $u_n^6 \to 1$ . . . 66
2.8. Moyenne de Cesàro d'une suite récurrente . . . 69
2.9. Théorème taubérien de Hardy . . . 70
2.10. Limites de tranches de Cauchy . . . 72
2.11. Valeurs d'adhérence d'une suite complexe . . . 75
2.12. Question de convergence . . . 76
2.13. Étude de convergence . . . 77
2.14. Minoration de la limite supérieure . . . 78
2.15. Suites et approximation diophantienne . . . 79
2.16. Équivalent d'une suite d'entiers . . . 81
2.17. Somme des puissances n-ièmes des n premiers entiers . . . 82
2.18. Racines itérées . . . 83
2.19. Un critère de convergence pour les systèmes dynamiques discrets . . . 86
2.20. Étude d'un système dynamique (1) . . . 88
2.21. Étude d'un système dynamique (2) . . . 90
2.22. Théorème de Sarkowski (1964) . . . 92
2.23. La fonction «tente» . . . 94
2.24. Un système dynamique discret et son analogue continu . . . 96
2.25. Équivalent d'une suite définie par récurrence . . . 99
2.26. Développement asymptotique d'une suite récurrente (1) . . . 100
2.27. Développement asymptotique d'une suite récurrente (2) . . . 102
2.28. Détermination d'une suite récurrente . . . 103
2.29. Étude de $x_{n+1} = 1 + \dfrac{n}{x_n}$ . . . 104
2.30. Un exercice d'Olympiades . . . 106
2.31. Équivalent d'une suite récurrente . . . 107
2.32. Suite arithmético-géométrique perturbée . . . 109
2.33. Système dynamique perturbé . . . 110
2.34. Suite récurrente linéaire d'ordre 2 . . . 111
2.35. Une équation fonctionnelle . . . 113
2.36. Suites vérifiant $u_{n+2} = (n+1)u_{n+1} + u_n$ . . . 113
2.37. Suites à récurrence linéaire (1) . . . 115
2.38. Suites à récurrence linéaire (2) . . . 118
2.39. Étude d'une suite récurrente . . . 119
2.40. Suite vérifiant $u_{n+1} = \ln(1+u_n) + \ln(1+u_{n-1})$ . . . 120
2.41. Suite vérifiant $u_{n+1} = \dfrac{f(u_1) + \cdots + f(u_n)}{n}$ . . . 121
2.42. Suites de Schwob . . . 122
2.43. Moyenne arithmético-harmonique . . . 123
2.44. Récurrences simultanées . . . 125
2.45. Suite définie implicitement (1) . . . 127
2.46. Suite définie implicitement (2) . . . 127
2.47. Suite définie implicitement (3) . . . 128
2.48. Suite définie implicitement (4) . . . 129

2.49. Zéros des polynômes de Taylor d'ordre impair de l'exponentielle 130

Chapitre 3. Séries numériques **133**

3.1. Étude de convergence 134
3.2. Une transformation d'Abel 135
3.3. Nature d'une série (1) 137
3.4. Nature d'une série (2) 138
3.5. Majoration à l'aide d'une intégrale 139
3.6. Sur la négligeabilité 141
3.7. Série et suite 142
3.8. Un théorème de comparaison 143
3.9. Étude de convergence 145
3.10. Condition nécessaire de convergence 145
3.11. Entiers qui s'écrivent sans le chiffre 9 147
3.12. Critère de condensation de Cauchy 147
3.13. Séries définies à l'aide d'une permutation de $\mathbb{N}^*$ 149
3.14. Recherche d'un équivalent (1) 151
3.15. Recherche d'un équivalent (2) 152
3.16. Sommation d'équivalents 153
3.17. Estimation d'une somme partielle 155
3.18. Développement asymptotique de la série harmonique . . . 156
3.19. Calcul de la somme d'une série (1) 159
3.20. Calcul de la somme d'une série (2) 160
3.21. Recherche d'un équivalent 163
3.22. Série des inverses des entiers premiers 166
3.23. Série définie à partir de ppcm 167
3.24. Nature d'une série 167
3.25. Inversion de Möbius 169
3.26. Nombre moyen de diviseurs des entiers inférieurs à x 172
3.27. Nature de $\sum \sin\left(\pi(2+\sqrt{3})^n\right)$ 175
3.28. Suite (a_n) telle $\sum a_n^k = 0$ pour tout entier k non nul 175
3.29. Construction d'une série complexe 177
3.30. Développement asymptotique du terme général 178
3.31. Autour des séries semi-convergentes 180
3.32. Calcul de la somme d'une série alternée 181
3.33. Transformation d'Euler 182
3.34. Sommation par paquets 185
3.35. Exemple de transformation d'Abel 188
3.36. Étude asymptotique à l'aide de transformations d'Abel . . 189
3.37. Séries de Hardy 192
3.38. Recherche d'un équivalent (1) 194
3.39. Recherche d'un équivalent (2) 196

3.40. Convergence en moyenne d'un produit de Cauchy 199
3.41. Application du théorème d'associativité 201
3.42. Étude de sommabilité 203
3.43. Convergence et somme d'une série double 204
3.44. Étude de séries doubles 205
3.45. Inégalité de Carleman (1923) 208
3.46. Inégalité de Hardy 210
3.47. Convergence commutative 213
3.48. Théorème de Riemann 217

Chapitre 4. Fonctions d'une variable réelle **219**

4.1. Applications propres 220
4.2. Zéros d'une combinaison linéaire d'exponentielles 220
4.3. Un calcul de limite 221
4.4. Fonctions sur-additives 222
4.5. Points de discontinuité d'une fonction réglée 224
4.6. Discontinuités d'une fonction monotone 226
4.7. Continuité au sens de Cesàro 227
4.8. Existence d'un point fixe 228
4.9. Un théorème de Paul Lévy (1934) 229
4.10. Une réciproque au théorème des valeurs intermédiaires . . . 232
4.11. Sous-groupes finis du groupe des homéomorphismes 233
4.12. Conjugaison de deux homéomorphismes 234
4.13. Nombre de rotation de Poincaré (1885) 235
4.14. Morphismes du groupe $(\mathbb{R}, +)$ respectant l'inverse 237
4.15. Une équation fonctionnelle (1) 238
4.16. Une équation fonctionnelle (2) 239
4.17. Une équation fonctionnelle (3) 240
4.18. Une équation fonctionnelle (4) 241
4.19. Une équation fonctionnelle (5) 241
4.20. Une équation fonctionnelle (6) 243
4.21. Valeurs propres d'un opérateur 244
4.22. Domination d'une fonction uniformément continue 246
4.23. Lemme de Croft 247
4.24. Étude d'un taux d'accroissement 248
4.25. Pseudo-dérivée 249
4.26. Limite d'une somme 251
4.27. La fonction sinus n'est pas rationnelle 252
4.28. Un cas simple du lemme de Sard 253
4.29. Théorème de Darboux 254
4.30. Sur la formule des accroissements finis 256
4.31. Maximum local 257
4.32. Dénombrabilité des maxima locaux 259

4.33. Majoration de $f^2 + f'^2$ 260
4.34. Interpolation par splines cubiques 260
4.35. Interpolation d'Hermite 264
4.36. Dérivées dominées par un polynôme 266
4.37. Une généralisation du théorème de Rolle 267
4.38. Minoration de la dérivée seconde 268
4.39. Zéros des dérivées successives 269
4.40. Théorème de Glaeser pour une variable (1963) 270
4.41. Étude du maximum de $f \longmapsto f(0) - 2f(\frac{1}{2}) + f(1)$ 272
4.42. Puissances entières 273
4.43. Inégalités de Kolmogorov 274
4.44. Théorème de division 277
4.45. Le théorème de réalisation de Borel 279
4.46. Théorème de Whitney 282
4.47. Convexité 284
4.48. Inégalité de convexité (1) 285
4.49. Inégalité de convexité (2) 285
4.50. Condition suffisante de convexité 286
4.51. Minimum d'une fonction convexe 287
4.52. Intervalle de monotonie d'une fonction strictement convexe 288
4.53. Étude asymptotique d'une fonction convexe 289
4.54. Combinaison convexe positive 290
4.55. Fonctions à variation bornée 292
4.56. Un développement asymptotique 294
4.57. Estimation du maximum d'une fonction polynôme 296
4.58. Dérivation d'équivalents 299
4.59. Intégration d'équivalents 301

Table des matières **303**

Index **309**

Introduction

Cet ouvrage est le premier tome d'analyse d'un recueil d'exercices de mathématiques destiné à la préparation des oraux des concours d'entrée aux Écoles normales supérieures et à l'École polytechnique. Il comportera six tomes, trois d'algèbre et trois d'analyse.

La vocation première des Écoles normales est de former des chercheurs ou des enseignants-chercheurs. Le concours d'entrée vise donc à détecter les qualités scientifiques du candidat, son aptitude à la recherche. À l'oral, on jugera avant tout la capacité de prendre des initiatives, d'utiliser une indication, de mener à bien une démarche. On ne sera pas surpris que les exercices posés aient un contenu mathématique riche, qu'ils soient très éloignés du simple exercice technique, d'application du cours, qu'ils soient souvent difficiles. Ils visent la plupart du temps à la démonstration d'un résultat mathématique significatif. Ils pourraient apparaître excessivement difficiles, si on perdait de vue le déroulement concret de l'épreuve. L'oral des ENS est un long dialogue (l'épreuve dure environ cinquante minutes, comme d'ailleurs à l'École polytechnique) entre le candidat et l'examinateur, qui tout au long de l'épreuve fournit des indications, quand c'est nécessaire, pour relancer la réflexion du candidat et tester ses réactions. Il est d'ailleurs impossible de rendre pleinement compte dans un recueil d'exercices du caractère oral de l'épreuve.

L'École polytechnique, quant à elle, est plus généraliste. Les exercices posés au concours sont de facture plus classique et, en règle générale, l'examinateur intervient moins. C'est au candidat de montrer sa maîtrise du programme dans la résolution d'un exercice dont la difficulté est cependant très variable. Certains sont proches des exercices d'ENS. Les énoncés circulent d'ailleurs d'un concours à l'autre, ou peuvent même être repris d'exercices d'Olympiades.

Les énoncés qui figurent dans ce recueil ont été donnés entre 1995 et 2007. Ils sont extraits pour l'essentiel des listes publiées chaque année par la RMS (*Revue des mathématiques de l'enseignement supérieur* aux éditions Vuibert jusqu'en 2003 et désormais *Revue de la filière Mathématiques* aux éditions e.net) dont nous remercions les auteurs pour l'aide précieuse qu'ils apportent ainsi aux élèves et aux professeurs des classes préparatoires. Il s'agit de versions communiquées par les étudiants, reflétant la compréhension que ceux-ci ont eue de l'exercice et le déroulement conjoncturel de leur oral, comme le montrent les variations d'une année à l'autre pour un même exercice. Nous n'avons pas hésité à les modifier, pour rectifier des erreurs, compléter un énoncé

quand manifestement l'exercice s'est arrêté avant que le résultat que l'examinateur avait en vue ne soit atteint, ou ajouter des indications.

Nous avons choisi de laisser quelques énoncés « bruts », ceux pour lesquels nous estimons qu'une démarche naturelle (qui peut être longue et ardue) permet de conduire à la solution. Pour d'autres exercices, nous avons pris la liberté de rajouter des questions intermédiaires, qui auraient pu être celles posées par l'examinateur. Quitte à perdre en concision, nous avons tenu à rédiger les solutions les plus pédagogiques possible, essayant d'exposer clairement les idées et démarches des raisonnements sans pour autant escamoter les détails ou calculs qui peuvent paraître évidents. On évite autant que possible l'introduction d'une astuce ou d'un objet *ad hoc* permettant d'atteindre rapidement la solution. S'il n'y a pas moyen d'expliquer l'origine de cette astuce, c'est que l'exercice est peu intéressant et que l'étudiant en tirera peu de profit.

À l'intérieur de chaque chapitre, les exercices ont été regroupés thématiquement, et à l'intérieur de chaque thème, souvent par ordre de difficulté croissante. Ainsi regroupés, ils apparaîtront plus accessibles, car plongés dans leur contexte mathématique, éclairés par d'autres exercices voisins. Les introductions historiques qui ouvrent chaque chapitre, outre leur intérêt propre, visent au même but. Enfin, nous avons agrémenté les énoncés de quelques remarques préliminaires. Sans faire de rappels de cours systématiques, nous avons énoncé, voire redémontré certains résultats : lemmes classiques, intervenant dans la résolution d'un grand nombre d'exercices, ou résultats au contraire à la lisière du programme, mais utiles, pour lesquels des éclaircissements étaient nécessaires. On trouvera aussi des remarques de synthèse ou des généralisations qui, nous l'espérons, pourront amener le candidat curieux à approfondir ses connaissances. Les quelques indications bibliographiques ont le même objectif.

Le lecteur ne tirera profit de ce livre d'exercices que s'il cherche des solutions personnelles avant d'en étudier les corrigés. Une bonne connaissance du cours est indispensable. En effet, les théorèmes du programme fournissent bon nombre de schémas de démonstration. Rappelons aussi quelques démarches générales qui peuvent faciliter l'appréhension des exercices difficiles :

▷ ne pas hésiter à faire des dessins pour visualiser les phénomènes : par exemple dans le domaine de la convexité, ou encore en représentant graphiquement l'évolution d'un système dynamique discret $u_{n+1} = f(u_n)$... ;

▷ établir le résultat pour un sous-ensemble et l'étendre par un argument de densité (par exemple pour passer de $\mathbb{Q}$ à $\mathbb{R}$, de l'ensemble des fonctions polynomiales aux fonctions continues...) ;

▷ ne pas hésiter dans la phase exploratoire à faire des approximations

qui peuvent sembler grossières, en vue de déterminer une limite ou un équivalent, à faire des interversions de sommations en vue de réaliser un calcul formel. Dans un second temps, il s'agira bien sûr de justifier les conjectures.

Au-delà des étudiants en classe préparatoire, ces ouvrages intéresseront aussi les candidats au CAPES et à l'Agrégation, qui y trouveront matière à réviser les principales notions du programme, ainsi que des exemples pour nourrir un développement pour leur oral.

Voyons maintenant plus précisément le contenu de ce tome 1 d'analyse. Les thèmes développés sont les nombres réels et complexes, les suites, les séries et enfin les fonctions d'une variable réelle (limites, continuité, dérivabilité, convexité...). Hormis le troisième chapitre (qui porte sur les séries numériques), ces exercices s'adressent aussi aux élèves de première année. La matière des différents chapitres est assez classique, sauf peut-être pour le premier qui fait appel à des techniques délicates et des idées originales. Le second tome d'analyse portera sur les suites et séries de fonctions, notamment les séries entières et les séries de Fourier. Quant au troisième, il abordera la topologie, les équations différentielles et les fonctions de plusieurs variables.

Nous remercions André Bellaïche, René Cori, Rached Mneimné, Bernard Randé et Gilles Godefroy, ainsi que nos élèves, pour leur relecture approfondie de l'ouvrage et leurs nombreuses suggestions, tant sur le fond que sur la forme.

Outre les notations habituelles, indiquons que $\mathrm{E}(x)$ désigne la partie entière du réel x.

Enfin, si vous souhaitez nous contacter pour nous faire part de vos remarques, vous pouvez envoyer un courriel à l'adresse *fgn.cassini@free.fr*.

Chapitre 1

Nombres réels et complexes
Topologie de $\mathbb{R}$ et $\mathbb{C}$

Si la définition rigoureuse du corps des réels est récente, la manipulation de nombres réels remonte à la plus haute Antiquité. Elle est présente dans toute mesure de grandeur et apparaît dès la mise en place d'un système de numération capable de noter les valeurs approchées des nombres réels (mathématiques babyloniennes). La mathématique grecque s'est intéressée essentiellement aux nombres entiers ; une fraction est moins considérée comme un nombre que comme un rapport, une relation entre deux entiers. On lui doit cependant quelques résultats essentiels pour les réels. La découverte de nombres irrationnels date de la fin du V^{e} *siècle avant notre ère. On attribue à l'école pythagoricienne la démonstration de l'irrationalité de* $\sqrt{2}$. *Dans le livre V des* Éléments *d'Euclide, on trouve l'exposé d'une théorie axiomatique des rapports de grandeurs, attribuée à Eudoxe. Y figurent, par exemple, des propositions qui équivalent à la distributivité de la multiplication par rapport à l'addition ou à l'associativité de la multiplication. Notons aussi l'énoncé suivant qui correspond essentiellement à ce qu'on appelle l'axiome d'Archimède : deux grandeurs ont entre elles un rapport si on peut trouver un multiple de l'une supérieur à l'autre. Avec le déclin de la mathématique grecque, on retourne à une conception naïve des nombres réels. Chez Diophante, par exemple, un nombre est l'inconnue d'un problème algébrique. Dans les siècles suivants, les progrès apparaîtront surtout au niveau du calcul. À la fin du* XVI^{e} *siècle, Stevin contribue à la généralisation de l'usage des fractions décimales. Napier propose, en 1617, d'utiliser le point ou la virgule pour séparer la partie entière de la partie décimale. Il choisit finalement le point, qui devient d'usage commun en Angleterre. La réflexion sur les réels va évoluer avec le développement de l'analyse au* XIX^{e} *: difficultés liées aux problèmes de limites ; progrès dans la connaissance des irrationnels et des nombres transcendants... Entre 1860 et 1880, Weierstrass, Dedekind, Méray et Cantor proposent des constructions de* $\mathbb{R}$. *Celle de Dedekind par les «coupures» de* $\mathbb{Q}$ *met au premier plan l'axiome de la borne supérieure, alors que celle de Méray et de Cantor qui fait intervenir des classes d'équivalence de suites de Cauchy fournit directement le caractère complet de* $\mathbb{R}$. *C'est à partir de ses recherches sur les réels que Cantor s'intéresse aux problèmes d'équipotence. Dans ce cadre, il énonce presque incidemment les principaux résultats*

élémentaires sur la topologie de la droite, la structure des ensembles ouverts et fermés... Borel et Baire prolongent l'étude des sous-ensembles de $\mathbb{R}$: le premier définit les ensembles de mesure nulle et les ensembles appelés de nos jours boréliens, *et montre qu'on peut définir sur ces derniers une mesure qui prolonge la longueur des intervalles ; le second définit les ensembles de première catégorie (on dit aussi* maigres*), c'est-à-dire les ensembles qu'on peut recouvrir par une famille dénombrable de fermés d'intérieur vide et montre qu'un tel ensemble est d'intérieur vide (c'est le théorème de Baire que le lecteur trouvera en exercice dans le troisième tome d'analyse). Un espace topologique où cette propriété est vraie est aujourd'hui appelé un espace de Baire.*

Nous terminerons cette introduction par quelques mots sur l'histoire des nombres complexes. Au XVI^{e} *siècle, Cardan et Tartaglia donnent une solution de l'équation du troisième degré $x^3 + px + q = 0$ sous la forme*

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}$$

Bombelli (1526-1573) a alors l'idée d'utiliser cette formule dans le cas où $4p^3 + 27q^2 < 0$. Il étudie notamment l'exemple de $x^3 - 15x - 4 = 0$ et obtient

$$x = \sqrt[3]{2 + \sqrt{-21}} + \sqrt[3]{2 - \sqrt{-21}} = 2 + \sqrt{-1} + 2 - \sqrt{-1} = 4$$

après quelques manipulations algébriques sur des expressions de la forme $a + b\sqrt{-1}$ que le lecteur pourra prendre plaisir à retrouver. Or, il est aisé de vérifier que 4 est bien solution de l'équation. Ainsi, la formule de Cardan permet de trouver des racines réelles par l'intermédiaire d'opérations effectuées sur des nombres impossibles, ou «imaginaires» (terme introduit par Descartes en 1637). Ces constatations empiriques ont conduit à un usage de plus en plus fréquent des nombres imaginaires et ce dès le début du XVII^{e} *siècle. Le symbole i utilisé de nos jours est introduit par Euler (1707-1783) en 1777. Il introduit également l'exponentielle complexe et les fonctions trigonométriques usuelles. C'est Moivre (1667-1754) qui met en évidence le lien entre les racines n-ièmes d'un nombre complexe et le partage d'un arc de cercle en n parties égales. Cependant il faut encore attendre quelques années avant que ne se développe véritablement l'usage de la représentation géométrique des nombres complexes. Gauss (1777-1855) avait sans doute cette vision géométrique de $\mathbb{C}$ mais il ne publia rien sur ce point avant 1831. Ce sont deux mathématiciens amateurs qui les premiers publièrent des mémoires sur ce sujet. Le premier est dû au danois Caspar Wessel (1797), mais il tomba rapidement dans l'oubli et ne fut redécouvert qu'un siècle plus*

tard. Le second date de 1806 et est dû au mathématicien suisse Argand. Ses travaux faillirent aussi passer inaperçus (l'article ne fut diffusé qu'en 1813) et furent accueillis très froidement par les grands de l'époque. Ce n'est qu'à partir de 1831, avec Gauss, et surtout en 1847 avec Cauchy, que ses idées furent adoptées, ouvrant la porte à de multiples développements.

Les trois premiers exercices sont issus d'un même oral et concernent le développement décimal d'un nombre réel. Dans le cursus scolaire, c'est très vite qu'on apprend à se représenter un réel par son développement décimal, par exemple $\pi = 3{,}14159265...$ *Cet objet n'est toutefois pas aussi simple qu'il y paraît, puisque la notation* $0{,}a_1a_2a_3...$ *représente en réalité une somme de série. De plus, on dit rarement dans le secondaire que ce développement n'est pas nécessairement unique : un nombre décimal (non nul) en a deux. Par exemple :* $9/25 = 0{,}36 = 0{,}3599999...$ *Le premier sera qualifié de propre et le second d'impropre. L'énoncé suivant, qui démontre l'existence et l'unicité du développement propre, est une question de cours.*

1.1. Développement décimal propre d'un réel

Démontrer que tout réel x de $[0,1[$ s'écrit de manière unique $x = \sum_{n=1}^{+\infty} \frac{a_n}{10^n}$ où $(a_n)_{n\geqslant 1}$ est une suite de $[\![0,9]\!]$ non stationnaire à 9 (c'est-à-dire que pour tout $N \geqslant 1$, il existe $n \geqslant N$ tel que $a_n \neq 9$).

(École polytechnique)

▷ **Solution.**

Le résultat, énoncé ici en base 10, est en fait valable pour toute base $b \geqslant 2$. Nous traiterons donc le cas général. Nous noterons $\mathcal{T}$ l'ensemble des suites $(a_n)_{n\geqslant 1}$ de $[\![0, b-1]\!]$ non stationnaires à $b-1$. Soit x un réel de l'intervalle $[0,1[$ (on se limite à ces réels, car tout réel est somme d'un entier, sa partie entière, et d'un réel de $[0,1[$).

• *Analyse.* Soit $(a_n)_{n\geqslant 1}$ dans $\mathcal{T}$ répondant au problème,

$$x = \sum_{n=1}^{+\infty} \frac{a_n}{b^n} = \frac{a_1}{b} + \frac{a_2}{b^2} + \cdots + \frac{a_n}{b^n} + \cdots$$

On sait bien, en base 10, comment récupérer le premier chiffre après la virgule a_1 : il suffit de multiplier par 10, ce qui a pour effet de décaler la virgule d'un cran vers la droite, puis de prendre la partie entière. Effectivement, en multipliant l'égalité ci-dessus par b il vient,

$$bx = a_1 + \sum_{n=2}^{+\infty} \frac{a_n}{b^{n-1}} = a_1 + \sum_{n=1}^{+\infty} \frac{a_{n+1}}{b^n}.$$

La somme qui apparaît est un réel de $[0,1[$, puisque

$$0 \leqslant \sum_{n=1}^{+\infty} \frac{a_{n+1}}{b^n} < \sum_{n=1}^{+\infty} \frac{b-1}{b^n} = \frac{b-1}{b} \frac{1}{1-\frac{1}{b}} = 1,$$

la seconde inégalité étant stricte puisque il existe au moins un entier $n > 1$ tel que $a_n < b-1$. On a donc nécessairement $a_1 = \mathrm{E}(bx)$. On a alors $bx - a_1 = \sum_{n=1}^{+\infty} \frac{a_{n+1}}{b^n}$, de sorte que par le même raisonnement, $a_2 = \mathrm{E}(bx - a_1)$. Plus généralement, si les a_k sont connus jusqu'au rang $p-1$ avec $p \geqslant 2$, on a

$$b^p x = a_1 b^{p-1} + a_2 b^{p-2} + \cdots + a_{p-1} b + a_p + \sum_{n=p+1}^{+\infty} \frac{a_n}{b^{n-p}},$$

soit

$$a_p = b^p x - a_1 b^{p-1} - a_2 b^{p-2} - \cdots - a_{p-1} b - \underbrace{\sum_{n=1}^{+\infty} \frac{a_{n+p}}{b^n}}_{\in [0,1[}.$$

Par conséquent $\boxed{a_p = \mathrm{E}\left(xb^p - a_1 b^{p-1} - a_2 b^{p-2} - \cdots - a_{p-1} b\right)}$. Si elle existe, la suite (a_n) est donc déterminée de manière unique.

• *Synthèse.* Le point précédent nous a fourni l'algorithme qui permet de définir les entiers a_p par récurrence. Il n'y a plus qu'à vérifier que cette suite convient. On considère donc la suite $(a_p)_{p \geqslant 1}$ définie par $a_1 = \mathrm{E}(xb)$ et $a_p = \mathrm{E}\left(xb^p - a_1 b^{p-1} - a_2 b^{p-2} - \cdots - a_{p-1} b\right)$ pour tout $p \geqslant 2$.
Nous allons démontrer successivement que :

(i) $\quad x = \sum_{n=1}^{+\infty} \frac{a_n}{b^n}$;
(ii) pour tout $p \geqslant 1, 0 \leqslant a_p \leqslant b-1$;
(iii) la suite (a_p) n'est pas stationnaire à $b-1$.

(i) Pour $p \geqslant 2$, on a par définition de la partie entière,

$$a_p \leqslant xb^p - a_1 b^{p-1} - a_2 b^{p-2} - \cdots - a_{p-1} b < a_p + 1$$

ce qui donne en divisant par b^p,

$$\frac{a_p}{b^p} \leqslant x - \sum_{n=1}^{p-1} \frac{a_n}{b^n} < \frac{a_p + 1}{b^p}, \quad \text{ou encore } \quad 0 \leqslant x - \sum_{n=1}^{p} \frac{a_n}{b^n} < \frac{1}{b^p} \quad (*).$$

Comme $\lim_{p\to+\infty} \frac{1}{b^p} = 0$, on en déduit que $x = \sum_{n=1}^{+\infty} \frac{a_n}{b^n}$.

(ii) Comme $x \in [0,1[$, $xb \in [0,b[$ et $0 \leqslant a_1 \leqslant b-1$. Soit $p \geqslant 2$. D'après $(*)$ écrite pour $p-1$ au lieu de p, on a

$$0 \leqslant x - \sum_{n=1}^{p-1} \frac{a_n}{b^n} < \frac{1}{b^{p-1}} \quad \text{et} \quad 0 \leqslant xb^p - \sum_{n=1}^{p-1} a_n b^{p-n} < b,$$

après multiplication par b^p. Comme l'entier a_p est la partie entière de $xb^p - \sum_{n=1}^{p-1} a_n b^{p-n}$, on a bien $0 \leqslant a_p \leqslant b-1$.

(iii) Supposons, par l'absurde, qu'il existe $N \geqslant 3$ tel que pour tout $n \geqslant N$, $a_n = b-1$. On a alors

$$\begin{aligned} x &= \sum_{n=1}^{+\infty} \frac{a_n}{b^n} = \sum_{n=1}^{N-1} \frac{a_n}{b^n} + \sum_{n=N}^{+\infty} \frac{b-1}{b^n} = \sum_{n=1}^{N-1} \frac{a_n}{b^n} + \frac{1}{b^N} \frac{b-1}{1-1/b} \\ &= \sum_{n=1}^{N-1} \frac{a_n}{b^n} + \frac{1}{b^{N-1}}. \end{aligned}$$

Ainsi $x - \sum_{n=1}^{N-1} \frac{a_n}{b^n} = \frac{1}{b^{N-1}}$, ce qui contredit l'inégalité $(*)$ pour $p = N-1$.

Conclusion. Pour tout $x \in [0,1[$, il existe une unique suite $(a_n)_{n\geqslant 1}$ de $\mathcal{T}$ telle que

$$x = \sum_{n=1}^{+\infty} \frac{a_n}{b^n}. \quad \triangleleft$$

On écrit alors $x = 0,a_1a_2...a_n...$ et on parle du développement propre de x en base b. Pour $b = 10$ on parle du développement décimal, pour $b = 2$ du développement dyadique et pour $b = 3$ du développement triadique. Pour un réel positif quelconque, on rajoute sa partie entière décomposée en base b à gauche de la virgule. Enfin pour un réel négatif, on met un signe moins devant le développement de sa valeur absolue.

Il est naturel de se demander ce qui se passe si on autorise les suites stationnaires à 9, ou à $b-1$ dans le cas général. Un calcul aisé montre que $0,a_1a_2...a_p9999...$ avec $a_p < 9$ est en fait égal à $0,a_1...a_{p-1}a'_p$ avec $a'_p = a_p + 1$ (pour $0,9999...$ on obtient 1). C'est un nombre décimal, i.e. *un rationnel de la forme $\frac{a}{10^k}$, $a \in \mathbb{Z}$. Un nombre décimal (non nul) admet donc exactement deux développements : son développement propre, qui est fini, et un unique développement stationnaire à 9, qualifié d'impropre. Dans le cas d'une base b quelconque les rationnels de la forme $\frac{a}{b^k}$, qui admettent deux développements, sont parfois appelés rationnels b-adiques.*

L'exercice suivant est extrait du même oral que le précédent. Un nombre rationnel x est déterminé par un nombre fini d'informations, puisque si on l'écrit sous la forme a/b, les entiers a et b sont par exemple déterminés par leurs chiffres en base 10 qui sont en nombre fini. Le développement décimal de x va également être déterminé par un nombre fini de données : il est périodique.

1.2. Caractérisation des rationnels

Soit $x \in [0,1[$. Montrer que x est rationnel si, et seulement si, le développement décimal propre de x est périodique à partir d'un certain rang.

(École polytechnique)

▷ **Solution.**

Nous allons encore démontrer le résultat en base $b \geqslant 2$ quelconque. Nous noterons $\mathcal{T}$ l'ensemble des suites $(a_n)_{n\geqslant 1}$ de $[\![0, b-1]\!]$ non stationnaires à $b-1$. Nous savons qu'il existe une unique suite $(a_n)_{n\geqslant 1}$ de $\mathcal{T}$ telle que $x = \sum_{n=1}^{+\infty} \frac{a_n}{b^n} = 0{,}a_1a_2a_3\ldots$

• Supposons tout d'abord le développement périodique à partir d'un certain rang : il existe $\mathrm{N} \geqslant 1$ et $d \geqslant 1$ tels que pour tout $n > \mathrm{N}$, $a_{n+d} = a_n$. On peut écrire

$$x = 0{,}a_1a_2\ldots a_{\mathrm{N}}\underbrace{a_{\mathrm{N}+1}\ldots a_{\mathrm{N}+d}}_{\text{période}}\underbrace{a_{\mathrm{N}+1}\ldots a_{\mathrm{N}+d}}\cdots$$

On va simplement calculer la somme de la série définissant x en regroupant les termes par paquets. On convient ici que $\overline{a_{\mathrm{N}+1}a_{\mathrm{N}+2}\ldots a_{\mathrm{N}+d}}$ désigne l'entier $a_{\mathrm{N}+d} + a_{\mathrm{N}+d-1}b + \cdots + a_{\mathrm{N}+1}b^{d-1}$ (c'est simplement l'écriture en base b de cet entier). On a alors

$$x = \frac{\overline{a_1\ldots a_{\mathrm{N}}}}{b^{\mathrm{N}}} + \sum_{k=1}^{+\infty} \frac{\overline{a_{\mathrm{N}+1}a_{\mathrm{N}+2}\ldots a_{\mathrm{N}+d}}}{b^{\mathrm{N}+kd}}.$$

La somme de la série géométrique $\sum_{k=1}^{+\infty} \frac{1}{b^{kd}}$ vaut $\frac{1}{b^d-1}$ et on obtient donc

$$x = \frac{\overline{a_1\ldots a_{\mathrm{N}}}}{b^{\mathrm{N}}} + \frac{\overline{a_{\mathrm{N}+1}a_{\mathrm{N}+2}\ldots a_{\mathrm{N}+d}}}{b^{\mathrm{N}}(b^d-1)},$$

qui est bien rationnel. Par exemple, en base $b = 10$, le rationnel dont le développement est $0{,}456123123123\ldots$ est $\frac{456}{10^3} + \frac{123}{10^3 \cdot 999}$.

• Inversement, supposons que x est rationnel et non nul (pour $x = 0$ le résultat est clair). Écrivons $x = \frac{p}{q}$ avec p et q dans $\mathbb{N}^*$ premiers entre eux. On va simplement appliquer l'algorithme donné dans l'exercice précédent, qui n'est rien d'autre que l'algorithme de division que l'on apprend dans les petites classes : tout le monde sait trouver le développement décimal de $\frac{9}{14}$. On multiplie 9 par 10 et on prend la partie entière qui vaut 6. C'est la première décimale après la virgule. On a alors $\frac{90}{14} - 6 = \frac{6}{14}$. On remultiplie par 10 et on obtient la deuxième décimale. C'est un 4. On continue ainsi et forcément, les numérateurs des fractions obtenues étant compris entre 1 et 13 on finit par tomber sur une fraction déjà obtenue. La suite devient périodique à partir de ce moment. Pour l'exemple de $\frac{9}{14}$ on obtient

$$\frac{9}{14} = 0{,}6\underbrace{428571}_{\text{période}}428571...$$

Dans le cas général, c'est pareil. La première décimale a_1 est la partie entière de $\frac{bp}{q}$, c'est-à-dire le quotient dans la division euclidienne de bp par q. Pour la seconde décimale on recommence alors avec $\frac{r}{q}$ où r est le reste de la division euclidienne de bp par q. Les restes modulo q étant compris entre 0 et $q-1$ la suite devient périodique à partir d'un certain rang. ◁

Le résultat de cet exercice fut établi par Wallis en 1693. Un problème intéressant est de déterminer la période. Nous invitons le lecteur à préciser les faits suivants. Dans le cas d'un rationnel b-adique (d'un décimal en base 10) le développement va être fini : dans l'algorithme ci-dessus on va obtenir un reste nul et à partir de ce moment toutes les décimales seront nulles. En fait dans ce cas, le développement est immédiatement donné par l'écriture en base b du numérateur. Par exemple, pour pour $x = \frac{17}{25}$ *et* $b = 10$, *on écrit* $x = \frac{17 \cdot 4}{10^2} = \frac{68}{10^2} = 0{,}68$. *Supposons maintenant que le dénominateur q de x est premier avec la base b. On va chercher le plus petit multiple de q de la forme* $b^d - 1$, *c'est-à-dire le plus petit exposant d telle que* $b^d \equiv 1\ [q]$. *Une telle puissance existe car la classe de b est inversible dans l'anneau* $\mathbb{Z}/q\mathbb{Z}$. *L'entier d recherché est l'ordre de cette classe dans le groupe des inversibles de* $\mathbb{Z}/q\mathbb{Z}$. *Par exemple si* $b = 10$ *et* $q = 7$, *on voit que* $\overline{3}$ *est d'ordre 6 dans* $(\mathbb{Z}/7\mathbb{Z})^*$ *(c'est un générateur). La période des fractions* $\frac{p}{7}$, $1 \leqslant p \leqslant 6$ *est égale à 6. Par exemple,*

$$\frac{1}{7} = \frac{142857}{7 \cdot 142857} = \frac{142857}{999999} = \frac{142857}{10^6 - 1} = 0{,}142857142857142857...$$

Dans ce cas le développement du rationnel $\frac{p}{q}$ est périodique dès le début. Supposons maintenant que le pgcd de q et b n'est pas égal à 1. On peut écrire x sous la forme $\frac{p'}{b^N q'}$ avec $\mathrm{pgcd}(b, q') = 1$ et b ne divisant par p'. On fait la division euclidienne de p' par q' pour se ramener à $x = \frac{a}{b^N} + \frac{r}{b^N q'}$ et on est ramené au cas précédent. Par exemple, si $x = \frac{39}{280}$ avec toujours $b = 10$, on écrit

$$\frac{39}{280} = \frac{39}{2^3 \cdot 5 \cdot 7} = \frac{39 \cdot 25}{10^3 \cdot 7} = \frac{975}{10^3 \cdot 7} = \frac{139}{10^3} + \frac{2}{10^3 \cdot 7}.$$

Le décimal $\frac{139}{10^3}$ donne les trois premiers chiffres après la virgule et le développement devient périodique à partir du rang 4 :

$$x = \frac{139}{10^3} + \frac{285714}{10^3 \cdot (10^6 - 1)} = 0{,}139285714285714...$$

L'énoncé suivant propose une première application des développements décimaux. On y utilise le procédé diagonal de Cantor pour démontrer que $\mathbb{R}$ n'est pas dénombrable.

1.3. Non-dénombrabilité de $\mathbb{R}$

Soit $f : \mathbb{N}^* \to [0, 1[$. Pour chaque $k \geqslant 1$, on appelle u_k la k-ième décimale dans le développement décimal propre de $f(k)$ et on pose $v_k = 0$ si $u_k = 1$ et $v_k = 1$ sinon. Montrer que $y = 0{,}v_1 v_2 ... v_n ...$ n'a pas d'antécédent par f. Conclusion ?

(École polytechnique)

▷ **Solution.**

Écrivons l'un sous l'autre les développements des réels $f(n)$, pour $n \geqslant 1$:

$$\begin{aligned}
f(1) &= 0,\ u_1 \cdots \\
f(2) &= 0,\ \bullet\ u_2 \cdots \\
f(3) &= 0,\ \bullet\ \bullet\ u_3 \cdots \\
&\vdots \qquad\qquad \ddots \\
f(k) &= 0,\ \bullet\ \cdots\cdots\ \bullet\ u_k \cdots
\end{aligned}$$

Le réel $y = 0{,}v_1 v_2 ... v_n ...$ est bien défini par son développement décimal propre car pour tout k, $v_k \neq 9$ (voir l'exercice 1.1). Par unicité du développement décimal propre on a $y \neq f(k)$ pour tout entier k

puisque la k-ième décimale de y est v_k et que $v_k \neq u_k$. Donc y n'a pas d'antécédent par f. Ainsi il n'existe aucune surjection de $\mathbb{N}$ sur $[0,1[$ et, *a fortiori*, sur $\mathbb{R}$. On conclut que $\mathbb{R}$ n'est pas dénombrable. $\lhd$

Un théorème général de Cantor assure qu'un ensemble E *quelconque n'est jamais équipotent à* $\mathcal{P}(\mathrm{E})$. *En effet, supposons qu'il existe une surjection* $f : \mathrm{E} \to \mathcal{P}(\mathrm{E})$ *et considérons la partie* $\mathrm{A} = \{x \in \mathrm{E}, x \notin f(x)\}$. *Soit* a *l'antécédent de* A *par* f. *Pour tout* x *de* E, *on a* $x \in \mathrm{A}$ *si et seulement si* $x \notin f(x)$, *autrement dit* $x \in f(a)$ *si et seulement si* $x \notin f(x)$. *Or cela donne une contradiction si on prend* $x = a$. *C'est sur le même type de raisonnement qu'est basé le paradoxe de Russell : dans le cadre axiomatique de la théorie des ensembles, supposer l'existence de l'ensemble de tous les ensembles conduit à une contradiction. En particulier, l'ensemble* $\mathcal{P}(\mathbb{N})$ *n'est pas dénombrable (on peut démontrer qu'il est en fait équipotent à* $\mathbb{R}$*). L'exercice suivant demande de construire une injection particulière de* $\mathbb{R}$ *dans* $\mathcal{P}(\mathbb{N})$. *Il peut être résolu à l'aide des développements décimaux.*

1.4. Injection de $\mathbb{R}$ dans $\mathcal{P}(\mathbb{N})$

Construire une injection $\psi : \mathbb{R} \to \mathcal{P}(\mathbb{N})$ telle que, pour tout x, $\psi(x)$ est infini, et pour tout couple $(x, y) \in \mathbb{R}^2$ tel que $x \neq y$, $\psi(x) \cap \psi(y)$ est fini.

(École normale supérieure)

Solution.

Observons pour commencer qu'une application $\psi : \mathbb{R} \to \mathcal{P}(\mathbb{N})$ telle que $\psi(x)$ est infini pour tout x et $\psi(x) \cap \psi(y)$ fini pour $x \neq y$ est nécessairement injective.

Pour définir ψ on va utiliser les développements décimaux, l'idée étant d'associer à un réel x, la suite des entiers qu'on lit après la virgule. Par exemple, à $\pi = 3{,}1415926...$ on a envie d'associer l'ensemble $\{1, 14, 141, 1415, 14159, ...\}$. Évidemment, deux réels qui ont la même partie fractionnaire donneront le même ensemble. On va donc se limiter à l'intervalle $]0,1[$, ce qui ne pose pas de problème puisqu'il est équipotent à $\mathbb{R}$ (il est facile de donner une bijection explicite entre $]0,1[$ et $\mathbb{R}$).

À tout réel x de l'intervalle $]0,1[$ on associe donc le sous-ensemble $\psi(x) = \{\mathrm{E}(10^n x),\ n \geqslant 1\}$. Comme la suite $(\mathrm{E}(10^n x))_{n \geqslant 1}$ tend vers l'infini, $\psi(x)$ est une partie infinie de $\mathbb{N}$. Il y a toutefois un problème car la seconde condition n'est pas vérifiée. En effet, si $x = 0{,}0a_1a_2... \in]0,1[$ le réel $\dfrac{x}{10} = 0{,}00a_1a_2...$ a la même image par ψ que x. Une manière simple d'éviter ce problème est de se limiter à l'intervalle $[1/10, 1[$ (qui

est toujours équipotent à $\mathbb{R}$), car tous les réels de cet intervalle ont leur premier chiffre après la virgule non nul. La restriction de l'application ψ à cet intervalle va alors vérifier la seconde condition par unicité du développement décimal propre (cf. exercice 1.1). En effet, si $x \in [1/10, 1[$, pour tout $n \geqslant 1$, $\mathrm{E}(10^n x)$ est l'unique entier de $\psi(x)$ qui s'écrit avec n chiffres (en base 10). Soit x et y deux réels de $[1/10, 1[$ tels que $\psi(x) \cap \psi(y)$ soit infini. On peut donc trouver $\varphi : \mathbb{N} \to \mathbb{N}^*$ strictement croissante telle que $\mathrm{E}(10^{\varphi(n)}x) = \mathrm{E}(10^{\varphi(n)}y)$ pour tout n. On en déduit que

$$x = \lim_{n \to +\infty} \frac{\mathrm{E}(10^{\varphi(n)}x)}{10^{\varphi(n)}} = \lim_{n \to +\infty} \frac{\mathrm{E}(10^{\varphi(n)}y)}{10^{\varphi(n)}} = y.$$

L'application ψ ainsi construite répond donc au problème posé. ◁

On peut observer que si $\psi(x) \cap \psi(y)$ est infini, on a nécessairement $\psi(x) = \psi(y)$. En effet, si $\mathrm{E}(10^n x) = \mathrm{E}(10^n y)$ pour une valeur de n on a aussi $\mathrm{E}(10^p x) = \mathrm{E}(10^p y)$ pour tout $p \leqslant n$ puisque $\mathrm{E}(10^p x)$ est obtenu en gardant les p chiffres les plus à gauche de $\mathrm{E}(10^n x)$.

Dans l'exercice suivant nous proposons plusieurs solutions dont l'une utilise les développements dyadiques.

1.5. Calcul d'une somme

Soit $x > 0$ un réel. Calculer $\displaystyle\sum_{n=0}^{+\infty} \mathrm{E}\left(\frac{x}{2^{n+1}} + \frac{1}{2}\right)$, où E désigne la partie entière.

(École normale supérieure)

▷ **Solution.**

• *Première solution.* Notons que la somme est en réalité finie, puisque la quantité dont on prend la partie entière tend vers $\frac{1}{2}$ lorsque n tend vers l'infini. Toutes ces puissances de 2 invitent assez naturellement à considérer le développement dyadique (propre) de x,

$$x = a_p 2^p + a_{p-1} 2^{p-1} + \cdots + a_1 2 + a_0 + \sum_{i=1}^{+\infty} \frac{a_{-i}}{2^i},$$

où les a_i sont dans $\{0, 1\}$. Le réel x est alors dans l'intervalle $]0, 2^{p+1}[$. Ainsi pour $k \geqslant p+1$, on a $0 < \dfrac{x}{2^{k+1}} + \dfrac{1}{2} < 1$ et donc $\mathrm{E}\left(\dfrac{x}{2^{k+1}} + \dfrac{1}{2}\right) = 0$. Prenons $k \in [\![0, p]\!]$. On a

$$\frac{x}{2^{k+1}} = \underbrace{a_p 2^{p-k-1} + \cdots + a_{k+1}}_{\in\mathbb{N}} + \frac{a_k}{2} + \underbrace{\frac{a_{k-1}}{2^2} + \cdots + \frac{a_0}{2^{k+1}} + \sum_{i=1}^{+\infty} \frac{a_{-i}}{2^{i+k+1}}}_{<1/2}$$

(la première partie n'apparaît pas lorsque $k = p$). La partie entière de $\frac{x}{2^{k+1}} + \frac{1}{2}$ vaut donc $a_p 2^{p-k-1} + \cdots + a_{k+1}$ si $a_k = 0$ et elle est égale à $a_p 2^{p-k-1} + \cdots + a_{k+1} + 1$ si $a_k = 1$. Dans les deux cas elle vaut

$$a_p 2^{p-k-1} + \cdots + a_{k+1} + a_k$$

(pour $k = p$, cela se réduit à a_p). Il en résulte que la somme recherchée vaut

$$\begin{aligned}\sum_{k=0}^{p-1}(a_p 2^{p-k-1} + \cdots + a_{k+1}) + \sum_{k=0}^{p} a_k &= \sum_{k=1}^{p} a_k \left(\sum_{l=0}^{k-1} 2^l\right) + \sum_{k=0}^{p} a_k \\ &= \sum_{k=0}^{p} a_k 2^k = \mathrm{E}(x).\end{aligned}$$

On a donc pour tout $x > 0$, $\boxed{\displaystyle\sum_{n=0}^{+\infty} \mathrm{E}\left(\frac{x}{2^{n+1}} + \frac{1}{2}\right) = \mathrm{E}(x)}$.

• *Seconde solution.* Elle est plus rapide et n'utilise pas les développements dyadiques, mais elle suppose la connaissance du résultat. Si on note $f(x)$ la somme à calculer, on observe que $f(2x) = f(x) + \mathrm{E}\left(x + \frac{1}{2}\right)$. Le lecteur vérifiera aisément que pour tout réel x, on a $\mathrm{E}(x) + \mathrm{E}\left(x + \frac{1}{2}\right) = \mathrm{E}(2x)$. Il en résulte que si $f(x) = \mathrm{E}(x)$ pour un $x > 0$, alors $f(2x) = \mathrm{E}(2x)$. Par suite, il suffit de vérifier que $f(x) = 0$ pour $x \in [0, 1[$, ce qui est immédiat.

• *Troisième solution.* Elle nous a été transmise par un lecteur et est de nature combinatoire. Montrons que $\mathrm{E}\left(\frac{x}{2^{k+1}} + \frac{1}{2}\right)$ correspond au nombre d'entiers m compris entre 1 et x de la forme $m = 2^k(2m' - 1)$ (c'est-à-dire dont la valuation en 2 est égale à k). En effet, on a

$$\begin{aligned}1 \leqslant 2^k(2m' - 1) \leqslant x &\iff 1 + 2^k \leqslant 2^{k+1} m' \leqslant x + 2^k \\ &\iff \frac{1 + 2^k}{2^{k+1}} \leqslant m' \leqslant \frac{x}{2^{k+1}} + \frac{1}{2}.\end{aligned}$$

Comme $0 \leqslant \frac{1 + 2^k}{2^{k+1}} \leqslant 1$, on en déduit que

$$1 \leqslant 2^k(2m' - 1) \leqslant x \iff 1 \leqslant m' \leqslant \mathrm{E}\left(\frac{x}{2^{k+1}} + \frac{1}{2}\right).$$

Comme $m' \longmapsto 2^k(2m'-1)$ est injective, on a le résultat annoncé.

Or, pour tout $m \in \mathbb{N}$ tel que $1 \leqslant m \leqslant x$, il existe un unique couple (k, m') tel que $m = 2^k(2m'-1)$. En faisant une partition de l'ensemble des entiers m compris entre 1 et x selon la valeur de k, on obtient

$$\sum_{k=0}^{+\infty} \mathrm{E}\left(\frac{x}{2^{k+1}} + \frac{1}{2}\right) = \mathrm{E}(x),$$

puisque $\mathrm{E}(x)$ est le nombre d'entiers entre 1 et x. $\lhd$

Voici maintenant une autre manière, due à Engel, de représenter les réels par une suite d'entiers. La caractérisation des rationnels dans cette représentation est particulièrement simple.

1.6. Développement en série de Engel

Soit $x \in]0,1]$.

1. Montrer qu'il existe une unique suite croissante d'entiers $(u_n)_{n\geqslant 0}$ telle que $u_0 \geqslant 2$ et $x = \sum_{n=0}^{+\infty} \frac{1}{u_0 u_1 \dots u_n}$. Il s'agit du développement en *série de Engel* du réel x.

2. Montrer que x est rationnel si et seulement si la suite $(u_n)_{n\geqslant 0}$ est stationnaire.

(École normale supérieure, École polytechnique)

▷ **Solution.**

1. Notons S l'ensemble des suites croissantes $(u_n)_{n\geqslant 0}$ de $\mathbb{N} \setminus \{0,1\}$. Soit $u = (u_n)_{n\geqslant 0} \in \mathrm{S}$. On a $\frac{1}{u_0 u_1 \dots u_k} \leqslant \frac{1}{2^{k+1}}$. La série géométrique $\sum \frac{1}{2^n}$ étant convergente, le théorème de comparaison des séries à termes positifs garantit l'existence de

$$\psi(u) = \sum_{k=0}^{+\infty} \frac{1}{u_0 u_1 \dots u_k}.$$

La somme est évidemment strictement positive et majorée par $\sum_{k=0}^{+\infty} \frac{1}{2^{k+1}} = 1$. L'exercice demande de prouver que $\psi : \mathrm{S} \to]0,1]$ est une application bijective. On se donne $x \in]0,1]$ et on va prouver que x admet un unique antécédent par ψ.

Supposons que $u \in \mathrm{S}$ vérifie $\psi(u) = x$. En multipliant cette égalité par u_0, on obtient

$$xu_0 = 1 + \sum_{k=1}^{+\infty} \frac{1}{u_1 \dots u_k} \leqslant 1 + \sum_{k=0}^{+\infty} \frac{1}{u_0 u_1 \dots u_k}$$

l'inégalité provenant de ce que la suite (u_n) est croissante. Il en résulte que $0 < xu_0 - 1 \leqslant x$ et donc que $u_0 = \mathrm{E}\left(\frac{1}{x}\right) + 1$. Si on pose

$$x_1 = xu_0 - 1 = \sum_{k=1}^{+\infty} \frac{1}{u_1 \dots u_k}$$

on a $x_1 = \psi((u_{n+1})_{n\geqslant 0})$. Le même raisonnement appliqué à x_1 donne $u_1 = \mathrm{E}\left(\frac{1}{x_1}\right)+1$. Une récurrence immédiate, montre que la suite $(u_n)_{n\geqslant 0}$ est nécessairement déterminée par $x_0 = x$ et pour tout n,

$$u_n = \mathrm{E}\left(\frac{1}{x_n}\right) + 1, \quad \text{et} \quad x_{n+1} = u_n x_n - 1.$$

Cela montre que x admet au plus un antécédent par ψ.

Inversement, montrons que les relations de récurrence ci-dessus définissent bien deux suites $(x_n)_{n\geqslant 0}$ et $(u_n)_{n\geqslant 0}$. Prouvons pour cela que x_n ne s'annule pas en vérifiant par récurrence sur n, que $x_n \in]0,1]$ pour tout n. C'est vrai au rang 0 par hypothèse. Si $x_n > 0$, on a, par définition de la partie entière, $\frac{1}{x_n} < u_n \leqslant \frac{1}{x_n} + 1$, et donc $0 < x_{n+1} = u_n x_n - 1 \leqslant x_n \leqslant 1$. D'où le résultat. On voit de plus que la suite (x_n) est décroissante et comme $u_n = \mathrm{E}\left(\frac{1}{x_n}\right) + 1$, la suite (u_n) est croissante. De plus $u_0 \geqslant 2$ donc $(u_n)_{n\geqslant 0}$ est bien un élément de S. Il ne reste plus qu'à vérifier que $\psi(u) = x$. On a $x_1 = u_0 x_0 - 1$, soit encore $x = x_0 = \frac{1}{u_0} + \frac{x_1}{u_0}$. La relation $x_2 = u_1 x_1 - 1$ permet de remplacer x_1 par $\frac{1}{u_1} + \frac{x_2}{u_1}$. On obtient $x = \frac{1}{u_0} + \frac{1}{u_0 u_1} + \frac{x_2}{u_0 u_1}$. On continue en remplaçant x_2. Une récurrence facile montre que pour tout $n \geqslant 0$,

$$x = x_0 = \frac{1}{u_0} + \frac{1}{u_0 u_1} + \dots + \frac{1}{u_0 u_1 \dots u_n} + \frac{x_{n+1}}{u_0 u_1 \dots u_n}.$$

La suite (u_n) étant croissante, on peut minorer ses termes par 2 et majorer x_{n+1} par 1. Il en résulte que le reste $\frac{x_{n+1}}{u_0 u_1 \dots u_n}$ est dominé par $\frac{1}{2^{n+1}}$ et tend donc vers 0. Ainsi,

$$\psi(u) = \sum_{k=0}^{+\infty} \frac{1}{u_0 u_1 \dots u_k} = x.$$

L'application ψ est donc une bijection de S sur $]0,1]$.

2. Supposons pour commencer le développement en série de Engel de x stationnaire à partir du rang N. On a alors

$$x = \frac{1}{u_0} + \frac{1}{u_0u_1} + \cdots + \frac{1}{u_0u_1\ldots u_{N-1}}\left(\sum_{k=0}^{+\infty}\frac{1}{u_N^k}\right)$$
$$= \frac{1}{u_0} + \frac{1}{u_0u_1} + \cdots + \frac{1}{u_0u_1\ldots u_{N-1}}\,\frac{u_N}{u_N-1} \in \mathbb{Q}.$$

Réciproquement, supposons que x est rationnel. On pose $x = \frac{p}{q}$, avec $p \leqslant q$, deux entiers naturels premiers entre eux. On a $\frac{1}{x} = \frac{q}{p}$. Effectuons la division euclidienne de q par p : $q = \alpha p + r$ avec $0 \leqslant r < p$. On a alors $u_0 = \alpha + 1$. Il en résulte que $x_1 = (\alpha+1)\frac{p}{q} - 1 = \frac{p-r}{q}$. Répétant cette opération, on en déduit que pour tout n, x_n s'écrit sous la forme $x_n = \frac{k_n}{q}$ où (k_n) est une suite décroissante d'entiers naturels non nuls. Cette suite est forcément stationnaire et la suite (u_n) aussi. ◁

Comme $e - 2 = \sum_{k=0}^{+\infty}\frac{1}{(n+2)!}$, *le développement en série de Engel de* $e-2$ *est donné par la suite* $u_n = n+2$. *On retrouve le fait que* $e \notin \mathbb{Q}$ *(voir l'exercice 1.8 pour une autre preuve de ce fait). L'algorithme permettant d'obtenir la suite* $\psi^{-1}(x)$ *est connu sous le nom d'algorithme de Briggs (1561-1630).*

1.7. Représentation des réels > 0 comme sommes de séries

Soit $(u_n)_{n\geqslant 0}$ une suite strictement décroissante de réels strictement positifs. On pose $S = \sum_{n=0}^{+\infty} u_n$; S est donc un élément de $]0, +\infty]$.

1. Déterminer une condition nécessaire et suffisante sur $(u_n)_{n\geqslant 0}$ pour que tout réel de $]0, S[$ puisse s'écrire d'au moins une manière sous la forme $\sum_{n=0}^{+\infty} u_{\varphi(n)}$, φ étant une extraction, c'est-à-dire une application strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$.

2. Déterminer une condition nécessaire et suffisante sur $(u_n)_{n\geqslant 0}$ pour qu'il y ait, quel que soit le réel de $]0, S[$, existence et unicité de l'écriture précédente.

(École normale supérieure)

Solution.

1. Cherchons des conditions nécessaires. La suite $(u_n)_{n\geqslant 0}$ décroît et est minorée par 0 ; elle converge donc vers un réel $\ell \geqslant 0$. Si ℓ n'est pas nul, aucune suite extraite de (u_n) ne converge vers 0 et pour toute extraction φ, $\sum_{n=0}^{+\infty} u_{\varphi(n)} = +\infty$. Aucun élément de $]0,S[$ n'a l'écriture voulue. On doit donc avoir $\lim\limits_{n\to+\infty} u_n = 0$.

Pour tout $n \in \mathbb{N}$, u_n appartient à $]0,S[$. S'il s'écrit $u_n = \sum\limits_{k=0}^{+\infty} u_{\varphi(k)}$, alors $\varphi(\mathbb{N}) \subset [\![n+1, +\infty[\![$, car pour tout $k \leqslant n$, on a $u_k > u_n$. On en déduit que $u_n \leqslant \sum\limits_{k=n+1}^{+\infty} u_k$.

On obtient donc les deux conditions suivantes :

(*i*) $\lim\limits_{n\to+\infty} u_n = 0$,

(*ii*) pour tout n de $\mathbb{N}$, $u_n \leqslant \sum\limits_{k=n+1}^{+\infty} u_k$.

Montrons que ces deux conditions sont suffisantes. Soit α dans $]0,S[$. On va construire une extraction φ telle que pour tout entier naturel n, $\sum_{k=0}^{n} u_{\varphi(k)} < \alpha$ en choisissant à chaque fois le plus grand terme possible.

• On prend pour $\varphi(0)$ le plus petit des entiers j tels que $u_j < \alpha$. Un tel entier existe car $\lim\limits_{n\to+\infty} u_n = 0$.

• Si $\varphi(0), \varphi(1), \ldots, \varphi(n)$ sont déterminés et vérifient $\sum\limits_{k=0}^{n} u_{\varphi(k)} < \alpha$, on définit $\varphi(n+1)$ comme le plus petit des entiers j tels que $j > \varphi(n)$ et $u_j < \alpha - \sum\limits_{k=0}^{n} u_{\varphi(k)}$. On a alors $\sum\limits_{k=0}^{n+1} u_{\varphi(k)} < \alpha$.

La série de terme général $u_{\varphi(k)}$ converge et $\sum\limits_{k=0}^{+\infty} u_{\varphi(k)} \leqslant \alpha$. Montrons que $\sum\limits_{k=0}^{+\infty} u_{\varphi(k)} = \alpha$. Remarquons que, par définition, si $\varphi(n+1) > \varphi(n)+1$, alors

$$u_{\varphi(n+1)-1} \geqslant \alpha - \sum_{k=0}^{n} u_{\varphi(k)} > \alpha - \sum_{k=0}^{+\infty} u_{\varphi(k)} \geqslant 0.$$

Cette inégalité fournit un contrôle de l'écart entre α et $\sum\limits_{k=0}^{n} u_{\varphi(k)}$. On va discuter selon qu'elle est réalisée pour une infinité d'indices n ou non. Posons pour cela $\mathrm{I} = \{n \in \mathbb{N}.\, \varphi(n+1) > \varphi(n)+1\}$.

• Si I est infini, il existe une extraction ψ telle que pour tout entier n,

$\psi(n) \in \mathrm{I}$. On a alors, pour tout entier n, $0 \leqslant \alpha - \sum_{k=0}^{+\infty} u_{\varphi(k)} \leqslant u_{\varphi(\psi(n)+1)-1}$. La suite $(u_{\varphi(\psi(n)+1)-1})_{n\geqslant 0}$, sous-suite de la suite $(u_n)_{n\geqslant 0}$ converge comme elle vers 0. On en déduit que $\sum_{k=0}^{+\infty} u_{\varphi(k)} = \alpha$.

• Si I est fini, notons n_0 son plus grand élément (si $\mathrm{I} = \emptyset$, on pose $n_0 = -1$). Puisque $\varphi(n+1) = \varphi(n) + 1$ si $n \geqslant n_0 + 1$, on a

$$\sum_{k=n_0+1}^{+\infty} u_{\varphi(k)} = \sum_{k=\varphi(n_0+1)}^{+\infty} u_k.$$

Or d'après la condition (ii), $\sum_{k=\varphi(n_0+1)}^{+\infty} u_k \geqslant u_{\varphi(n_0+1)-1}$ (si $\mathrm{I} = \emptyset$ on a $\varphi(0) > 0$, car sinon $\alpha = \mathrm{S}$). Puisque n_0 appartient à I on a, d'autre part, $u_{\varphi(n_0+1)-1} \geqslant \alpha - \sum_{k=0}^{n_0} u_{\varphi(k)}$. On déduit de ces deux inégalités que $\sum_{k=0}^{+\infty} u_{\varphi(k)} \geqslant \alpha$ et donc que $\sum_{k=0}^{+\infty} u_{\varphi(k)} = \alpha$.

2. Montrons qu'il y a, pour tout $\alpha \in]0, \mathrm{S}[$, existence et unicité de l'écriture si et seulement si pour tout entier naturel n, $u_n = \sum_{k=n+1}^{+\infty} u_k$ (ce qui impose en particulier que S est fini et que u_n tend vers 0).

• Montrons que c'est nécessaire. D'après la première question les conditions (i) et (ii) doivent être vérifiées pour avoir l'existence. Supposons qu'il existe un entier n_0 pour lequel $u_{n_0} < \sum_{k=n_0+1}^{+\infty} u_k$ et considérons $\alpha \in \left]u_{n_0}, \sum_{k=n_0+1}^{+\infty} u_k\right[$. Notons que la suite $(u_k)_{k\geqslant n_0+1}$ est strictement décroissante et vérifie aussi les conditions (i) et (ii). Il existe donc une extraction φ telle que $\varphi(\mathbb{N}) \subset [\![n_0+1, +\infty[\![$ et $\alpha = \sum_{k=0}^{+\infty} u_{\varphi(k)}$ (puisque $\alpha \in \left]0, \sum_{k=n_0+1}^{+\infty} u_k\right[$). Mais on a également $\alpha - u_{n_0} \in \left]0, \sum_{k=n_0+1}^{+\infty} u_k\right[$ et il existe une extraction ψ telle que $\psi(\mathbb{N}) \subset [\![n_0+1, +\infty[\![$ et $\alpha - u_{n_0} = \sum_{k=0}^{+\infty} u_{\psi(k)}$. L'égalité $\alpha = \sum_{k=0}^{+\infty} u_{\varphi(k)} = u_{n_0} + \sum_{k=0}^{+\infty} u_{\psi(k)}$ montre que l'écriture de α n'est pas unique.

• Montrons que cette condition est suffisante. Si cette condition est réalisée, alors la série de terme général u_n converge et pour tout entier n, $u_n = \sum_{k=n+1}^{+\infty} u_k = u_{n+1} + \sum_{k=n+2}^{+\infty} u_k = 2u_{n+1}$. La suite $(u_n)_{n\geqslant 0}$ est une

suite géométrique de raison $\frac{1}{2}$. Pour tout entier n, $u_n = \frac{u_0}{2^n}$ et $S = 2u_0$. L'unicité demandée se ramène à l'unicité de la représentation d'un réel de $]0,1[$ sous la forme $\sum_{k=1}^{+\infty} \frac{a_k}{2^k}$, avec pour tout k de $\mathbb{N}^*$ $a_k = 0$ ou 1 et $\{k, a_k = 1\}$ infini. Il s'agit de l'unicité de la décomposition de tout réel de $]0,1[$ en base 2, en prenant, si α est de la forme $\frac{n}{2^k}$ $(k \in \mathbb{N}^*, 0 < n < 2^k)$, le développement impropre. $\triangleleft$

Les exercices suivant concernent des questions de rationalité. Le lecteur trouvera dans l'exercice 5.48 du tome 1 d'algèbre une preuve de la transcendance de e c'est-à-dire du fait que la suite $(e^n)_{n\geqslant 0}$ est $\mathbb{Q}$-libre. Ici, on montre élémentairement que la suite $(1, e, e^2)$ est $\mathbb{Q}$-libre, i.e. *que e n'est racine d'aucune équation du second degré à coefficients entiers.*

1.8. Sur le nombre e

1. Montrer que e est irrationnel.
2. Montrer que la famille $(1, e, e^2)$ est $\mathbb{Q}$-libre.
(**École normale supérieure**)

Solution.

1. Rappelons que e est défini par $e = \sum_{n=0}^{+\infty} \frac{1}{n!}$. On procède par l'absurde en supposant que $e = \frac{p}{q}$ avec $p, q \geqslant 1$ et $\operatorname{pgcd}(p,q) = 1$. On a alors

$$q!p = qq!e = \sum_{k=0}^{q} \frac{qq!}{k!} + \underbrace{q \sum_{k=q+1}^{+\infty} \frac{q!}{k!}}_{\varepsilon_q}.$$

La première somme est un entier naturel de même que $q!p$. On va voir que ε_q est un réel de $]0,1[$ ce qui fournira la contradiction souhaitée. On a $\varepsilon_q > 0$ et

$$\varepsilon_q < q \sum_{h=1}^{+\infty} \frac{1}{(q+1)^h} = \frac{q}{q+1} \frac{1}{1 - \frac{1}{q+1}} = 1.$$

2. Montrons que la famille $(1, e, e^2)$ est libre. Supposons qu'il y ait une relation de liaison $\alpha e + \beta e^2 + \gamma = 0$ avec α, β et γ rationnels non tous nuls. En divisant par e et en multipliant cette relation par un entier convenable, on peut donc trouver trois entiers relatifs a, b, c non tous nuls, tels que $ae + be^{-1} = c$. On peut même supposer que a et b sont non

nuls puisque e est irrationnel et, quitte à multiplier par -1, que $a \in \mathbb{N}^*$. On va approcher e et e^{-1} par les sommes partielles des séries qui les définissent. Cela se fait directement en utilisant la formule de Taylor-Lagrange en 0 pour la fonction $f(x) = ae^x + be^{-x}$. Pour tout $n \geqslant 1$, il existe $\alpha_n \in]0,1[$ tel que

$$c = ae + be^{-1} = f(1) = a\sum_{k=0}^{n} \frac{1}{k!} + b\sum_{k=0}^{n} \frac{(-1)^k}{k!} + \frac{ae^{\alpha_n} + (-1)^{n+1}be^{-\alpha_n}}{(n+1)!}.$$

En multipliant cette égalité par $n!$, on voit que le nombre

$$A_n = \frac{ae^{\alpha_n} + (-1)^{n+1}be^{-\alpha_n}}{n+1}$$

est un entier. Si b est positif (comme a) la suite extraite (A_{2n+1}) est formée d'entiers naturels non nuls et tend vers 0 ; c'est impossible. Si b est négatif on prend la suite (A_{2n}) et on obtient la même contradiction. ◁

Dans la plupart des démonstrations d'irrationalité ou de transcendance, on raisonne par l'absurde et l'argument fondamental consiste souvent à construire une suite d'entiers > 0 qui tend vers 0 (voir aussi l'exercice suivant). Expliquons pourquoi il a été indispensable de travailler avec e^{-1} au lieu de e^2. Si on utilise la fonction $f(x) = ae^x + be^{2x}$ et qu'on lui applique la formule de Taylor comme ci-dessus, le reste contient au numérateur un terme de la forme $ae^{\alpha_n} + b2^{n+1}e^{2\alpha_n}$. Et le 2^{n+1} rend ce reste divergent. La remarque qu'on vient de faire laisse penser que la méthode employée ici ne peut se généraliser pour établir la transcendance de e.

Le nombre π est également transcendant : c'est un théorème de Lindemann (1882). L'exercice qui suit établit son irrationalité, prouvée pour la première fois en 1761 par Lambert à l'aide du développement en fraction continue de la fonction tangente.

1.9. Irrationalité de π

À tout polynôme $f \in \mathbb{R}[X]$, on associe $F = \sum_{k \geqslant 0} (-1)^k f^{(2k)}$.

1. Calculer $\int_0^\pi f(t)\sin t\,dt$ en fonction de $F(0)$ et $F(\pi)$.

2. On suppose que $\pi = \frac{a}{b}$ avec a, b entiers > 0 et on pose $f_n(x) = \frac{x^n(a-bx)^n}{n!}$. Tracer le graphe de f_n sur $[0,\pi]$.

3. Prouver que $\int_0^\pi f_n(t)\sin t dt$ est entier et en déduire que π est irrationnel.

(École normale supérieure)

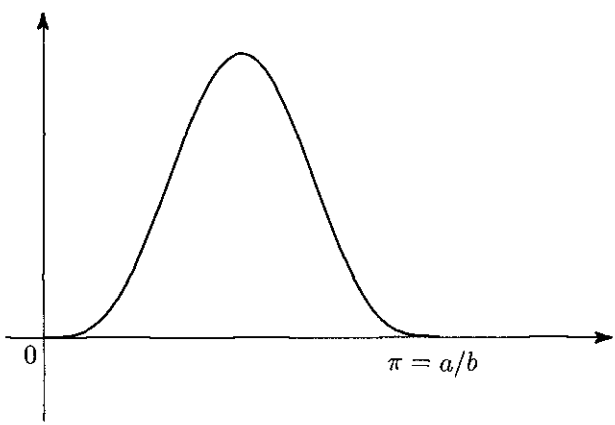
Solution.

1. Notons que la somme qui définit F est finie. Si on effectue une double intégration par parties dans l'intégrale, on obtient

$$\int_0^\pi f(t)\sin t\mathrm{d}t = [f(\pi)+f(0)] - \int_0^\pi f''(t)\sin t\mathrm{d}t.$$

En itérant le calcul on a donc $\int_0^\pi f(t)\sin t\mathrm{d}t = \mathrm{F}(\pi)+\mathrm{F}(0)$.

2. La fonction f_n est nulle en 0 et π et strictement positive sur l'intervalle $]0,\pi[$. Ses dérivées jusqu'à l'ordre $n-1$ s'annulent en 0 et en a/b.



3. Pour montrer que $\mathrm{I}_n = \int_0^\pi f_n(t)\sin t\mathrm{d}t$ est entier, il suffit, d'après la question 1, de montrer que f_n et toutes ses dérivées prennent des valeurs entières en 0 et π. Traitons le cas de 0. Comme 0 est une racine de f_n de multiplicité n, toutes les dérivées d'ordre $p<n$ sont nulles en 0. Soit $p \geqslant n$. On a par la formule de Leibniz,

$$f_n^{(p)}(x) = \frac{1}{n!}\sum_{k=0}^{p} \mathrm{C}_p^k (x^n)^{(k)}[(a-bx)^n]^{(p-k)}.$$

Quand on évalue en 0, il ne reste qu'un seul terme dans la somme, correspondant à $k=n$,

$$f_n^{(p)}(0) = \frac{1}{n!}\mathrm{C}_p^n n![(a-bx)^n]^{(p-n)}(0) = \mathrm{C}_p^n[(a-bx)^n]^{(p-n)}(0) \in \mathbb{Z}$$

car $(a-bx)^n$ est un polynôme à coefficients entiers. On montre de même que les dérivées de f_n en $\pi = \dfrac{a}{b}$ sont toutes entières.

Comme $f_n(t)\sin t$ est strictement positif pour $t \in]0,\pi[$, on peut affirmer que pour tout n, I_n est un entier naturel non nul. Or, comme $x(a-bx) \leqslant \dfrac{a^2}{4b} = \mathrm{M}$ pour tout $x \in [0,\pi]$, on a

$$1 \leqslant \mathrm{I}_n \leqslant \int_0^\pi \frac{\mathrm{M}^n}{n!}\mathrm{d}t = \frac{\pi \mathrm{M}^n}{n!} \xrightarrow[n\to+\infty]{} 0$$

ce qui est impossible. Donc $\pi \notin \mathbb{Q}$. ◁

L'exercice suivant est extrait d'une épreuve écrite du concours d'entrée à l'ENS de Saint-Cloud (1985), consacrée au troisième problème de Hilbert. Il s'agissait de savoir si lorsqu'on a deux polyèdres de même volume, par exemple un cube et un tétraèdre régulier, on peut découper le premier en un nombre fini de tétraèdres qui, réassemblés autrement, donnent le second (on dit alors que les deux polyèdres sont équidécomposables). La réponse est négative en général et notamment dans le cas du cube et du tétraèdre régulier. Pour le montrer, le mathématicien Max Dehn (élève de Hilbert) a associé à chaque polyèdre des quantités invariantes par les transformations décrites ci-dessus (appelées depuis invariants de Dehn). L'irrationalité du nombre $(\arccos 1/3)/\pi$ *permet de construire un tel invariant qui ne prend pas la même valeur pour le tétraèdre régulier et pour le cube. Ils ne sont donc pas équidécomposables*[1].

1.10. Irrationalité de $(\arccos 1/3)/\pi$

On se propose de montrer que $\alpha = \dfrac{\arccos 1/3}{\pi} \notin \mathbb{Q}$.

1. Calculer $e^{i\alpha\pi}$.

2. Montrer que $\alpha \in \mathbb{Q}$ si et seulement si il existe un entier naturel non nul n tel que $(1+2i\sqrt{2})^n = 3^n$.

3. Montrer que $(1+2i\sqrt{2})^n = a_n + ib_n\sqrt{2}$, où a_n et b_n sont des entiers vérifiant $a_n - b_n \not\equiv 0 \pmod 3$. Conclure.

(**École normale supérieure**)

1. Le lecteur intéressé pourra résoudre le problème mentionné ci-dessus ou consulter le chapitre 7 de AIGNER (M.) & ZIEGLER (G.M.), *Proofs from the Book*, Springer, 1998. Trad. fr. : *Raisonnements divins*, Springer, 2002.

Solution.

1. Par définition de α, on a $\cos \alpha\pi = \frac{1}{3}$. Comme $\alpha\pi$ appartient à l'intervalle $]0, \pi[$ son sinus est positif et vaut donc $\sqrt{1 - \frac{1}{9}} = \frac{2\sqrt{2}}{3}$. Il en résulte que $e^{i\alpha\pi} = \frac{1 + 2i\sqrt{2}}{3}$.

2. Le nombre complexe $e^{i\alpha\pi}$ est une racine de l'unité si et seulement si α est rationnel. Ainsi, α est rationnel si et seulement s'il existe un entier naturel non nul n tel que $e^{in\alpha\pi} = 1$, c'est-à-dire $(1 + 2i\sqrt{2})^n = 3^n$.

3. L'ensemble des complexes de la forme $a + ib\sqrt{2}$ avec $(a, b) \in \mathbb{Z}^2$ est clairement un sous-anneau de $\mathbb{C}$. Pour tout n, il existe donc des entiers a_n et b_n, qui sont d'ailleurs uniques, tels que $(1 + 2i\sqrt{2})^n = a_n + ib_n\sqrt{2}$. Comme

$$\begin{aligned}(1 + 2i\sqrt{2})^{n+1} &= (1 + 2i\sqrt{2})^n(1 + 2i\sqrt{2}) = (a_n + ib_n\sqrt{2})(1 + 2i\sqrt{2}) \\ &= (a_n - 4b_n) + i\sqrt{2}(b_n + 2a_n)\end{aligned}$$

on a les relations de récurrence suivantes : $a_{n+1} = a_n - 4b_n$ et $b_{n+1} = b_n + 2a_n$. On en déduit que $a_{n+1} - b_{n+1} = -a_n - 5b_n \equiv -(a_n - b_n) \pmod 3$. Comme $a_1 - b_1 \equiv -1 \pmod 3$, il en résulte que pour tout $n \geqslant 1$, $a_n - b_n \not\equiv 0 \pmod 3$. Il est donc impossible de trouver un entier $n \geqslant 1$ tel que $(1 + 2i\sqrt{2})^n = 3^n$, ce qui prouve que α est irrationnel. ◁

On peut montrer que le nombre $\frac{1}{\pi} \arccos \frac{1}{\sqrt{n}}$ *est irrationnel sauf pour* $n \in \{1, 2, 4\}$.

Liouville est le premier à produire des nombres transcendants, en 1844. La transcendance de e ne sera prouvée qu'en 1873 par Hermite et celle de π *en 1882 par Lindemann.*

1.11. Nombres de Liouville

1. Soit $P \in \mathbb{Z}[X]$, de degré $m \geqslant 1$, et x une racine réelle de P. Montrer qu'il existe $K > 0$ tel que, pour tout rationnel $\frac{a}{b}$ de $]x - 1, x + 1]$ vérifiant $P(\frac{a}{b}) \neq 0$, on ait $\left|x - \frac{a}{b}\right| \geqslant \frac{K}{|b|^m}$.

2. Soit $(u_n) \in [\![0, 9]\!]^{\mathbb{N}}$ une suite strictement positive à partir d'un certain rang. On pose $x = \sum_{n=0}^{+\infty} \frac{u_n}{10^{n!}}$ (x est un *nombre de Liouville*) et $s_n = \sum_{k=0}^{n} \frac{u_k}{10^{k!}}$. Montrer que $|x - s_n| \leqslant \frac{1}{10^{nn!}}$ pour tout n. En déduire que x est transcendant sur $\mathbb{Q}$.

(École normale supérieure)

▷ **Solution.**

1. On va utiliser l'inégalité des accroissements finis. La fonction P' est bornée sur $[x-1, x+1]$. Soit $M > 0$ un majorant de $|P'|$ sur l'intervalle $[x-1, x+1]$. On a

$$\left|P\left(\frac{a}{b}\right)\right| = \left|P(x) - P\left(\frac{a}{b}\right)\right| \leqslant M\left|x - \frac{a}{b}\right|.$$

La quantité $P\left(\frac{a}{b}\right)$ n'est pas nulle par hypothèse. On va la minorer. Si le polynôme P s'écrit $P = \sum_{k=0}^{m} a_k X^k$, on voit que $b^m P\left(\frac{a}{b}\right) = \sum_{k=0}^{m} a_k a^k b^{m-k}$ est un entier. Comme $\frac{a}{b}$ n'est pas racine de P il est non nul, donc plus grand que 1 en valeur absolue. On a donc $\left|b^m P\left(\frac{a}{b}\right)\right| \geqslant 1$, ce qui, avec l'inégalité précédente, conduit au résultat demandé en prenant $K = \frac{1}{M}$.

2. La série définissant x est clairement convergente et elle donne le développement décimal propre de x. On a, pour tout $n \in \mathbb{N}$,

$$|x - s_n| = x - s_n = \sum_{k=n+1}^{+\infty} \frac{u_k}{10^{k!}} \leqslant 9 \sum_{k=n+1}^{+\infty} \frac{1}{10^{k!}}.$$

Pour $k \geqslant n+1$, on a $k! \geqslant kn!$ et donc $\frac{1}{10^{k!}} \leqslant \frac{1}{10^{kn!}} = \left(\frac{1}{10^{n!}}\right)^k$. On en déduit que

$$|x - s_n| \leqslant 9 \sum_{k=n+1}^{+\infty} \left(\frac{1}{10^{n!}}\right)^k = 9\left(\frac{1}{10^{n!}}\right)^{n+1} \frac{1}{1 - \frac{1}{10^{n!}}}$$
$$\leqslant 9\left(\frac{1}{10^{n!}}\right)^n \frac{1}{10^{n!} - 1} \leqslant \frac{1}{10^{nn!}},$$

ce qui est le résultat demandé.

Montrons que x est transcendant. On raisonne par l'absurde et on suppose qu'il existe $P \in \mathbb{Z}[X]$, non nul, tel que $P(x) = 0$. On choisit $K > 0$ comme dans la question 1. La suite (s_n) est évidemment à valeurs dans $\mathbb{Q}$ et converge vers x. Le polynôme P n'étant pas nul, il existe un voisinage de x sur lequel P ne s'annule qu'en x. On en déduit qu'il existe un entier naturel n_0 tel que, pour $n \geqslant n_0$, s_n soit dans $[x-1, x+1]$ et $P(s_n) \neq 0$. La fraction s_n étant de dénominateur $10^{n!}$, on déduit de la première question que, pour $n \geqslant n_0$, on a $|x - s_n| \geqslant \frac{K}{(10^{n!})^m}$. On a donc, en combinant cela avec l'inégalité précédente, $\frac{K}{(10^{n!})^m} \leqslant \frac{1}{10^{nn!}}$ ou encore $(10^{n!})^{n-m} \leqslant \frac{1}{K}$, pour $n \geqslant n_0$. C'est impossible car le premier membre tend vers $+\infty$.

Conclusion. Tous les nombres de Liouville sont transcendants. ◁

La première question de l'exercice a montré qu'un irrationnel algébrique, c'est-à-dire qui est racine d'un polynôme non nul à coefficients entiers, s'approche assez mal par des nombres rationnels. C'est un théorème dû à Liouville que le lecteur trouvera aussi dans l'exercice 5.47 du tome 1 d'algèbre. Les nombres de Liouville sont construits pour être très bien approchés par des rationnels et c'est ce qui les rend transcendants.

1.12. Nombres de Pisot

1. Soit A une partie de $\mathbb{N}^{\mathbb{N}}$ telle qu'il existe une application $n_0 : \mathrm{A} \to \mathbb{N}$ vérifiant, pour tout $(u, v) \in \mathrm{A}^2$,

$$(\forall k \leqslant \max(n_0(u), n_0(v)),\quad u_k = v_k) \Longleftrightarrow u = v$$

Montrer que A est au plus dénombrable.

2. Considérons $\mathrm{S} = \{t > 1,\ \exists \alpha > 0,\ \lim\limits_{n\to+\infty} d(\alpha t^n, \mathbb{Z}) = 0\}$. Soit $t \in \mathrm{S}$ et a une suite d'entiers telle que $|\alpha t^n - a_n| \leqslant \dfrac{1}{2}$ pour tout $n \in \mathbb{N}$. Montrer que $a_{n+2} - \dfrac{a_{n+1}^2}{a_n}$ tend vers 0. Montrer que S est au plus dénombrable.

3. Soit $t > 1$ un réel, racine d'un polynôme unitaire $\mathrm{P} \in \mathbb{Z}[\mathrm{X}]$ tel que toutes les autres racines complexes de P soient de module strictement inférieur à 1 (un tel réel t est appelé un *nombre de Pisot*). Montrer que $t \in \mathrm{S}$.

(École normale supérieure)

Solution.

1. Pour $p \in \mathbb{N}$, l'ensemble A_p formé des suites u de A telles que $n_0(u) \leqslant p$ est au plus dénombrable, car l'application qui à une suite $u \in \mathrm{A}_p$ associe le $(p+1)$-uplet $(u_0, u_1, \ldots, u_p)$ est injective (et on sait que $\mathbb{N}^{p+1}$ est dénombrable). Comme $\mathrm{A} = \bigcup\limits_{p\geqslant 0} \mathrm{A}_p$ est une réunion dénombrable d'ensembles au plus dénombrables, A est au plus dénombrable.

2. Notons que comme $d(\alpha t^n, \mathbb{Z})$ tend vers 0, à partir d'un certain rang l'entier a_n est déterminé de manière unique. Par hypothèse, on a $a_n = \alpha t^n + o(1)$. Ainsi,

$$\frac{a_{n+1}^2}{a_n} = \frac{\alpha^2 t^{2n+2} + o(t^n)}{\alpha t^n + o(1)} = \frac{\alpha t^{n+2} + o(1)}{1 + o(t^{-n})} = \alpha t^{n+2} + o(1)$$

Il en résulte donc que $a_{n+2} - \dfrac{a_{n+1}^2}{a_n} = o(1)$.

Soit ψ une application qui à chaque $t \in \mathrm{S}$ associe une suite (a_n) comme-ci-dessus. On a $t = \lim_{n\to+\infty} \frac{a_{n+1}}{a_n}$ de sorte que ψ est injective. Montrons alors que l'ensemble des suites obtenues est au plus dénombrable. Pour cela, on va montrer que $\psi(\mathrm{S})$ vérifie les hypothèses de la question 1. Soit $t \in \mathrm{S}$ et $(a_n) = \psi(t)$. Il existe un rang $n_0(a)$ tel que pour $n \geqslant n_0(a)$, $\left|a_{n+2} - \frac{a_{n+1}^2}{a_n}\right| < \frac{1}{2}$. Considérons $(b_n) = \psi(t')$ une seconde suite et supposons que $a_k = b_k$ pour $k \leqslant \max(n_0(a), n_0(b)) = \mathrm{N}$. Alors, par l'inégalité triangulaire, puisque $\frac{a_{\mathrm{N}}^2}{a_{\mathrm{N}-1}} = \frac{b_{\mathrm{N}}^2}{b_{\mathrm{N}-1}}$, on obtient

$$|a_{\mathrm{N}+1} - b_{\mathrm{N}+1}| \leqslant \left|a_{\mathrm{N}+1} - \frac{a_{\mathrm{N}}^2}{a_{\mathrm{N}-1}}\right| + \left|b_{\mathrm{N}+1} - \frac{b_{\mathrm{N}}^2}{b_{\mathrm{N}-1}}\right| < \frac{1}{2} + \frac{1}{2} = 1.$$

Comme, $a_{\mathrm{N}+1}$ et $b_{\mathrm{N}+1}$ sont entiers on a $a_{\mathrm{N}+1} = b_{\mathrm{N}+1}$. Une récurrence facile montre alors que $a = b$. D'après la question 1, $\psi(\mathrm{S})$ est au plus dénombrable, donc S aussi.

3. Notons $z_1, \ldots, z_p$ les racines de P autres que t. Montrons que pour tout n, la somme de Newton $\mathrm{S}_n = t^n + z_1^n + \cdots + z_p^n$ est dans $\mathbb{Z}$. En effet, toutes les fonctions symétriques élémentaires des racines de P sont dans $\mathbb{Z}$ car P est unitaire et à coefficients dans $\mathbb{Z}$. Comme S_n est un polynôme symétrique en les racines de P à coefficients entiers, il peut s'écrire comme un polynôme à coefficients entiers en les fonctions symétriques élémentaires de ces racines. Il en résulte que $\mathrm{S}_n \in \mathbb{Z}$ pour tout n. Notons que cela résulte aussi des formules de Newton (cf. exercice 5.26 du tome 1 d'algèbre) et d'une récurrence sur n. Comme $|z_k| < 1$ pour tout k, les suites $(z_k^n)_{n\geqslant 0}$ tendent vers 0 et $d(t^n, \mathbb{Z})$ tend donc vers 0. Par suite, $t \in \mathrm{S}$. ◁

Le nombre d'or $\Phi = \frac{1+\sqrt{5}}{2}$ *est par exemple un nombre de Pisot. Il résulte de l'exercice que l'ensemble des nombres de Pisot est dénombrable. En 1944, Raphaël Salem a démontré qu'il est fermé. Le lecteur intéressé par les nombres de Pisot se reportera au problème de six heures posé au concours de la rue d'Ulm en 1989. On y démontre notamment le théorème de Pisot qui affirme qu'un réel* $t > 1$ *est un nombre de Pisot si et seulement si il existe* $\lambda > 0$ *tel que la série* $\sum d(\lambda t^n, \mathbb{Z})^2$ *converge.*

Le prochain exercice, très classique, rappelle la structure des sous-groupes additifs de $\mathbb{R}$.

1.13. Sous-groupes additifs de $\mathbb{R}$

1. Soit G un sous-groupe de $(\mathbb{R}, +)$ non réduit à $\{0\}$. Montrer que G est soit dense dans $\mathbb{R}$, soit de la forme $a\mathbb{Z}$ avec $a > 0$.

2. Soit α et β deux réels non nuls. Discuter la nature du sous-groupe additif qu'ils engendrent.

3. Soit $\beta \notin \mathbb{Q}$. Montrer que $\mathbb{N}\beta + \mathbb{Z}$ est dense dans $\mathbb{R}$.

4. Soit $\theta \notin 2\pi\mathbb{Q}$. Montrer que $\{e^{in\theta},\ n \in \mathbb{N}\}$ est dense dans le cercle unité S^1 de $\mathbb{C}$. En déduire l'ensemble des valeurs d'adhérence de la suite $(\sin n)_{n\geqslant 0}$.

(École polytechnique)

Solution.

1. Si G est monogène égal à $a\mathbb{Z}$ (avec $a > 0$), a est le plus petit élément strictement positif de G. Si G est dense dans $\mathbb{R}$, $G \cap \mathbb{R}_+^*$ n'a pas de plus petit élément mais une borne inférieure nulle. Il est donc naturel d'introduire $G_+ = G \cap \mathbb{R}_+^*$ et $a = \inf G_+$. Le réel $a \geqslant 0$ est bien défini car G_+ est non vide et minoré. En effet, il existe $x \in G$ non nul et donc x ou $-x$ est dans G_+, qui est minoré par 0. Les remarques précédentes invitent alors à distinguer deux cas.

• Supposons $a > 0$. On va montrer que $a \in G$ puis que $G = a\mathbb{Z}$. Raisonnons par l'absurde et supposons que a ne soit pas dans G.

Comme $a > 0$, on a $2a > a$ et il existe $x \in G_+$, tel que $x < 2a$. On a donc $a < x < 2a$, puisque $a \notin G$. Il existe alors $y \in G_+$ tel que $y < x$. On a $a < y < x < 2a$ et $x - y$ en tant que différence d'éléments de G est dans G, et même dans G_+. Or $x - y < a$, ce qui contredit la définition de a. Donc $a \in G$.

Par conséquent $a\mathbb{Z}$, le groupe engendré par a, est inclus dans G. Réciproquement, soit $x \in G$ et $k = E\left(\frac{x}{a}\right) \in \mathbb{Z}$. Comme G est un groupe le réel $x - ka$ est dans G, et comme $k \leqslant \frac{x}{a} < k + 1$, il vérifie $0 \leqslant x - ka < a = \min G_+$. Nécessairement, $x - ka = 0$ et $x = ka \in a\mathbb{Z}$. Il en résulte que $G = a\mathbb{Z}$.

• Supposons $a = 0$. On va montrer que G est dense dans $\mathbb{R}$ c'est-à-dire que G rencontre tout intervalle ouvert. Soit $I =]\alpha, \beta[$ un tel intervalle. Comme $a = 0$, il existe un élément $g \in G$ tel que $0 < g < \beta - \alpha$. Le sous-groupe $g\mathbb{Z}$ engendré par g est inclus dans G et rencontre nécessairement I (sinon il existerait un entier $k \in \mathbb{Z}$ tel que $I \subset]kg, (k+1)g[$ ce qui contredirait l'inégalité $g < \beta - \alpha$). Cela prouve la densité de G dans $\mathbb{R}$.

Conclusion. Un sous-groupe additif de $\mathbb{R}$ est soit monogène (et donc discret), soit dense dans $\mathbb{R}$.

2. Il s'agit d'étudier le sous-groupe $G = \alpha\mathbb{Z} + \beta\mathbb{Z} \neq \{0\}$.

Supposons qu'il existe $a > 0$ tel que $G = a\mathbb{Z}$. Comme α et β sont

dans G, on peut trouver $(k,l) \in \mathbb{Z}^2$ tel que $\alpha = ka$ et $\beta = la$, et en faisant le rapport $\dfrac{\alpha}{\beta} = \dfrac{k}{l} \in \mathbb{Q}$.

Réciproquement, supposons $\dfrac{\alpha}{\beta}$ rationnel. Écrivons $\dfrac{\alpha}{\beta} = \dfrac{k}{l}$, avec k et l premiers entre eux. On a

$$\alpha\mathbb{Z} + \beta\mathbb{Z} = \beta\left(\frac{k}{l}\mathbb{Z} + \mathbb{Z}\right) = \frac{\beta}{l}(k\mathbb{Z} + l\mathbb{Z}) = \frac{\beta}{l}\mathbb{Z},$$

car k et l sont premiers entre eux.

Conclusion. Si $\dfrac{\alpha}{\beta} \in \mathbb{Q}$, G est monogène et sinon G est dense dans $\mathbb{R}$.

3. Le sous-groupe additif $G = \mathbb{Z} + \beta\mathbb{Z}$ de $\mathbb{R}$ est dense d'après la question précédente. On va montrer que l'ensemble $\mathbb{N}\beta + \mathbb{Z}$ reste encore dense. Soit $a < b$ deux réels. On peut trouver un élément $x = v\beta + u \in G$ tel que $0 < x < b - a$.

• Supposons d'abord que v est un entier naturel, c'est-à-dire que $x \in \mathbb{N}\beta + \mathbb{Z}$. On choisit un entier $n_0 < a$. Les éléments de la suite $(kx + n_0)_{k\geqslant 0}$ sont tous dans $\mathbb{N}\beta + \mathbb{Z}$ et il est clair que l'un d'eux au moins appartient à l'intervalle $]a, b[$ (c'est le même argument que dans la question 1).

• Supposons que $v < 0$. Alors $-x \in \mathbb{N}\beta + \mathbb{Z}$ et $-(b-a) < -x < 0$. On choisit $n_0 \in \mathbb{Z}$ avec $n_0 > b$ et il existe au moins un des éléments de la suite $(n_0 - kx)_{k\geqslant 0}$ qui appartient à $]a, b[$.

4. Posons $X = \{e^{in\theta},\ n \in \mathbb{N}\}$. Il s'agit de l'image par l'application $f : x \longmapsto e^{2i\pi x}$ de l'ensemble $\mathbb{Z} + \dfrac{\theta}{2\pi}\mathbb{N}$. Comme f est continue et que cet ensemble est dense dans $\mathbb{R}$ d'après la question précédente, son image est dense dans $f(\mathbb{R}) = S^1$.

Dans le cas $\theta = 1$, on a bien $\dfrac{1}{\pi} \notin \mathbb{Q}$ et l'ensemble $\{e^{in},\ n \in \mathbb{N}\}$ est dense dans S^1. Comme l'application qui à un nombre complexe associe sa partie imaginaire est continue, l'ensemble $\{\sin n,\ n \in \mathbb{N}\}$ est dense dans le segment $[-1,1]$. Pour tout $a \in [-1,1]$, tout $\varepsilon > 0$ et tout $N \in \mathbb{N}$, on est alors assuré de trouver un entier $n \geqslant N$ tel que $|\sin n - a| \leqslant \varepsilon$. Cela signifie que tout réel de $[-1,1]$ est une valeur d'adhérence de la suite $(\sin n)_{n\geqslant 0}$. L'autre inclusion étant évidente, on conclut que l'ensemble des valeurs d'adhérence de la suite est le segment $[-1,1]$. ◁

On peut déduire de la question 1 le fait qu'un sous-groupe G *de* S^1 *est soit fini (égal dans ce cas au groupe des racines n-ièmes de l'unité, où* $n = |G|$*), soit dense dans* S^1.

Les exercices qui suivent sont consacrés à quelques inégalités. Le lecteur pourra aussi se reporter au chapitre 4 où il trouvera des exercices portant sur des inégalités de convexité.

1.14. Une inégalité

Montrer que parmi treize réels distincts on peut toujours en choisir deux, disons x et y, tels que $0 < \dfrac{x-y}{1+xy} < 2-\sqrt{3}$.

(École polytechnique)

Solution.

Notons $x_1 < x_2 < \cdots < x_{13}$ les réels en question. L'expression $\dfrac{x-y}{1+xy}$ rappelle le développement de $\tan(\theta - \theta')$. Considérons les treize réels $\theta_i = \arctan x_i$ pour $1 \leqslant i \leqslant 13$. Ils sont dans $\mathrm{I} = \left]-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right[$ et rangés dans l'ordre croissant. Comme la longueur de I est égale à π, il existe $k \in [\![1, 12]\!]$ tel que $\theta_{k+1} - \theta_k < \dfrac{\pi}{12}$. Comme la fonction tangente est strictement croissante sur $\left[0, \dfrac{\pi}{2}\right[$, on a $0 < \tan(\theta_{k+1} - \theta_k) < \tan\dfrac{\pi}{12}$. Or, nous savons que

$$\tan(\theta_{k+1} - \theta_k) = \frac{\tan\theta_{k+1} - \tan\theta_k}{1 + \tan\theta_k \tan\theta_{k+1}} = \frac{x_{k+1} - x_k}{1 + x_k x_{k+1}}.$$

Le réel $\alpha = \tan\dfrac{\pi}{12}$ vérifie l'équation

$$\frac{2\alpha}{1-\alpha^2} = \tan\frac{\pi}{6} = \frac{1}{\sqrt{3}}.$$

En résolvant l'équation du second degré qui en découle, on a $\alpha = -\sqrt{3} \pm 2$. Comme $\alpha > 0$, on trouve $\alpha = 2 - \sqrt{3}$. Finalement, on obtient

$$0 < \frac{x_{k+1} - x_k}{1 + x_k x_{k+1}} < 2 - \sqrt{3}.$$

D'où le résultat. ◁

1.15. Inégalité du réordonnement

Soit $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_n$ et $y_1 \geqslant y_2 \geqslant \cdots \geqslant y_n$ des réels. Soit $(z_1, \ldots, z_n)$ une permutation de $(y_1, \ldots, y_n)$. Montrer que

$$\sum_{i=1}^{n} (x_i - y_i)^2 \leqslant \sum_{i=1}^{n} (x_i - z_i)^2.$$

(École polytechnique)

▷ **Solution.**

On commence par développer les carrés. Dans l'inégalité que l'on souhaite prouver, on peut alors simplifier de part et d'autre $\sum_{i=1}^{n} x_i^2$ et $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} z_i^2$, de sorte qu'il est équivalent de montrer l'inégalité

$$\sum_{i=1}^{n} x_i z_i \leqslant \sum_{i=1}^{n} x_i y_i.$$

Supposons dans un premier temps que les inégalités entre les x_i sont toutes strictes : $x_1 > x_2 > \cdots > x_n$. Si les z_i ne sont pas rangés en ordre décroissant, c'est-à-dire si on peut trouver $i < j$ tel que $z_i < z_j$, on augmente (strictement) la somme $\sum_{i=1}^{n} x_i z_i$ en échangeant z_i et z_j. En effet,

$$x_i z_i + x_j z_j < x_i z_j + x_j z_i,$$

car cette inégalité se réécrit $x_i(z_i - z_j) + x_j(z_j - z_i) < 0$ ce qui équivaut encore à $(x_i - x_j)(z_i - z_j) < 0$. Parmi les $n!$ sommes obtenues en permutant $y_1, \ldots, y_n$, la valeur maximale est donc obtenue pour une permutation $(z_1, \ldots, z_n)$ telle que $z_1 \geqslant z_2 \geqslant \cdots \geqslant z_n$. Auquel cas, on a $z_1 = y_1$, $z_2 = y_2, \ldots, z_n = y_n$. Cela démontre le résultat dans ce cas.

Pour le cas général nous proposons deux rédactions.

• On peut se ramener au cas où les inégalités entre les x_i sont strictes en les perturbant un petit peu. On pose, pour $\varepsilon > 0$, $x'_n = x_n$, $x'_{n-1} = x_{n-1} + \varepsilon$, $x'_{n-2} = x_{n-2} + 2\varepsilon, \ldots, x'_1 = x_1 + (n-1)\varepsilon$. On a alors $x'_n < x'_{n-1} < \cdots < x'_1$ et d'après ce qui précède,

$$\sum_{i=1}^{n} x'_i z_i \leqslant \sum_{i=1}^{n} x'_i y_i$$

pour tout n-uplet $(z_1, \ldots, z_n)$ obtenu en permutant $(y_1, \ldots, y_n)$. Cela étant valable pour tout $\varepsilon > 0$, il suffit de faire tendre ε vers 0^+ pour obtenir le résultat souhaité.

• On peut aussi raisonner directement en regroupant les indices $1, \ldots, n$ en paquets $(1, \ldots, i_1)$, $(i_1 + 1, \ldots, i_2)$, $\ldots$, $(i_{p-1} + 1, \ldots, i_p)$ (avec $i_p = n$) de façon à avoir

$$x_1 = \cdots = x_{i_1} < x_{i_1+1} = \cdots = x_{i_2} < \cdots < x_{i_{p-1}+1} = \cdots = x_{i_p}.$$

Sans changer la somme $\sum_{i=1}^{n} x_i z_i$, on peut réarranger les z_i à l'intérieur de chaque paquet d'indices, en les classant par ordre décroissant de façon à avoir $z_{i_{k-1}+1} \geqslant \cdots \geqslant z_{i_k}$. Si après ce réarrangement les z_i ne sont pas

globalement rangés par ordre décroissant, alors si i et j sont tels que $i < j$ et $z_i < z_j$, les indices i et j n'appartiennent pas au même paquet et on a aussi $x_i > x_j$. On peut alors conclure comme dans le premier cas.

1.16. Inégalité de Tchebychev

1. Soient $(x_1, \ldots, x_n)$ et $(y_1, \ldots, y_n)$ deux suites monotones de réels. Comparer

$$\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \times \left(\frac{1}{n}\sum_{k=1}^{n} y_k\right) \quad \text{à} \quad \left(\frac{1}{n}\sum_{k=1}^{n} x_k y_k\right).$$

2. Soient f et g deux fonctions monotones et continues par morceaux sur $[a, b]$. Comparer $\int_a^b f \times \int_a^b g$ avec $(b-a)\int_a^b fg$.

(École polytechnique)

Solution.

1. Il est naturel de faire la différence des deux quantités proposées. En multipliant par n^2 on doit chercher le signe de

$$\mathrm{D} = n\sum_{k=1}^{n} x_k y_k - \sum_{1 \leqslant i,j \leqslant n} x_i y_j.$$

En sommant d'abord selon les x_i il vient

$$\begin{aligned} \mathrm{D} &= \sum_{i=1}^{n} x_i \Big((y_i - y_1) + (y_i - y_2) + \cdots + (y_i - y_n) \Big) \\ &= \sum_{j<i} (y_i - y_j)(x_i - x_j). \end{aligned}$$

Par hypothèses les différences $x_i - x_j$ (resp. $y_i - y_j$) pour $i < j$ ont toutes le même signe. On en déduit que si les deux suites ont la même monotonie alors $\mathrm{D} \geqslant 0$ et

$$\left(\frac{1}{n}\sum_{k=1}^{n} x_k\right) \times \left(\frac{1}{n}\sum_{k=1}^{n} y_k\right) \leqslant \left(\frac{1}{n}\sum_{k=1}^{n} x_k y_k\right).$$

L'inégalité est inversée si les deux suites ont des monotonies opposées.

2. La première question suggère d'utiliser des sommes de Riemann. Notons $(x_{k,n})_{0 \leqslant k \leqslant n}$ les points de la subdivision régulière du segment $[a, b]$ en n intervalles. Supposons par exemple que f et g ont la même monotonie. On a alors pour tout n,

$$\left(\frac{b-a}{n}\sum_{k=1}^{n} f(x_{k,n})\right)\times\left(\frac{b-a}{n}\sum_{k=1}^{n} g(x_{k,n})\right) \leqslant \frac{(b-a)^2}{n}\sum_{k=1}^{n} f(x_{k,n})g(x_{k,n})$$

d'après la première question. En passant à la limite on obtient donc

$$\int_a^b f(t)\mathrm{d}t \times \int_a^b g(t)\mathrm{d}t \leqslant (b-a)\int_a^b f(t)g(t)\mathrm{d}t$$

et l'inégalité est inversée lorsque f et g ont des monotonies opposées. ◁

Notons qu'il est très facile de déduire l'inégalité de Tchebychev à partir de l'inégalité du réordonnement de l'exercice précédent. Supposons par exemple les suites de même monotonie. On a alors les inégalités suivantes :

$$\begin{aligned}
x_1y_1 + x_2y_2 + \cdots + x_ny_n &\leqslant x_1y_1 + x_2y_2 + \cdots + x_ny_n\\
x_1y_2 + x_2y_3 + \cdots + x_ny_1 &\leqslant x_1y_1 + x_2y_2 + \cdots + x_ny_n\\
&\vdots\\
x_1y_n + x_2y_1 + \cdots + x_ny_{n-1} &\leqslant x_1y_1 + x_2y_2 + \cdots + x_ny_n
\end{aligned}$$

En sommant il vient $(x_1+\cdots+x_n)(y_1+\cdots+y_n) \leqslant n(x_1y_1+\cdots+x_ny_n)$ *ce qui est le résultat voulu.*

Les trois exercices qui suivent sont consacrés à des inégalités assez fines sur des nombres complexes.

1.17. Inégalité dans $\mathbb{C}$ (1)

Soit $n \in \mathbb{N}^*$, et $(z_1, \ldots, z_n) \in \mathbb{C}^n$. Montrer qu'il existe une partie $\mathrm{I} \subset [\![1, n]\!]$ telle que

$$\left|\sum_{k\in \mathrm{I}} z_k\right| \geqslant \frac{1}{\pi}\sum_{k=1}^{n} |z_k|.$$

On introduira, pour $\theta \in \mathbb{R}$,

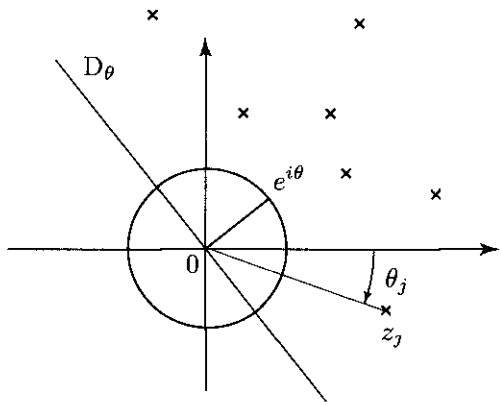
$$\mathrm{S}_\theta = \left\{ j \in [\![1, n]\!],\ \theta - \frac{\pi}{2} \leqslant \arg z_j \leqslant \theta + \frac{\pi}{2} \right\},$$

ainsi que la fonction $f(\theta) = \left|\sum_{k\in \mathrm{S}_\theta} z_k\right|$. On cherchera une minoration de f, que l'on intégrera sur $[0, 2\pi]$.

(École polytechnique)

Solution.

On peut évidemment supposer que les z_i sont tous non nuls. Notons D_θ la droite du plan complexe passant par 0 et de vecteur normal $e^{i\theta}$. L'ensemble S_θ est alors l'ensemble des indices $j \in [\![1, n]\!]$ tels que z_j soit dans le demi-plan fermé délimité par D_θ et contenant $e^{i\theta}$. Posons $z_k = r_k e^{i\theta_k}$ avec $r_k = |z_k| > 0$ et $\theta_k \in [0, 2\pi[$.



On a alors

$$\Big|\sum_{j\in S_\theta} z_j\Big| = \Big|\sum_{j\in S_\theta} z_j e^{-i\theta}\Big| \geqslant \mathrm{Re}\Big(\sum_{j\in S_\theta} z_j e^{-i\theta}\Big) = \sum_{j\in S_\theta} r_j \cos(\theta_j - \theta).$$

La fonction f est une fonction en escalier sur $[0, 2\pi]$ et la fonction $\theta \longmapsto \sum_{j\in S_\theta} r_j \cos(\theta_j - \theta)$ est continue par morceaux. Il vient

$$\int_0^{2\pi} f(\theta)\mathrm{d}\theta \geqslant \int_0^{2\pi} \sum_{j\in S_\theta} r_j \cos(\theta_j - \theta)\mathrm{d}\theta = \sum_{j=1}^{n} r_j \int_0^{2\pi} \cos(\theta - \theta_j)\chi_j(\theta)\mathrm{d}\theta$$

où la fonction χ_j est définie par $\chi_j(\theta) = 1$ si $j \in S_\theta$ et $\chi_j(\theta) = 0$ sinon. Or, si j est fixé, j est dans S_θ si et seulement si θ appartient à $\left[\theta_j - \frac{\pi}{2}, \theta_j + \frac{\pi}{2}\right]$ (modulo 2π). Plus précisément, $\chi_j^{-1}(1) = \left[\theta_j - \frac{\pi}{2}, \theta_j + \frac{\pi}{2}\right]$ si $\theta_j \in \left[\frac{\pi}{2}, \frac{3\pi}{2}\right]$, $\chi_j^{-1}(1) = \left[0, \theta_j + \frac{\pi}{2}\right] \cup \left[\theta_j + \frac{3\pi}{2}, 2\pi\right]$ si $\theta_j \in \left[0, \frac{\pi}{2}\right]$ et $\chi_j^{-1}(1) = \left[\theta_j - \frac{\pi}{2}, 2\pi\right] \cup \left[0, \theta_j - \frac{3\pi}{2}\right]$ si $\theta_j \in \left[\frac{3\pi}{2}, 2\pi\right]$. Par 2π-périodicité de la fonction cosinus, on a dans chacun des trois cas,

$$\int_0^{2\pi} \cos(\theta - \theta_j)\chi_j(\theta)\mathrm{d}\theta = \int_{\theta_j - \frac{\pi}{2}}^{\theta_j + \frac{\pi}{2}} \cos(\theta - \theta_j)\mathrm{d}\theta = 2.$$

Il en résulte que

$$\int_0^{2\pi} f(\theta)\mathrm{d}\theta \geqslant 2\sum_{j=1}^{n} r_j = 2\sum_{j=1}^{n} |z_j|.$$

Il existe donc nécessairement $\theta_0 \in [0, 2\pi]$ tel que $2\pi f(\theta_0) \geqslant 2 \sum_{j=1}^{n} |z_j|$. La partie $I = S_{\theta_0}$ répond alors à la question posée. $\lhd$

1.18. Inégalité dans $\mathbb{C}$ (2)

Soit $n \in \mathbb{N}^*$, et $z_1, \ldots, z_n$ des nombres complexes non nuls, dont on notera, pour $1 \leqslant k \leqslant n$, ρ_k le module et θ_k un argument. On désigne par I_n l'ensemble des applications de $[\![1, n]\!]$ dans $\{-1, 1\}$.

1. Montrer que

$$\max_{\varepsilon \in I_n} \left| \sum_{k=1}^{n} \varepsilon(k) z_k \right| = \sup_{\theta \in \mathbb{R}} \sum_{k=1}^{n} \rho_k |\cos(\theta_k - \theta)|.$$

2. Montrer que

$$\sum_{k=1}^{n} |z_k| \leqslant \frac{\pi}{2} \max_{\varepsilon \in I_n} \left| \sum_{k=1}^{n} \varepsilon(k) z_k \right|.$$

3. Prouver que $\dfrac{\pi}{2}$ est la plus petite constante strictement positive telle que l'inégalité ci-dessus ait lieu pour tout n et tout n-uplet $(z_1, \ldots, z_n)$ de nombres complexes non nuls.

(École normale supérieure)

$\rhd$ **Solution.**

1. Soit $\theta \in \mathbb{R}$. Notons D_θ la droite du plan complexe passant par 0 et de vecteur normal $e^{i\theta}$. On choisit les signes $\varepsilon(k)$ de sorte que $\varepsilon(k) z_k$ appartienne au demi-plan fermé délimité par D_θ et contenant $e^{i\theta}$. Dans ce cas, $\varepsilon(k)$ est exactement le signe de $\cos(\theta - \theta_k)$ (lorsque cette quantité est non nulle *i.e.* si $z_k \notin D_\theta$). Il en résulte que

$$\begin{aligned} \sum_{k=1}^{n} \rho_k |\cos(\theta_k - \theta)| &= \sum_{k=1}^{n} \varepsilon(k) \rho_k \cos(\theta_k - \theta) = \operatorname{Re}\left(\sum_{k=1}^{n} \varepsilon(k) z_k e^{-i\theta} \right) \\ &\leqslant \left| e^{-i\theta} \sum_{k=1}^{n} \varepsilon(k) z_k \right| = \left| \sum_{k=1}^{n} \varepsilon(k) z_k \right|. \end{aligned}$$

Ce qui précède étant valable pour tout $\theta \in \mathbb{R}$, on a

$$\sup_{\theta \in \mathbb{R}} \sum_{k=1}^{n} \rho_k |\cos(\theta_k - \theta)| \leqslant \max_{\varepsilon \in I_n} \left| \sum_{k=1}^{n} \varepsilon(k) z_k \right|.$$

Inversement, soit $\varepsilon \in I_n$. On écrit $\sum_{k=1}^{n} \varepsilon(k) z_k = r e^{i\theta}$ où $r \geqslant 0$ et $\theta \in \mathbb{R}$. On a alors

$$\left|\sum_{k=1}^{n}\varepsilon(k)z_k\right| = r = \sum_{k=1}^{n}\varepsilon(k)z_k e^{-i\theta} = \mathrm{Re}\left(\sum_{k=1}^{n}\varepsilon(k)z_k e^{-i\theta}\right)$$
$$= \sum_{k=1}^{n}\varepsilon(k)\rho_k\cos(\theta_k-\theta) \leqslant \sum_{k=1}^{n}\rho_k|\cos(\theta_k-\theta)|,$$

ce qui prouve que

$$\max_{\varepsilon\in \mathrm{I}_n}\left|\sum_{k=1}^{n}\varepsilon(k)z_k\right| \leqslant \sup_{\theta\in\mathbb{R}}\sum_{k=1}^{n}\rho_k|\cos(\theta_k-\theta)|$$

et l'égalité demandée en résulte.

2. On a, d'après la question 1,

$$\int_0^{\pi}\sum_{k=1}^{n}\rho_k|\cos(\theta_k-\theta)|\mathrm{d}\theta \leqslant \pi\max_{\varepsilon\in \mathrm{I}_n}\left|\sum_{k=1}^{n}\varepsilon(k)z_k\right|.$$

On va calculer l'intégrale de gauche. Comme $u \longmapsto |\cos u|$ est une fonction π-périodique, on a pour tout k,

$$\int_0^{\pi}|\cos(\theta-\theta_k)|\mathrm{d}\theta = \int_{-\theta_k}^{\pi-\theta_k}|\cos u|\mathrm{d}u = \int_0^{\pi}|\cos u|\mathrm{d}u = 2.$$

On obtient donc

$$2\sum_{k=1}^{n}|z_k| = 2\sum_{k=1}^{n}\rho_k \leqslant \pi\max_{\varepsilon\in \mathrm{I}_n}\left|\sum_{k=1}^{n}\varepsilon(k)z_k\right|,$$

ce qui est l'inégalité demandée.

3. On va prendre pour $z_1,\dots,z_n$ les racines $2n$-ièmes de l'unité qui se trouvent dans le demi-plan défini par $\mathrm{Im}\, z \geqslant 0$ (hormis -1), c'est-à-dire les complexes $z_k = e^{\frac{i(k-1)\pi}{2n}}$ pour $k \in [\![1,n]\!]$. On a dans ce cas $\sum_{k=1}^{n}|z_k| = n$. Pour estimer le second membre de l'inégalité obtenue en 2, on va utiliser la question 1. Notons f la fonction définie par

$$f(\theta) = \sum_{k=0}^{n-1}\left|\cos\Big(\theta-\frac{k\pi}{n}\Big)\right|.$$

On voit facilement que f est $\dfrac{\pi}{n}$-périodique ce qui permet de se placer sur $[0,\pi/n]$ pour calculer son maximum. Pour simplifier les calculs, on va supposer en plus que $n=2p$ est pair. On a alors

$$f(\theta) = \sum_{k=0}^{p-1}\cos\Big(\theta-\frac{k\pi}{n}\Big) - \sum_{k=p}^{n-1}\cos\Big(\theta-\frac{k\pi}{n}\Big)$$
$$= \sum_{k=0}^{p-1}\cos\Big(\theta-\frac{k\pi}{n}\Big) + \cos\Big(\theta+\frac{\pi}{2}-\frac{k\pi}{n}\Big)$$

En utilisant la somme des termes d'une suite géométrique, on obtient

$$\sum_{k=0}^{p-1} \cos\Big(\theta - \frac{k\pi}{n}\Big) = \mathrm{Re}\Big(e^{i\theta}\frac{1+i}{1-e^{-i\pi/n}}\Big) = \frac{\sqrt{2}\cos(\theta - \frac{\pi}{4} + \frac{\pi}{4p})}{2\sin\frac{\pi}{4p}}.$$

Il en résulte que

$$f(\theta) = \frac{\sqrt{2}}{2\sin\frac{\pi}{4p}}\Big(\cos\Big(\theta - \frac{\pi}{4} + \frac{\pi}{4p}\Big) + \cos\Big(\theta + \frac{\pi}{4} + \frac{\pi}{4p}\Big)\Big) = \frac{\cos(\theta + \frac{\pi}{4p})}{\sin\frac{\pi}{4p}}.$$

Sur $\big[0, \frac{\pi}{n}\big]$ le maximum de f est donc $\operatorname{cotan}\frac{\pi}{4p} = \operatorname{cotan}\frac{\pi}{2n}$. Si C est une constante qui convient pour l'inégalité de la question 2, on doit donc avoir pour tout n pair,

$$n \leqslant \mathrm{C}\operatorname{cotan}\frac{\pi}{2n}$$

soit $n\tan\frac{\pi}{2n} \leqslant \mathrm{C}$. En passant à la limite, on obtient $\frac{\pi}{2} \leqslant \mathrm{C}$. ◁

Dans l'exercice 3.48 le lecteur trouvera un théorème de Riemann qui affirme que si $\sum u_n$ est une série réelle semi-convergente, on peut trouver pour tout réel λ une permutation σ de $\mathbb{N}$ telle que $\sum_{n=0}^{+\infty} u_{\sigma(n)} = \lambda$. On peut se poser la même question dans $\mathbb{C}$. Un théorème de Steinitz affirme que si une série complexe $\sum z_n$ converge, l'ensemble des sommes des séries convergentes de la forme $\sum z_{\sigma(n)}$ (où σ est une permutation de $\mathbb{N}$) est un sous-espace affine réel de $\mathbb{C}$. On le démontre à partir du lemme de confinement suivant : si $z_1, \ldots, z_n$ sont des nombres complexes de module $\leqslant 1$ et de somme nulle, il existe une permutation $\sigma \in \mathcal{S}_n$ telle que $\Big|\sum_{i=1}^{p} z_{\sigma(i)}\Big| \leqslant \sqrt{5}$ pour tout p (le lecteur pourra trouver ces différents résultats dans GONNORD (S.) & TOSEL (N.), Thèmes d'analyse pour l'Agrégation, *Topologie et analyse fonctionnelle, Ellipses, 1996, p. 18-22). Dans l'exercice suivant est établi un lemme de confinement analogue. On ne s'autorise plus à permuter les z_i mais on peut leur attribuer un signe. Pour un bon choix de ces signes, les sommes partielles auront toutes un module $\leqslant \sqrt{3}$. La deuxième question en donne une application spectaculaire aux séries : toute série complexe $\sum z_n$ dont le terme général tend vers 0 peut être rendue convergente en «signant» convenablement les z_n.*

1.19. Un lemme de confinement

1. Soit $z_1, z_2, \ldots, z_n$ des nombres complexes de module inférieur ou égal à 1. Montrer qu'il existe un n-uplet $(\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ tel que pour tout $p \in [\![1, n]\!]$, $|\varepsilon_1 z_1 + \cdots + \varepsilon_p z_p| \leqslant \sqrt{3}$.

2. Soit $(z_n)_{n \geqslant 0}$ une suite de nombres complexes tendant vers 0. Montrer qu'il existe une suite $(\varepsilon_n)_{n \geqslant 0}$ d'éléments de $\{\pm 1\}$ telle que la série $\sum \varepsilon_n z_n$ converge.

(École normale supérieure)

Solution.

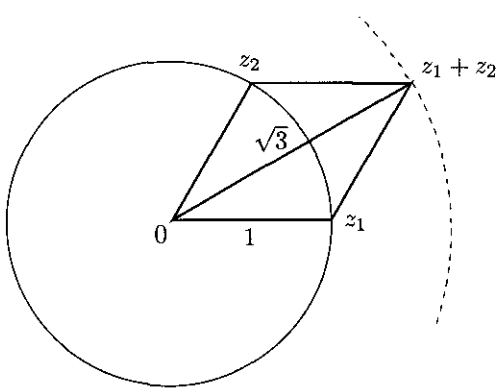
1. Notons $\mathrm{D} = \{z \in \mathbb{C},\ |z| \leqslant 1\}$ le disque unité fermé de $\mathbb{C}$. On va montrer par récurrence sur $n \geqslant 1$ la validité de la proposition (H_n) suivante

$$(\mathrm{H}_n) \quad \begin{cases} \text{Pour tout } (z_1, \ldots, z_n) \in \mathrm{D}^n, \quad \text{il existe } (\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n \\ \quad \text{tel que pour tout } k \in [\![1, n]\!], \quad \left| \sum_{i=1}^{k} \varepsilon_i z_i \right| \leqslant \sqrt{3}. \end{cases}$$

• (H_1) étant triviale, montrons (H_2). Soit $(z_1, z_2) \in \mathrm{D}^2$. Si $|z_1 - z_2| \leqslant 1$ alors le choix $\varepsilon_1 = 1$ et $\varepsilon_2 = -1$ convient. Supposons $|z_1 - z_2| > 1$. On a alors $|z_1 - z_2|^2 = |z_1|^2 + |z_2|^2 - 2\,\mathrm{Re}(\overline{z_1} z_2) > 1$ de sorte que,

$$|z_1 + z_2|^2 = |z_1|^2 + |z_2|^2 + 2\,\mathrm{Re}(\overline{z_1} z_2) \leqslant 2(|z_1|^2 + |z_2|^2) - 1 \leqslant 4 - 1 = 3$$

Le choix $\varepsilon_1 = \varepsilon_2 = 1$ convient.



La figure ci-dessus montre le cas d'égalité : on a alors $|z_1| = |z_2| = |z_1 - z_2| = 1$.

Pour prouver (H_2) *on aurait aussi pu utiliser l'égalité du parallélogramme :* $|z_1 - z_2|^2 + |z_1 + z_2|^2 = 2(|z_1|^2 + |z_2|^2) \leqslant 4$ *donc* $|z_1 - z_2| \leqslant \sqrt{2}$ *ou* $|z_1 + z_2| \leqslant \sqrt{2}$.

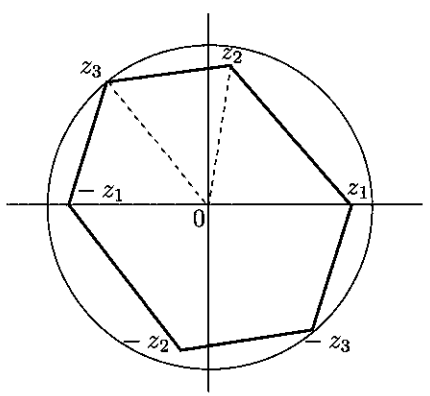
• On suppose (H_{n-1}) vérifiée avec $n \geqslant 3$ et on se donne $(z_1, \ldots, z_n)$ dans D^n. L'idée est de regrouper ensembles deux des z_i pour se ramener à l'hypothèse de récurrence. Distinguons deux cas.

⋆ Cas 1 : il existe $\varepsilon \in \{\pm 1\}$ tel que $z_1 + \varepsilon z_2 \in D$. On applique l'hypothèse de récurrence (H_{n-1}) à $(z_1 + \varepsilon z_2, z_3, \ldots, z_n) \in D^{n-1}$. Notons $(\varepsilon_2, \ldots, \varepsilon_n)$ une suite de signes donnée par l'hypothèse de récurrence pour ce $(n-1)$-uplet. La suite $(\varepsilon_2, \varepsilon\varepsilon_2, \varepsilon_3, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ convient pour $(z_1, \ldots, z_n)$. En effet, $|\varepsilon_2 z_1| \leqslant 1 \leqslant \sqrt{3}$, $|\varepsilon_2 z_1 + \varepsilon_2 \varepsilon z_2| \leqslant 1 < \sqrt{3}$ et pour $k \geqslant 3$,

$$\left|\varepsilon_2(z_1 + \varepsilon z_2) + \sum_{i=3}^{k} \varepsilon_i z_i\right| \leqslant \sqrt{3}.$$

⋆ Cas 2 : les nombres complexes $z_1 + z_2$ et $z_1 - z_2$ ne sont pas dans D. D'après la preuve du cas $n = 2$, on a $|z_1 + z_2| \leqslant \sqrt{3}$ et $|z_1 - z_2| \leqslant \sqrt{3}$. Ainsi, tous les choix possibles pour ε_1 et ε_2 conviendront. Pour se ramener à l'hypothèse de récurrence on va utiliser z_3.

Commençons par prouver que l'un des nombres complexes $z_1 - z_3$, $z_1 + z_3$, $z_2 - z_3$ ou $z_2 + z_3$ est dans D. On raisonne par l'absurde. Si ce n'est pas le cas, tous les nombres $z_1 \pm z_2$, $z_1 \pm z_3$, $z_2 \pm z_3$ sont de module > 1. Les nombres $\pm z_1$, $\pm z_2$, $\pm z_3$ sont alors non nuls, et d'arguments différents. Quitte à permuter les indices ou à remplacer z_i par $-z_i$, on peut supposer que $z_1, z_2, z_3, -z_1, -z_2, -z_3$ est un hexagone non croisé :



Les côtés de cet hexagone sont exactement les six nombres $|z_1 \pm z_2|$, $|z_1 \pm z_3|$, $|z_2 \pm z_3|$. Nécessairement, l'un des angles au centre de cet hexagone est $\leqslant \frac{\pi}{3}$. Le côté correspondant a une longueur $\leqslant 1$ (écrire l'identité d'Al Kachi pour s'en convaincre).

Revenons maintenant à la preuve de (H_n) en traitant par exemple le cas où $z_1 + \varepsilon z_3 \in D$ $(\varepsilon = \pm 1)$, le cas où $z_2 + \varepsilon z_3 \in D$ se faisant

de même. On applique l'hypothèse de récurrence (H_{n-1}) au $(n-1)$-uplet $(z_1 + \varepsilon z_3, z_2, z_4, \ldots, z_n)$. Notons $(\varepsilon_1, \varepsilon_3, \varepsilon_4, \ldots, \varepsilon_n)$ un $(n-1)$-uplet de signes ainsi obtenu. Alors on prend $(\varepsilon_1, \varepsilon_2, \varepsilon\varepsilon_1, \varepsilon_4, \ldots, \varepsilon_n)$ pour $(z_1, \ldots, z_n)$. Vérifions que ce choix convient : $|\varepsilon_1 z_1| = 1 < \sqrt{3}$, $|\varepsilon_1 z_1 + \varepsilon_2 z_2| \leqslant \sqrt{3}$ (d'après la remarque faite plus haut), $|\varepsilon_1 z_1 + \varepsilon_2 z_2 + \varepsilon\varepsilon_1 z_3| = |\varepsilon_1(z_1 + \varepsilon z_3) + \varepsilon_2 z_2| \leqslant \sqrt{3}$, etc.

2. Il résulte de ce qui précède, que si $a_1, \ldots, a_n$ sont des nombres complexes bornés par $M > 0$, alors il existe $(\varepsilon_1, \ldots, \varepsilon_n) \in \{\pm 1\}^n$ tel que

$$\forall k \in [\![1, n]\!], \quad \left|\sum_{i=1}^{k} \varepsilon_i a_i\right| \leqslant M\sqrt{3}.$$

La suite (z_n) étant bornée, on peut se donner $M > 0$ tel que pour tout $n \in \mathbb{N}$, $|z_n| \leqslant M$. On considère alors une suite strictement croissante d'entiers N_k, telle que $N_0 = 0$ et, pour tout $k \in \mathbb{N}$, on ait : $\forall n \geqslant N_k$, $|z_n| \leqslant \dfrac{M}{2^k}$. Cela est clairement possible puisque (z_n) tend vers 0.

On définit maintenant la suite (ε_n) sur chaque tranche $[\![N_k, N_{k+1} - 1]\!]$ de manière à avoir pour tout $p \in [\![N_k, N_{k+1} - 1]\!]$,

$$\left|\sum_{j=N_k}^{p} \varepsilon_j z_j\right| \leqslant \frac{\sqrt{3}M}{2^k}.$$

Notons U_k la somme de la tranche de Cauchy de la série $\sum \varepsilon_n z_n$ d'indices entre N_k et $N_{k+1} - 1$:

$$U_k = \sum_{j=N_k}^{N_{k+1}-1} \varepsilon_j z_j.$$

Comme $|U_k| \leqslant \dfrac{\sqrt{3}M}{2^k}$, la série $\sum U_k$ est absolument convergente. Si pour $n \in \mathbb{N}$, on note $\varphi(n)$ l'unique indice k tel que $N_k \leqslant n < N_{k+1}$, on a

$$\sum_{i=1}^{n} \varepsilon_i z_i = \sum_{k=0}^{\varphi(n)-1} U_k + \underbrace{\sum_{j=N_k}^{n} \varepsilon_i z_i}_{\leqslant \frac{\sqrt{3}M}{2^{\varphi(n)}}}.$$

Cela prouve que la série $\sum \varepsilon_i z_i$ converge. ◁

L'exercice suivant propose une équation fonctionnelle dans $\mathbb{R}[X]$.

1.20. Équation fonctionnelle de Shapiro

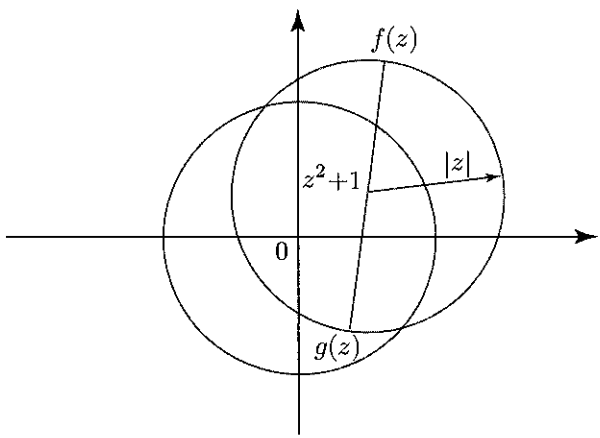
1. Déterminer les parties finies non vides A de $\mathbb{C}$ stables par $f : z \longmapsto z^2 + z + 1$ et $g : z \longmapsto z^2 - z + 1$.

2. En déduire $\{P \in \mathbb{R}[X],\ P(X^2 + X + 1) = P(X)P(X + 1)\}$.

(**École polytechnique**)

▷ **Solution.**

1. Les points $f(z)$ et $g(z)$ sont sur le cercle de centre $z^2 + 1$ et de rayon $|z|$ et comme $|f(z) - g(z)| = 2|z|$, ils y sont même diamétralement opposés.



Si $z^2 + 1 \neq 0$, on est certain que l'un des deux nombres complexes $f(z)$ ou $g(z)$ est de module strictement supérieur à $|z|$.

Il en résulte que si on choisit $z \in A$ de module maximal, on a nécessairement $z^2+1 = 0$ *i.e.* $z \in \{-i, i\}$. Comme $g(i) = -i$ et $g(-i) = i$, on a donc $\{-i, i\} \subset A$ et de plus tout point de $A \setminus \{-i, i\}$ est de module strictement inférieur à 1. On va montrer par l'absurde que A est égal à l'ensemble $\{-i, i\}$. Supposons donc qu'il existe $z \in A \setminus \{-i, i\}$. On le choisit de module maximal. Alors $f(z)$ ou $g(z)$ est de module strictement plus grand donc égal à i ou à $-i$. Or les antécédents de i par f sont i et $-1-i$, ceux de $-i$ sont $-i$ et $-1+i$. Pour g on trouve les opposés. Ils sont tous de module $\geqslant 1$, ce qui fournit la contradiction. Donc $A = \{-i, i\}$.

2. Les polynômes constants solutions sont 0 et 1. Soit P une solution avec $\deg P \geqslant 1$. Alors P est unitaire et l'ensemble des racines de P est stable par f et par g (car $(z-1)^2 + (z-1) + 1 = z^2 - z + 1$). Il s'agit donc de $\{-i, i\}$. Comme P est réel, les multiplicités des racines i et $-i$ sont égales et P est de la forme $(X^2 + 1)^n$, $n \in \mathbb{N}^*$. On vérifie alors très facilement que tous ces polynômes conviennent. ◁

1.21. Homographies laissant invariant le disque unité

Déterminer le groupe des homographies f de $\mathbb{C}$ laissant invariant D_0 (*i.e.* telles que $f(D_0) = D_0$), où D_0 est le disque fermé de centre 0 et de rayon 1.

(École polytechnique)

Solution.

Une homographie est une application de la forme $z \longmapsto \dfrac{az+b}{cz+d}$, où $ad - bc \neq 0$ (si $ad - bc = 0$, l'homographie dégénère en une application constante). Une telle application réalise une bijection de $\mathbb{C}$ sur $\mathbb{C}$ si $c = 0$ et de $\mathbb{C} \setminus \left\{ -\dfrac{d}{c} \right\}$ sur $\mathbb{C} \setminus \left\{ \dfrac{a}{c} \right\}$ sinon.

Si une telle homographie f envoie D_0 sur D_0, elle envoie aussi le complémentaire de D_0 dans le complémentaire de D_0. Notons C_0 le cercle de centre 0 et de rayon 1. Tout point z de C_0 est limite d'une suite (z_n) de $\mathbb{C} \setminus D_0$. Par continuité, $f(z)$ est limite de la suite $(f(z_n))$ de $\mathbb{C} \setminus D_0$. Ainsi $f(z)$ appartient à l'adhérence de $\mathbb{C} \setminus D_0$. Par ailleurs, $f(z)$ appartient à D_0, donc z appartient à C_0. On a donc $f(C_0) \subset C_0$. Nous allons commencer par chercher les homographies qui envoient C_0 dans C_0.

On doit avoir, pour tout $t \in \mathbb{R}$, $|f(e^{it})| = 1$, soit $|ae^{it} + b| = |ce^{it} + d|$. On obtient

$$|a|^2 + |b|^2 + 2\,\mathrm{Re}(a\bar{b}e^{it}) = |c|^2 + |d|^2 + 2\,\mathrm{Re}(c\bar{d}e^{it})$$

et il existe $t_0 \in \mathbb{R}$ tel que

$$|a|^2 + |b|^2 - |c|^2 - |d|^2 = 2\,\mathrm{Re}((c\bar{d} - a\bar{b})e^{it}) = 2|c\bar{d} - a\bar{b}|e^{i(t_0+t)}.$$

On a nécessairement

$$c\bar{d} - a\bar{b} = 0 \quad \text{et} \quad |a|^2 + |b|^2 = |c|^2 + |d|^2,$$

sinon le membre de gauche de l'égalité dépend de t.

Si $a = 0$, alors $d = 0$ (car $ad - bc \neq 0$) et $|b| = |c|$. On obtient, pour tout $z \in \mathbb{C}^*$,

$$|f(z)| = \left| \frac{b}{cz} \right| = \frac{1}{|z|}.$$

L'application f envoie D_0 dans son complémentaire et ne convient pas.

Si $a \neq 0$, on écrit $\bar{b} = \dfrac{c\bar{d}}{a}$ et on a

$$|a|^2+|b|^2-|c|^2-|d|^2 = |a|^2+\frac{|c|^2|d|^2}{|a|^2}-|c|^2-|d|^2 = \frac{(|a|^2-|c|^2)(|a|^2-|d|^2)}{|a|^2}.$$

Cette quantité étant nulle on a donc $|a| = |c|$ ou $|a| = |d|$.

Si $|a| = |c|$, il existe $\alpha \in \mathbb{R}$ tel que $c = ae^{i\alpha}$. On obtient ensuite $\overline{b} = \overline{d}e^{i\alpha}$ et donc $d = be^{i\alpha}$, ce qui donne $ad - bc = 0$. Il faut donc $|a| = |d| \neq 0$. On obtient alors

$$f(z) = \frac{a}{d}\left(\frac{z + \dfrac{b}{a}}{1 + \dfrac{c}{d}z}\right).$$

Il existe $\alpha \in \mathbb{R}$ tel que $\dfrac{a}{d} = e^{i\alpha}$. On pose $\dfrac{b}{a} = -z_0$. On a alors

$$\frac{c}{d} = \frac{a\overline{b}}{d\overline{d}} = \frac{a\overline{b}}{a\overline{a}} = \frac{\overline{b}}{\overline{a}} = -\overline{z_0}$$

et

$$f(z) = e^{i\alpha}\frac{z - z_0}{1 - \overline{z_0}z}.$$

Notons de plus que $z_0 = f^{-1}(0)$ donc z_0 appartient à $\mathrm{D}_0 \setminus \mathrm{C}_0$ et $|z_0| < 1$.

Il ne reste plus qu'à vérifier que réciproquement si $\alpha \in \mathbb{R}$ et $z_0 \in \mathbb{C}$, avec $|z_0| < 1$, l'homographie $f : z \longmapsto e^{i\alpha}\dfrac{z - z_0}{1 - \overline{z_0}z}$ envoie D_0 sur D_0. La fonction f est définie sur D_0 et, pour $z \in \mathrm{D}_0$,

$$|f(z)|^2 = \frac{|z - z_0|^2}{|1 - \overline{z_0}z|^2} = \frac{|z|^2 + |z_0|^2 - 2\,\mathrm{Re}(z\overline{z_0})}{1 + |z\overline{z_0}|^2 - 2\,\mathrm{Re}(z\overline{z_0})} \leqslant 1,$$

car $|z|^2 + |z_0|^2 - 1 - |z|^2|z_0|^2 = (|z|^2 - 1)(1 - |z_0|^2) \leqslant 0$. On a donc $f(\mathrm{D}_0) \subset \mathrm{D}_0$. On montre aisément que f^{-1} est définie par

$$f^{-1}(z) = e^{-i\alpha}\frac{z + e^{i\alpha}z_0}{1 + \overline{z_0}e^{-i\alpha}}.$$

L'homographie f^{-1} a la même forme que f, α et z_0 étant remplacé par $-\alpha$ et $-e^{i\alpha}z_0$. On a donc également $f^{-1}(\mathrm{D}_0) \subset \mathrm{D}_0$, soit $\mathrm{D}_0 \subset f(\mathrm{D}_0)$ et finalement $f(\mathrm{D}_0) = \mathrm{D}_0$.

Conclusion. Une homographie laisse D_0 invariant s'il existe $\alpha \in \mathbb{R}$ et $z_0 \in \mathbb{C}$, avec $|z_0| < 1$ tels que f soit la fonction $z \longmapsto e^{i\alpha}\dfrac{z - z_0}{1 - \overline{z_0}z}$. ◁

Si on cherche les homographies qui laissent invariant le disque ouvert de centre 0 *et de rayon* 1, *on trouve le même résultat.*

L'exercice suivant fait appel à de nombreuses notions. On y utilise les nombres complexes pour donner une interprétation géométrique du problème posé mais ensuite on constate qu'il s'agit essentiellement d'étudier une suite récurrente linéaire, donc de calculer les puissances d'une matrice.

1.22. Suites de polygones

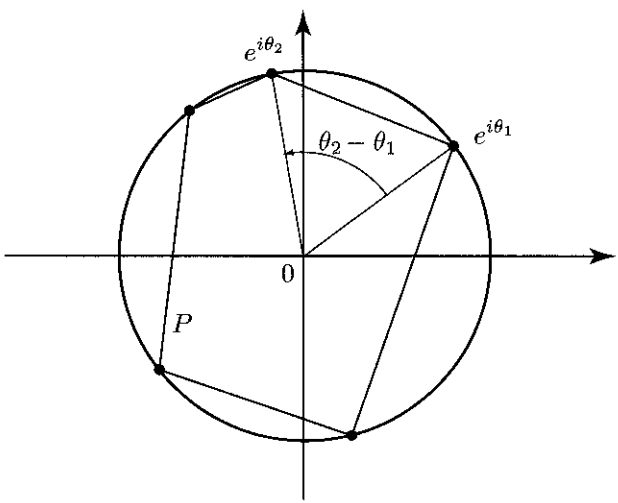
Soit $0 \leqslant \theta_1 \leqslant \cdots \leqslant \theta_n < 2\pi$ et a, b des réels positifs ou nuls de somme 1. On définit $\theta_j^{(0)} = \theta_j$, puis $\theta_j^{(k+1)} = a\theta_j^{(k)} + b\theta_{j+1}^{(k)} \pmod{2\pi}$ si $1 \leqslant j < n$ et $\theta_n^{(k+1)} = a\theta_n^{(k)} + b\theta_1^{(k)} \pmod{2\pi}$.

Étudier la limite, lorsque k tend vers $+\infty$, de $\theta_{j+1}^{(k)} - \theta_j^{(k)}$.

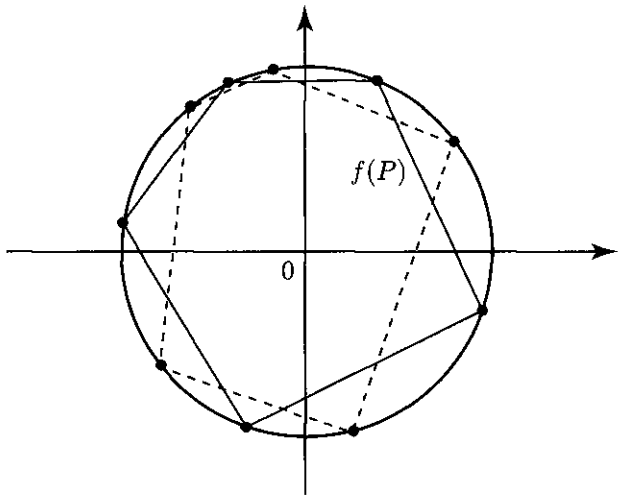
(**École normale supérieure**)

Solution.

Il importe de visualiser ce problème géométriquement. Aux n réels θ_j on associe le polygone convexe P inscrit dans le cercle unité de $\mathbb{C}$ dont les sommets sont les $e^{i\theta_j}$, $1 \leqslant j \leqslant n$. Les quantités $\alpha_j = \theta_{j+1} - \theta_j$ s'interprètent alors comme les angles au centre de ce polygone. On posera $\alpha_n = \theta_1 - \theta_n + 2\pi$ de manière à ce que la somme des α_j soit toujours égale à 2π.



La transformation étudiée s'interprète alors de la manière suivante. Au polygone $\mathrm{P}_k = (\theta_j^{(k)})_{1\leqslant j\leqslant n}$, on associe le polygone $f(\mathrm{P}_k) = (\theta_j^{(k+1)})_{1\leqslant j\leqslant n}$ où $\theta_j^{(k+1)}$ est le barycentre de $\theta_j^{(k)}$ et $\theta_{j+1}^{(k)}$, affectés des coefficients positifs a et b. Par exemple, pour $a = b = 1/2$, les sommets de $f(\mathrm{P}_k)$ sont obtenus en prenant l'intersection avec le cercle unité des médiatrices des côtés de P_k. Voici par exemple l'image par f du polygone précédent dans le cas $a = b = 1/2$:



L'exercice demande d'étudier le comportement, lorsqu'on fait tendre k vers l'infini, des angles $\alpha_j^{(k)} = \theta_{j+1}^{(k)} - \theta_j^{(k)}$.

Bien entendu, si $a = 1$ la fonction f est constante et les suites $\alpha_j^{(k)}$ sont nulles. Dans le cas $a = 0$ et $b = 1$, le polygone $f(\mathrm{P})$ est le même que P, mais la numérotation des sommets y a subi un décalage d'une unité. Dans ce cas, les suites $\alpha_j^{(k)}$ sont périodiques de période n. Dans la suite on va donc se préoccuper du cas $0 < a < 1$. L'observation de quelques exemples va nous permettre d'avoir une intuition du résultat.

• Si on prend $n = 2$ et $a = b = 1/2$, on obtient deux points alignés après une itération de la transformation f. Et les suites $\alpha_j^{(k)} = \theta_{j+1}^{(k)} - \theta_j^{(k)}$ pour $j = 1, 2$ sont alors stationnaires à π.

• Plus généralement, si les n points forment un polygone régulier P, alors $f(\mathrm{P})$ est un autre polygone régulier obtenu par rotation à partir de P. Dans ce cas les suites $\alpha_j^{(k)}$ sont constantes égales à $\dfrac{2\pi}{n}$.

On va en fait montrer que lorsque $0 < a < 1$, toutes les suites $\alpha_j^{(k)}$ convergent vers $\dfrac{2\pi}{n}$. Dans ce qui suit, l'indice j sera lu modulo n. On a, pour tout j et tout k,

$$\alpha_j^{(k+1)} = \theta_{j+1}^{(k+1)} - \theta_j^{(k+1)} = a(\theta_{j+1}^{(k)} - \theta_j^{(k)}) + b(\theta_{j+2}^{(k)} - \theta_{j+1}^{(k)}) = a\alpha_j^{(k)} + b\alpha_{j+1}^{(k)}$$

Si, pour simplifier, on note X_k le vecteur de $\mathbb{R}^n$ de coordonnées $(\alpha_1^{(k)}, \ldots, \alpha_n^{(k)})$, on a donc tout simplement $\mathrm{X}_{k+1} = \mathrm{A}\mathrm{X}_k$, où A est la matrice

$$\begin{pmatrix} a & b & 0 & \dots & 0 \\ 0 & a & b & & \\ \vdots & & \ddots & \ddots & \\ 0 & & & \ddots & b \\ b & 0 & \dots & 0 & a \end{pmatrix} = a\mathrm{I}_n + b\mathrm{M}.$$

On est ramené à un problème d'algèbre linéaire, à savoir le calcul des puissances de A. Or, la matrice M est une matrice de permutation, de polynôme caractéristique $\mathrm{X}^n - 1$. Elle est diagonalisable, ses valeurs propres étant les racines n-ièmes de l'unité. Il en résulte que les valeurs propres de A sont les $\lambda_\omega = a + b\omega$ où $\omega^n = 1$. Elles sont deux à deux distinctes et de module inférieur à 1 (λ_ω est sur le segment joignant 1 à ω), la seule qui est de module 1 étant $\lambda_1 = a + b = 1$. Ainsi, si on note p_ω la projection sur la droite propre pour λ_ω parallèlement à la somme des autres sous-espaces propres de A, on a $\mathrm{A} = \sum_{\omega^n=1} \lambda_\omega p_\omega$ et pour tout k, $\mathrm{A}^k = \sum_{\omega^n=1} \lambda_\omega^k p_\omega$. Par suite, A^k tend vers la projection p_1 sur la droite propre pour la valeur propre 1, à savoir la droite engendrée par le vecteur $(1, 1, \dots, 1)$ parallèlement à la somme des autres sous-espaces propres.

Ainsi, $\mathrm{A}^k \mathrm{X}_0$ tend vers un vecteur dont toutes les composantes sont égales. Comme on a $\sum_{j=1}^{n} \alpha_j^{(k)} = 2\pi$ pour tout k, ces composantes valent $\frac{2\pi}{n}$. Le polygone $f^k(\mathrm{P})$ «converge» donc vers un polygone régulier. ◁

Les deux exercices qui suivent sont liés à la notion de mesure mais ne font intervenir que la longueur pour un intervalle. Le premier définit la mesure extérieure de Lebesgue d'une partie de $\mathbb{R}$. *On y utilise le théorème de Borel-Lebesgue : de tout recouvrement d'un compact de* $\mathbb{R}$ *par une famille d'ouverts on peut extraire un sous-recouvrement fini.*

1.23. Mesure extérieure de Lebesgue

Soit $\mathrm{E} \subset \mathbb{R}$. Pour toute famille au plus dénombrable $\mathrm{F} = (]a_i, b_i[)_{i\in\mathrm{I}}$ d'intervalles ouverts recouvrant E, on considère $l(\mathrm{F}) = \sum_{i\in\mathrm{I}}(b_i - a_i) \in \mathbb{R} \cup \{+\infty\}$. On définit alors la *mesure extérieure de Lebesgue* de E par la formule $\mu(\mathrm{E}) = \inf_{\mathrm{F}} l(\mathrm{F})$, où la borne inférieure est prise sur toutes les familles au plus dénombrables d'intervalles ouverts recouvrant E.

1. Quelle est la mesure extérieure d'un intervalle ?

2. Montrer qu'une réunion dénombrable de parties de mesure extérieure nulle est encore de mesure extérieure nulle.

3. Quelle est la mesure extérieure de $\mathbb{Q}$?

(École normale supérieure)

▷ **Solution.**

1. Nous allons démontrer que la mesure extérieure d'un intervalle est sa longueur, ce qui n'est pas évident sur la définition.

Supposons d'abord que E soit un intervalle compact $[a,b]$. Si F est un recouvrement de E, on peut, d'après le théorème de Borel-Lebesgue, en extraire un sous-recouvrement fini $\mathrm{F}' = (]a_{i_k}, b_{i_k}[)_{1\leqslant k\leqslant n}$. De l'inclusion $[a,b] \subset \bigcup_{1\leqslant k\leqslant n}]a_{i_k}, b_{i_k}[$, on déduit que $\chi_{[a,b]} \leqslant \sum_{k=1}^{n} \chi_{]a_{i_k},b_{i_k}[}$ (où χ_{I} désigne la fonction caractéristique de I)[2] et donc

$$b-a = \int_{\mathbb{R}} \chi_{[a,b]}(x)\mathrm{d}x \leqslant \sum_{k=1}^{n} \int_{\mathbb{R}} \chi_{]a_{i_k},b_{i_k}[}(x)\mathrm{d}x = \sum_{k=1}^{n} (b_{i_k}-a_{i_k}) = l(\mathrm{F}').$$

Par définition de la somme de la famille $(b_i - a_i)_{i\in\mathrm{I}}$, on a, *a fortiori*, $b-a \leqslant l(\mathrm{F})$. Ceci étant vrai pour tout famille F d'intervalles ouverts recouvrant $[a,b]$, on en déduit que $b-a \leqslant \mu([a,b])$.

Afin de montrer une inégalité de sens inverse, on remarque que, pour tout $\varepsilon > 0$, la famille à un élément $\mathrm{F} = (]a-\varepsilon, b+\varepsilon[)$ recouvre $[a,b]$. Il s'ensuit que $\mu([a,b]) \leqslant l(\mathrm{F}) = b-a+2\varepsilon$. Ceci étant vrai pour tout $\varepsilon > 0$, on en déduit que $\mu([a,b]) \leqslant b-a$ et donc que $\mu([a,b]) = b-a$.

Pour traiter le cas général, remarquons que la fonction μ est croissante pour l'inclusion. En effet, si $\mathrm{E} \subset \mathrm{E}'$ et si F est une famille au plus dénombrable d'intervalles ouverts recouvrant E', alors F recouvre E. On en déduit que $\mu(\mathrm{E}) \leqslant l(\mathrm{F})$, puis, par définition de $l(\mathrm{E}')$, que $\mu(\mathrm{E}) \leqslant \mu(\mathrm{E}')$. Si donc E est un intervalle borné d'extrémités a et b, on a, pour $\varepsilon \in]0, \dfrac{b-a}{2}[$,

$$b-a-2\varepsilon = \mu([a+\varepsilon, b-\varepsilon]) \leqslant \mu(]a,b[) \leqslant \mu(\mathrm{E}) \leqslant \mu([a,b]) = b-a$$

et donc $\mu(\mathrm{E}) = b-a$.

Si enfin E est un intervalle non borné, on peut, pour tout $\mathrm{A} > 0$, trouver un intervalle borné I, inclus dans E, de longueur supérieure à A. On en déduit que $\mu(\mathrm{E}) \geqslant \mu(\mathrm{I}) \geqslant \mathrm{A}$ et donc que $\mu(\mathrm{E}) = +\infty$.

---

2. L'utilisation des fonctions caractéristiques et des intégrales est très commode mais pas indispensable. On peut aussi raisonner, comme nous le faisons dans l'exercice suivant, en extrayant de F' un sous-recouvrement F'' de cardinal minimal.

2. Soit $(\mathrm{E}_n)_{n\in\mathbb{N}}$ une famille dénombrable de parties telles que $\mu(\mathrm{E}_n) = 0$ pour tout n et $\mathrm{E} = \bigcup_{n\in\mathbb{N}} \mathrm{E}_n$. Donnons-nous $\varepsilon > 0$. Par définition, pour tout $n \in \mathbb{N}$, il existe une famille au plus dénombrable $\mathrm{F}_n = (]a_{n,i}, b_{n,i}[)_{i\in\mathrm{I}_n}$ d'intervalles ouverts recouvrant E_n telle que $l(\mathrm{F}_n) \leqslant \frac{\varepsilon}{2^{n+1}}$. L'ensemble $\mathrm{I} = \bigcup_{n\in\mathbb{N}} \{n\} \times \mathrm{I}_n$ est au plus dénombrable, car réunion dénombrable d'ensembles au plus dénombrables. La famille $\mathrm{F} = (]a_{n,i}, b_{n,i}[)_{(n,i)\in\mathrm{I}}$ recouvre E et on a, d'après les propriétés des familles sommables,

$$\begin{aligned}\mu(\mathrm{E}) &\leqslant \sum_{(n,i)\in\mathrm{I}} (b_{n,i} - a_{n,i}) = \sum_{n\in\mathbb{N}} \left(\sum_{i\in\mathrm{I}_n} (b_{n,i} - a_{n,i}) \right) = \sum_{n\in\mathbb{N}} l(\mathrm{F}_n) \\ &\leqslant \sum_{n\in\mathbb{N}} \frac{\varepsilon}{2^{n+1}} = \varepsilon.\end{aligned}$$

Par définition de μ, on a donc $\mu(\mathrm{E}) = 0$.

3. L'ensemble $\mathbb{Q}$ est dénombrable donc réunion d'une famille dénombrable de singletons, c'est-à-dire de segments de longueur nulle. Il résulte alors des questions 1 et 2 que $\mu(\mathbb{Q}) = 0$. ◁

En réalité, toutes les parties envisagées dans cet exercice sont des parties mesurables au sens de Lebesgue. Pour de telles parties, la mesure extérieure coïncide avec la mesure de Lebesgue. Notons que les deux premières questions fournissent une autre preuve de la non-dénombrabilité d'un intervalle non trivial (voir l'exercice 1.3) puisque les singletons ont une mesure extérieure nulle.

1.24. Recouvrements

On suppose donné pour tout réel x un intervalle ouvert $\mathrm{I}(x)$ centré en x et dont la longueur, notée $\mu(\mathrm{I}(x))$, est majorée par $c > 0$ fixé. Soit $a < b$ deux réels. Montrer qu'il existe un nombre fini de réels $x_1, \ldots, x_n$ tels que

$$(i)\ [a,b] \subset \bigcup_{i=1}^{n} \mathrm{I}(x_i) \text{ et } (ii)\ \sum_{i=1}^{n} \mu(\mathrm{I}(x_i)) \leqslant 2(b-a) + c.$$

(École polytechnique)

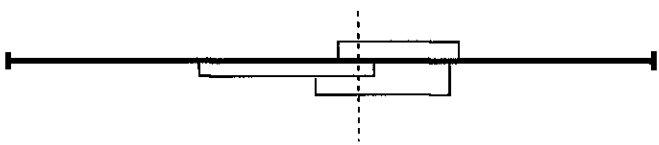
Solution.

Le fait que l'on puisse extraire de la famille $(\mathrm{I}(x))_{x\in[a,b]}$ un recouvrement fini du segment $[a,b]$ résulte du théorème de Borel-Lebesgue, le segment $[a,b]$ étant compact. Cela permet de réaliser la condition (i). Pour satisfaire la condition (ii), il faut toutefois ne pas prendre trop d'intervalles. Il semble donc raisonnable de choisir parmi tous les sous-recouvrements finis de cette famille, un sous-recouvrement de cardinal

minimal n. On note $\mathrm{I}(x_1), \ldots, \mathrm{I}(x_n)$ les intervalles de ce recouvrement et, pour tout i, $\mathrm{I}(x_i) =]a_i, b_i[$.

Examinons les propriétés de ce recouvrement. Elles nous feront comprendre d'où vient le majorant $2(b-a)+c$.

(a) Tout point de $]a,b[$ appartient au plus à *deux* des intervalles de ce recouvrement. En effet, si α appartient à trois intervalles, l'un de ces trois intervalles est inclus dans la réunion des deux autres et peut donc être éliminé, ce qui contredit la minimalité de n.

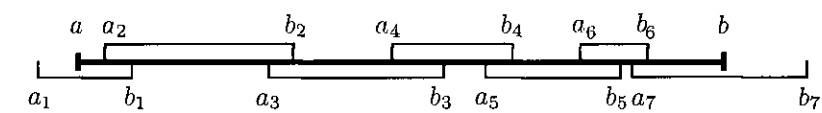


(b) Les points a et b appartiennent chacun à au plus *un* des intervalles du recouvrement. En effet, si le point a appartient à deux intervalles $]a_{n_1}, b_{n_1}[$ et $]a_{n_2}, b_{n_2}[$ et que par exemple $b_{n_1} \leq b_{n_2}$, on peut retrancher l'intervalle $\mathrm{I}(x_{n_1})$ du sous-recouvrement sans qu'il cesse d'être un recouvrement, ce qui contredit à nouveau la minimalité de n. Il en est de même en b.

(c) Aucun des intervalles du recouvrement n'est contenu dans un autre intervalle du recouvrement.

On peut alors supposer que les $\mathrm{I}(x_i)$ sont rangés dans «l'ordre croissant», c'est-à-dire que

$$a_1 < a_2 < b_1 < a_3 < b_2 < a_4 < b_3 < \cdots < a_n < b_{n-1} < b_n.$$



D'après la propriété (b) on a $a_1 < a < a_2$ et $b_{n-1} < b < b_n$.

Comme, par construction, le centre x_1 de l'intervalle $]a_1, b_1[$ est dans le segment $[a,b]$, on a $a - a_1 \leqslant \dfrac{c}{2}$ (il peut y avoir égalité si $x_1 = a$). On a, pour les mêmes raisons, $b_n - b \leqslant \dfrac{c}{2}$.

Puisque l'un au plus des intervalles du recouvrement fini dépasse à gauche de a (celui qui contient a), et qu'il dépasse d'au plus $\dfrac{c}{2}$ (son centre est à droite de a), qu'un intervalle au plus dépasse à droite de b, et d'au plus $\dfrac{c}{2}$, et que le reste de $]a,b[$ est recouvert au plus deux fois, il semble clair que (voir dessin)

$$\sum_{i=1}^{n} \mu(\mathrm{I}(x_i)) \leq 2(b-a) + \frac{c}{2} + \frac{c}{2}.$$

Montrons-le précisément. On se ramène à des intervalles disjoints en écrivant :

$$\sum_{i=1}^{n}(b_i - a_i) = (b_n - a_n) + \sum_{i=1}^{n-1}(b_i - a_{i+1}) + \sum_{i=1}^{n-1}(a_{i+1} - a_i)$$
$$= b_n - a_1 + \sum_{i=1}^{n-1}(b_i - a_{i+1}).$$

Or on a $b_n - a_1 = (b_n - b) + (b - a) + (a - a_1) \leqslant (b-a) + c$, d'après ce qui précède, et comme les intervalles $]a_{i+1}, b_i[$ pour $i \in [\![1, n-1]\!]$ sont deux à deux disjoints et inclus dans $[a, b]$, on a aussi $\sum_{i=1}^{n-1}(b_i - a_{i+1}) \leqslant b - a$. On obtient donc $\sum_{i=1}^{n}(b_i - a_i) \leqslant 2(b-a) + c$, ce qui est l'inégalité (ii) souhaitée. $\lhd$

Nous terminons ce chapitre avec quelques exercices de nature topologique. Le premier permet de rappeler que les ensembles $\mathbb{Q}$ et $\mathbb{R} \setminus \mathbb{Q}$ sont denses dans $\mathbb{R}$.

1.25. Irrationnels denses dans une partie

Soit M une partie majorée de $\mathbb{R}_+^*$ contenant au moins deux éléments et telle que pour tout $(a, b) \in M^2$, $\sqrt{ab} \in M$. Montrer que $M \cap (\mathbb{R} \setminus \mathbb{Q})$ est dense dans $[\inf M, \sup M]$.

(École normale supérieure)

Solution.

Par hypothèse on a $\inf M < \sup M$. Montrons tout d'abord que M est dense dans le segment $I = [\inf M, \sup M]$. Raisonnons par l'absurde et supposons qu'il existe $\inf M \leqslant a < b \leqslant \sup M$ tel que $]a, b[\cap M$ soit vide. On va agrandir notre intervalle $]a, b[$ au maximum en conservant cette propriété : posons

$$\alpha = \inf\{x \geqslant \inf M,\ ]x, b[\cap M = \emptyset\} \text{ et } \beta = \sup\{x \leqslant \sup M,\ ]a, x[\cap M = \emptyset\}.$$

L'intervalle ouvert $]\alpha, \beta[$ a toujours une intersection vide avec M mais, par construction, il existe des éléments de M aussi près qu'on veut de α (à gauche) et de β (à droite). En fait cet intervalle ouvert $]\alpha, \beta[$ n'est autre que la composante connexe de l'ouvert $\mathbb{R} \setminus \overline{M}$ qui contient $\dfrac{\alpha+\beta}{2}$ (voir cette notion dans la première question de l'exercice 1.27). On peut donc se donner une suite (a_n) de M qui converge vers α et une suite (b_n)

de M qui converge vers β. Les réels $\sqrt{a_n b_n}$ sont dans M par hypothèse et tendent vers $\sqrt{\alpha\beta} \in]\alpha, \beta[$. Pour n assez grand ils sont donc dans l'intervalle ouvert $]\alpha, \beta[$ ce qui amène notre contradiction.

Montons maintenant que $\mathrm{M} \cap (\mathbb{R} \setminus \mathbb{Q})$ est dense dans $[\inf \mathrm{M}, \sup \mathrm{M}]$. Raisonnons encore par l'absurde en supposant qu'il existe un intervalle ouvert $]a, b[\subset \mathrm{I}$ tel que tous les éléments de $\mathrm{M} \cap]a, b[$ soient rationnels.

En utilisant la décomposition des entiers en facteur premiers, il est facile de voir que tout nombre rationnel $x > 0$ s'écrit de manière unique sous la forme $x = \prod_{p \in \mathcal{P}} p^{\nu_p(x)}$ où $\mathcal{P}$ est l'ensemble des nombres premiers et où $(\nu_p(x))$ est une suite d'entiers relatifs presque nulle (c'est-à-dire dont seul un nombre fini de termes sont non nuls). Par exemple $\frac{12}{25} = 2^2 \cdot 3 \cdot 5^{-2}$. On aura besoin de la constatation suivante : si x est un nombre rationnel, alors $\sqrt{x} \in \mathbb{Q}$ si et seulement si les entiers $\nu_p(x)$ sont tous pairs. La condition est clairement suffisante et elle est nécessaire car si $y = \sqrt{x} \in \mathbb{Q}$ alors $x = y^2$ et $\nu_p(x) = 2\nu_p(y)$ pour tout nombre premier p.

Revenons alors à l'exercice. Par densité de M notre intervalle $]a, b[$ contient deux éléments $x < y$ de M (qui sont donc rationnels). Par hypothèse $\sqrt{xy}$ est encore dans M et comme $\sqrt{xy} \in]x, y[\subset]a, b[$ il est aussi rationnel. Recommençons en remplaçant y par $y_1 = \sqrt{xy}$ et ainsi de suite. Tous les réels de la suite définie par $y_0 = y$ et $y_{n+1} = \sqrt{xy_n}$ sont dans $\mathrm{M} \cap \mathbb{Q}$. Soit p un nombre premier. La suite $\nu_p(y_n)$ est à valeurs dans $\mathbb{Z}$ et vérifie la relation de récurrence $\nu_p(y_{n+1}) = \frac{\nu_p(x) + \nu_p(y_n)}{2}$. Il s'agit d'une suite arithmético-géométrique qui converge vers $\nu_p(x)$. Par conséquent elle est stationnaire et par suite constante (car le seul antécédent de $\nu_p(x)$ par la fonction $t \mapsto 1/2(\nu_p(x) + t)$ est $\nu_p(x)$). On a donc $\nu_p(y) = \nu_p(x)$. Comme cela vaut pour tout p on a $y = x$ et c'est une contradiction.

Conclusion. L'ensemble $\mathrm{M} \cap (\mathbb{R} \setminus \mathbb{Q})$ est dense dans I. $\lhd$

1.26. Application ouverte

Soit $\mathrm{P} \in \mathbb{C}[\mathrm{X}]$ non constant. Montrer que l'image par P d'une partie fermée (resp. ouverte) de $\mathbb{C}$ est fermée (resp. ouverte).
(École normale supérieure)

$\rhd$ **Solution.**

Il suffit de prouver le résultat pour les parties fermées. En effet, si U est un ouvert de complémentaire le fermé F alors $\mathrm{P}(\mathrm{U}) = \mathbb{C}^2 \setminus \mathrm{P}(\mathrm{F})$ car P est surjectif (le théorème de d'Alembert-Gauss assure l'existence d'une racine de $\mathrm{P}(\mathrm{X}) - w$ pour tout $w \in \mathbb{C}$). Donc $\mathrm{P}(\mathrm{U})$ est ouvert si et seulement si $\mathrm{P}(\mathrm{F})$ est fermé.

Montrons que P(F) est fermé. On suppose F non vide. Soit w_n une suite de points de P(F) qui converge vers $w \in \mathbb{C}$. On veut prouver que $w \in$ P(F). Pour tout n on choisit $z_n \in$ F tel que $P(z_n) = w_n$. La suite $(z_n)_{n\geqslant 0}$ est bornée sinon on pourrait en extraire une sous-suite $(z_{\varphi(n)})$ telle que $|z_{\varphi(n)}| \to +\infty$ et comme $|P(z)| \to +\infty$ lorsque $|z| \to +\infty$ on aurait $|w_{\varphi(n)}| \to +\infty$ ce qui est absurde. On peut donc extraire une sous-suite $(z_{\psi(n)})$ qui converge vers un point $a \in$ F. Par continuité de P la suite $(w_{\psi(n)})$ converge vers $P(a)$. Mais elle converge aussi vers w. Par unicité de la limite $w = P(a) \in$ P(F). $\lhd$

Plus généralement, si f est une fonction continue propre (c'est-à-dire telle que l'image réciproque de tout compact est compact : voir l'exercice 4.1) alors l'image directe d'un fermé par f est fermée.

La seconde question de ce dernier exercice est tirée du cours de topologie de Gustave Choquet (éditions Masson).

1.27. Partition dénombrable de [0, 1]

1. Montrer qu'un ouvert non vide de $\mathbb{R}$ s'écrit de manière unique comme réunion d'une famille d'intervalles ouverts non vides deux à deux disjoints et que cette famille est au plus dénombrable.

2. Montrer que le segment $[0, 1]$ n'est pas réunion dénombrable de fermés non vides deux à deux disjoints.

(**École normale supérieure**)

Solution.

1. Soit U un ouvert non vide de $\mathbb{R}$. On cherche à partitionner U en intervalles ouverts. On considère pour cela la relation binaire R définie sur U de la manière suivante : pour $(x, y) \in U^2$, $x\,R\,y$ si et seulement si le segment d'extrémités x et y est inclus dans U. Il s'agit visiblement d'une relation d'équivalence sur U. Montrons que les classes d'équivalence sont des intervalles ouverts. Par définition une classe d'équivalence I est convexe : c'est donc un intervalle de $\mathbb{R}$. Si $x \in$ I, comme U est ouvert, on peut trouver $\eta > 0$ tel que $]x - \eta, x + \eta[\subset$ U. Mais alors tout point de $]x - \eta, x + \eta[$ est en relation avec x, donc $]x - \eta, x + \eta[\subset$ I. Cela montre que I est ouvert. Comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, on peut trouver dans chaque classe d'équivalence un nombre rationnel, et définir ainsi une injection de U/R dans $\mathbb{Q}$. Comme $\mathbb{Q}$ est dénombrable, on en déduit que U/R est au plus dénombrable. Enfin, comme par définition on a $U = \bigcup_{I\in U/R} I$, on a bien montré que U est réunion d'une famille au plus dénombrable d'intervalles ouverts deux à deux disjoints.

Il reste à prouver qu'il s'agit de la seule manière d'écrire U comme une réunion d'intervalles ouverts deux à deux disjoints. Supposons que $U = \bigcup_{j\in J}]a_j, b_j[$, où les $]a_j, b_j[$ sont deux à deux disjoints. Soit $j \in J$. Tous les éléments de $]a_j, b_j[$ étant visiblement en relation pour R, $]a_j, b_j[$ est inclus dans une classe $I =]a, b[$ pour la relation R. En fait il y a égalité. En effet, supposons par exemple $a < a_j$. Alors $a_j \in U$ et il existe $k \in J$ tel que $a_j \in]a_k, b_k[$, ce qui est impossible car dans ce cas $]a_j, b_j[\cap]a_k, b_k[\neq \emptyset$. De la même manière on montre que $b_j = b$. Les intervalles $]a_j, b_j[$ sont donc des classes d'équivalence pour R et on les obtient toutes sans quoi la réunion des $]a_j, b_j[$ ne serait pas égale à U. D'où le résultat.

Les intervalles ouverts de U/R *sont appelés les composantes connexes de* U. *Ce qui précède montre qu'il s'agit des intervalles ouverts inclus dans* U *qui sont maximaux au sens de l'inclusion.*

2. Supposons par l'absurde qu'une telle partition soit possible et posons $[0, 1] = \bigcup_{n\geqslant 0} F_n$ où les F_n sont des fermés non vides.

L'idée de la démonstration consiste à construire une suite (I_n) d'intervalles ouverts telle que

(*i*) $I_n \subset \bar{I}_n \subset I_{n-1}$;

(*ii*) I_n ne rencontre pas F_n.

La première condition assure que $\bigcap_{n=0}^{\infty} I_n = \bigcap_{n=0}^{\infty} \bar{I}_n$. Comme l'intersection d'une suite décroissante de segments est non vide, il existe au moins un point α qui appartient à tous les I_n, et qui donc n'appartient à aucun F_n, ce qui apporte la contradiction désirée. La construction de la suite (I_n) n'est toutefois pas si simple. Il faut s'assurer à chaque étape que l'intervalle I_n ne se retrouve pas entièrement contenu dans un fermé F_p, car cela bloquerait la récurrence.

Quitte à regrouper deux des fermés et à effectuer une renumérotation on peut très bien supposer que 0 et 1 sont dans F_0 (cela permet simplement d'éviter des considérations de topologie induite).

L'ensemble $[0, 1] \setminus F_0 =]0, 1[\setminus F_0$ est donc ouvert et non vide. On prend pour I_0 l'une des composantes connexes de cet ouvert.

Pour définir I_1, on considère le premier indice k tel que F_k rencontre I_0, soit k_1. L'ensemble $I_0 \setminus F_{k_1}$ est ouvert, et il est non vide : sinon, on aurait $I_0 \subset F_{k_1}$, d'où $\bar{I}_0 \subset F_{k_1}$ puisque F_{k_1} est fermé ; or les extrémités de I_0 appartiennent à F_0, et F_0 et F_{k_1} sont disjoints.

Soit J l'une des composantes connexes de $I_0 \setminus F_{k_1}$. De deux choses l'une. Ou bien les deux extrémités de J sont dans F_{k_1} ; ce sont en particulier des points de I_0, et on a $\bar{J} \subset I_0$; on pose alors $I_1 = J$. Ou bien l'une des extrémités de J est une extrémité de I_0 et l'autre appartient à

F_{k_1} ; dans ce cas, on remplace J par un intervalle plus petit, disons J', en conservant l'extrémité qui appartient à F_{k_1}, et en rapprochant l'autre extrémité, de façon que $\overline{J'} \subset I_0$; on pose alors $I_1 = J'$.

On continue en construisant par récurrence une suite d'intervalles $(I_n)_{n\geqslant 0}$ et une suite d'entiers $0 = k_0 < k_1 < k_2 < \cdots < k_n < \cdots$ tels que

(i') $I_n \subset \bar{I}_n \subset I_{n-1}$ (avec $I_{-1} = [0, 1]$) ;

(ii') I_n ne rencontre pas F_0, F_1, ..., F_{k_n} ;

(iii') $\bar{I}_n$ rencontre F_{k_n}.

La suite $(\bar{I}_n)_{n\geqslant 0}$ est une suite décroissante de segments, donc son intersection est non vide. Si $\alpha \in [0, 1]$ est dans cette intersection, il n'est dans aucun des fermés $F_0, F_1, \ldots, F_{k_n}$. Comme cela vaut pour tout n, α n'est dans aucun F_n ce qui apporte la contradiction souhaitée. ◁

Le lecteur qui connaît le théorème de Baire pourra rédiger une preuve légèrement plus courte de ce résultat, preuve qui aura le mérite de se généraliser. En effet, on peut montrer qu'un espace métrique complet connexe et localement connexe ne peut se partitionner en une réunion dénombrable de fermés non vides[3].

3. Voir par exemple GONNORD (S.) & TOSEL (N.), *Thèmes d'analyse pour l'Agrégation*, Topologie et analyse fonctionnelle, Ellipses, 1996, p. 41.

Chapitre 2

Suites réelles et complexes

Bien avant qu'elle ne soit conceptualisée, on a utilisé des itérations où est sous-jacente la notion de suite. Par exemple, Archimède quand il cherche une valeur approchée de π considère les suites (p_n) et (P_n) des périmètres des polygones inscrits et circonscrits à un cercle de rayon 1 et aboutit à des formules qui équivalent à $p_{2n} = \sqrt{p_n \mathrm{P}_n}$ et $\mathrm{P}_{2n} = \dfrac{2\mathrm{P}_n p_{2n}}{p_{2n} + \mathrm{P}_n}$.

La théorie des suites au sens moderne est établie au début du XIX^e *siècle quand s'affirme la volonté de donner à l'analyse des bases rigoureuses qui la débarrassent des notions métaphysiques d'infiniment petits ou de quantités évanouissantes. Dans ses* Notions fondamentales de la théorie des suites, *rédigées vers 1800 et restées inédites, Gauss donne la définition moderne d'une suite (application de $\mathbb{N}$ dans $\mathbb{R}$). Il définit les notions de majorant et de borne supérieure d'une suite. Plus intéressant encore, il donne les définitions de la limite supérieure et de la limite inférieure d'une suite ($\limsup a_n = \lim\limits_{n\to+\infty} \sup\limits_{p\geqslant n} a_p$ et $\liminf a_n = \lim\limits_{n\to+\infty} \inf\limits_{p\geqslant n} a_p$, dans le langage d'aujourd'hui) et, quand ces deux quantités sont égales, appelle leur valeur commune la limite de la suite.*

C'est le Cours d'analyse *de Cauchy (1821) qui ouvre la voie à l'analyse moderne. Dans le chapitre des* Préliminaires, *il donne les définitions d'une suite et de la limite d'une suite, les premières définitions précises de $+\infty$ et $-\infty$, introduit la notion de valeur d'adhérence. Le «critère de Cauchy» de convergence, déjà connu de Bolzano est explicité pour les séries.*

Pendant une grande partie du XIX^e *siècle, la convergence d'une suite de Cauchy ou d'une suite croissante majorée sont présentés comme des axiomes qui constituent le fondement de toutes les questions où intervient la notion de limite. Ce point de vue va être remis en cause par Méray (1868) puis par Cantor (1872) qui, voulant en donner des justifications précises, construisent $\mathbb{R}$ à partir des suites de Cauchy de $\mathbb{Q}$.*

La théorie des suites réelles dont tous les concepts sont parfaitement définis depuis la fin du XIX^e *a connu récemment des développements importants avec l'étude des systèmes dynamiques, qui apportent un regard nouveau sur les suites définies par une relation de récurrence de la forme $u_{n+1} = f(u_n)$. Si la fonction f est continue et si la suite converge, sa limite l est nécessairement un point fixe de f. Ce fait,*

démontré par Cauchy, est à la base de toutes les méthodes numériques itératives. L'intérêt pour de telles suites est ancien. Dans le traité De la méthode des fluxions et des suites infinies *(1740), pour obtenir une valeur approchée d'une solution de l'équation $g(x) = 0$, Newton expose ce qu'on appelle depuis «méthode de Newton» : prenant a_0 proche de la solution de l'équation, on considère une suite vérifiant*

$$a_{n+1} = a_n - \frac{g(a_n)}{g'(a_n)}.$$

C'est lors de l'étude de certains systèmes dynamiques discrets[1] *qu'est apparue la notion de chaos, qui a connu ces dernières décennies un grand succès. Pour des fonctions f très simples (par exemple une fonction trinôme), le système dynamique peut avoir un comportement qui semble aléatoire. L'exemple le plus connu est la* suite logistique *vérifiant une relation de la forme*

$$u_{n+1} = (1+\alpha)u_n - \alpha u_n^2.$$

Cette suite été utilisée par Verhulst en 1845 pour décrire un modèle de croissance de la population. Pour $0 < \alpha \leqslant 2$ et une population u_0 pas trop importante, la suite (u_n) converge vers la population stable 1. *Mais comme l'a démontré en 1963 le météorologiste E.N. Lorenz, pour des valeurs plus grandes de α, cette loi décrit certains aspects des flux turbulents.*

Dans les années 1980, de grands progrès ont été accomplis dans l'étude de ces systèmes dynamiques grâce à la puissance des ordinateurs. Par exemple, pour la suite logistique, on observe que, pour $2 < \alpha < 2{,}5$, le comportement de la suite tend vers une oscillation régulière entre deux valeurs (cycle d'ordre 2) ; puis pour $2{,}5 \leqslant \alpha < 2{,}55$, vers un cycle d'ordre 4 ; ensuite quand α augmente, vers des cycles d'ordre 8, 16... Au-delà de 2,57 environ, le système devient chaotique. Feigenbaum a montré en 1981 que, pour une classe assez large d'applications f de $[-1,1]$ dans $[-1,1]$ et $f_\lambda = \lambda f$, $0 < \lambda < 1$, le système dynamique défini par $u_{n+1} = f_\lambda(u_n)$ a un comportement comparable : il existe une suite croissante de valeurs λ_j du paramètre λ pour lesquelles la dynamique change (le nombre de points d'un cycle double quand λ, supposé passage de λ_j) jusqu'à une valeur critique λ_∞, de telle manière que $\dfrac{\lambda_{j+1} - \lambda_j}{\lambda_{j+2} - \lambda_{j+1}}$ tende vers $\delta = 4{,}669...$, constante universelle indépendante de f. Au-delà de λ_∞, on retrouve des cycles stables de période $3 \cdot 2^j$ et des points de bifurcation.

On s'est aussi intéressé à l'itération de fonctions complexes, en particulier les fonctions $f : x \mapsto x^2 + c$, où $c \in \mathbb{C}$. On étudie l'ensemble

1. Qui revient à l'étude du comportement des applications itérées $f^n : X \to X$.

des nombres complexes z pour lesquels la suite de premier terme z est bornée. On note K_c cet ensemble, et on l'appelle ensemble de Julia. *Sa frontière présente des formes très belles et très variées selon les valeurs de c. Les premiers résultats, établis entre 1905 et 1920, sont dus à Fatou et Julia (évidemment sans aucun moyen informatique). En 1980, Mandelbrot étudia l'ensemble des points c pour lesquels 0 est dans K_c (le célèbre* ensemble de Mandelbrot*). Les ensembles de Julia ont les propriétés des fractales, en particulier l'autosimilitude. En revanche, l'ensemble de Mandelbrot est extraordinairement varié ; plus l'échelle est grande, plus l'image se complique. On a montré un caractère universel de l'ensemble de Mandelbrot : pour diverses fonctions complexes à un paramètre, on trouve des copies déformées de cet ensemble.*

Les premiers exercices sont assez généraux : étude de la monotonie, de la convergence d'une suite.

2.1. Une étude de monotonie

Étudier la monotonie de la suite de terme général

$$u_n = \sum_{k=1}^{n} (-1)^k \mathrm{C}_n^k \ln k.$$

(École polytechnique)

Solution.

Pour $n \geqslant 1$, on a

$$\begin{aligned} u_{n+1} &= \sum_{k=1}^{n+1} (-1)^k \mathrm{C}_{n+1}^k \ln k \\ &= \sum_{k=1}^{n} (-1)^k (\mathrm{C}_n^k + \mathrm{C}_n^{k-1}) \ln k + (-1)^{n+1} \ln(n+1) \\ &= u_n + \sum_{k=0}^{n} (-1)^{k+1} \mathrm{C}_n^k \ln(k+1), \end{aligned}$$

et $v_n = u_{n+1} - u_n = \sum_{k=0}^{n} (-1)^{k+1} \mathrm{C}_n^k \ln(k+1)$. Pour étudier le signe de v_n, les logarithmes nous gênent. Pour les faire disparaître, on songe à dériver, mais pour cela il faut introduire une fonction. On remplace donc 1 par x et on pose, pour $x > 0$,

$$f(x) = \sum_{k=0}^{n} (-1)^{k+1} \mathrm{C}_n^k \ln(x+k).$$

Pour tout $x > 0$, on a $f'(x) = \sum_{k=0}^{n} (-1)^{k+1} \frac{\mathrm{C}_n^k}{x+k}$. Considérons la fraction rationnelle $\mathrm{F} = \sum_{k=0}^{n} (-1)^{k+1} \frac{\mathrm{C}_n^k}{\mathrm{X}+k}$. Il existe un polynôme $\mathrm{P} \in \mathbb{R}_n[\mathrm{X}]$ tel que $\mathrm{F} = \frac{\mathrm{P}}{\mathrm{X}(\mathrm{X}+1)\dots(\mathrm{X}+n)}$. En multipliant F par $\mathrm{X}+k$ et en évaluant en $-k$, on trouve, pour $0 \leqslant k \leqslant n$,

$$\begin{aligned}(-1)^{k-1}\mathrm{C}_n^k &= \frac{\mathrm{P}(-k)}{(-k)(-k+1)\dots(-k+k-1)(-k+k+1)\dots(-k+n)}, \\ &= \frac{\mathrm{P}(-k)}{(n-k)!(-1)^k k!} = \mathrm{P}(-k)(-1)^k \frac{\mathrm{C}_n^k}{n!}\end{aligned}$$

et donc $\mathrm{P}(-k) = -n!$. Comme P est de degré $\leqslant n$ et qu'il prend la valeur $-n!$ au moins $n+1$ fois, on en déduit que $\mathrm{P} = -n!$. On a donc, pour tout $x > 0$, $f'(x) = \frac{-n!}{x(x+1)\dots(x+n)} < 0$. La fonction f décroît sur $\mathbb{R}_+^*$. Calculons sa limite en $+\infty$. Pour $x > 0$, on a

$$\begin{aligned}f(x) &= \left(\sum_{k=0}^{n} (-1)^{k-1} \mathrm{C}_n^k\right) \ln x + \sum_{k=0}^{n} (-1)^{k-1} \mathrm{C}_n^k \ln\left(1+\frac{k}{x}\right) \\ &= \sum_{k=0}^{n} (-1)^{k-1} \mathrm{C}_n^k \ln\left(1+\frac{k}{x}\right),\end{aligned}$$

puisque $\sum_{k=0}^{n} (-1)^{k-1} \mathrm{C}_n^k = (1-1)^n = 0$ pour $n \geqslant 1$. On en déduit que $\lim_{x \to +\infty} f(x) = 0$. On conclut que, pour tout $x > 0$, $f(x) > 0$. En particulier, $v_n = f(1) > 0$ et la suite $(u_n)_{n \geqslant 1}$ est donc strictement croissante. ◁

Notons qu'on a démontré que, pour $n \geqslant 1$,

$$u_{n+1} - u_n = \int_1^{+\infty} \frac{n!}{x(x+1)\dots(x+n)}\, dx.$$

2.2. Permutation des termes d'une suite

Caractériser les suites réelles $(u_n)_{n \geqslant 0}$ telles qu'existe une permutation σ pour laquelle la suite $(u_{\sigma(n)})_{n \geqslant 0}$ est monotone à partir d'un certain rang.

(École normale supérieure)

Solution.

• *Analyse.* Soit $(u_n)_{n\geqslant 0}$ une suite réelle et σ une permutation de $\mathbb{N}$ telle que la suite $(u_{\sigma(n)})_{n\geqslant 0}$ soit monotone à partir d'un certain rang. Par le théorème des limites monotones, la suite $(u_{\sigma(n)})_{n\geqslant 0}$ admet une limite ℓ dans $\overline{\mathbb{R}} = \mathbb{R}\cup\{-\infty,+\infty\}$. Montrons que la suite $(u_n)_{n\geqslant 0}$ converge également vers ℓ. Traitons le cas où $\ell\in\mathbb{R}$. Soit $\varepsilon>0$. Par hypothèse on peut trouver $\mathrm{N}\in\mathbb{N}$ tel que $|u_{\sigma(n)}-\ell|\leqslant\varepsilon$ pour $n\geqslant \mathrm{N}$. Soit alors $\mathrm{N}' = \max_{0\leqslant k\leqslant \mathrm{N}-1}\sigma(k)$. Pour tout $n>\mathrm{N}'$ on a $|u_n-\ell|\leqslant\varepsilon$ ce qui prouve le résultat. Il est facile d'adapter cette preuve au cas où ℓ est infini. Notons par ailleurs qu'à partir d'un certain rang tous les termes de la suite $(u_n)_{n\geqslant 0}$ se trouvent du même côté de la limite ℓ et même strictement si la suite n'est pas stationnaire.

• *Synthèse.* Considérons une suite $(u_n)_{n\geqslant 0}$ qui admet une limite ℓ dans $\overline{\mathbb{R}}$, qui n'est pas stationnaire (ce cas étant évident puisque $\sigma=\mathrm{Id}_{\mathbb{N}}$ convient) et telle qu'à partir d'un certain rang N_0 tous les termes de la suite sont strictement du même côté de ℓ. On supposera par exemple que $u_n<\ell$ pour $n\geqslant \mathrm{N}_0$. Traitons le cas où $\ell\in\mathbb{R}$. Posons tout d'abord $\sigma(n)=n$ pour $0\leqslant n\leqslant \mathrm{N}_0-1$. Considérons ensuite les ensembles $\Omega_0=\{n\geqslant \mathrm{N}_0,\ u_n<\ell-1\}$ et $\Omega_k=\{n\geqslant \mathrm{N}_0,\ \ell-\dfrac{1}{k}\leqslant u_n<\ell-\dfrac{1}{k+1}\}$, pour tout $k\in\mathbb{N}^*$. Par hypothèse, les ensembles Ω_k sont tous finis et forment une partition de l'intervalle $[\![\mathrm{N}_0,+\infty[\![$. Il suffit de définir $\sigma(k)$ pour $k\geqslant \mathrm{N}_0$ en prenant les éléments de Ω_0 puis ceux de Ω_1, et ainsi de suite et en s'arrangeant pour que la suite $(u_{\sigma(k)})_{k\geqslant \mathrm{N}_0}$ soit croissante (on classe par ordre croissant les éléments de chaque ensemble Ω_k). Le raisonnement s'adapte aisément au cas où $\ell=+\infty$ (on considère un rang N_0 à partir duquel les termes de la suite sont strictement positifs et pour tout $k\in\mathbb{N}$, $\Omega_k=\{n\geqslant \mathrm{N}_0,\ k<u_n\leqslant k+1\}$) et au cas où $u_n>\ell$ à partir d'un certain rang.

Conclusion. Les suites réelles $(u_n)_{n\geqslant 0}$ pour lesquelles il existe une permutation σ telle que $(u_{\sigma(n)})_{n\geqslant 0}$ soit monotone à partir d'un certain rang sont les suites stationnaires et les suites ayant une limite ℓ dans $\overline{\mathbb{R}}$ et qui sont strictement inférieures ou supérieures à ℓ à partir d'un certain rang.

L'exercice qui suit utilise simplement la définition de la convergence d'une suite. Il est à rapprocher de l'exercice 4.4 sur les fonctions sur-additives. L'idée est la même.

2.3. Suites sous-additives

Soit $u = (u_n)_{n\geqslant 0}$ une suite réelle vérifiant $u_{n+m} \leqslant u_n + u_m$ pour tout couple (m,n) d'entiers naturels. Montrer que $\left(\frac{u_n}{n}\right)_{n\geqslant 1}$ converge vers $\ell = \inf\limits_{n\geqslant 1} \frac{u_n}{n} \in \mathbb{R}\cup\{-\infty\}$.

(École normale supérieure)

▷ **Solution.**

De l'hypothèse, on déduit immédiatement que, pour tout entier n, on a $u_{2n} \leqslant 2u_n$, puis par une récurrence simple, que $u_{mn} \leqslant mu_n$, pour m et n dans $\mathbb{N}^*$.

Commençons par le cas où ℓ est réel. Pour tout $\varepsilon > 0$, on peut trouver m tel que $\ell \leqslant \frac{u_m}{m} \leqslant \ell + \varepsilon$. Si n est un entier quelconque, on en fait la division euclidienne par m et on obtient $n = qm + r$, où $0 \leqslant r < m$. On a alors $u_n \leqslant u_{mq} + u_r \leqslant qu_m + u_r$ et donc, en divisant par n,

$$\ell \leqslant \frac{u_n}{n} \leqslant \frac{q}{n}u_m + \frac{u_r}{n} \leqslant \frac{(\ell+\varepsilon)qm}{n} + \frac{\mathrm{M}}{n} \leqslant (\ell+\varepsilon)\left(1 - \frac{r}{n}\right) + \frac{\mathrm{M}}{n},$$

où $\mathrm{M} = \max\limits_{0\leqslant k<m} u_k$. Puisque $0 \leqslant r < m$, $\frac{r}{n}$ tend vers 0 lorsque n tend vers $+\infty$ et $(\ell+\varepsilon)\left(1 - \frac{r}{n}\right) + \frac{\mathrm{M}}{n}$ tend vers $\ell+\varepsilon$. On a donc pour n assez grand,

$$\ell \leqslant \frac{u_n}{n} \leqslant \ell + 2\varepsilon,$$

ce qui prouve que $\lim\limits_{n\to+\infty} \frac{u_n}{n} = \ell$.

Le cas où $\ell = -\infty$ se traite de même. Pour tout $\mathrm{A} < 0$, il existe m tel que $\frac{u_m}{m} \leqslant \mathrm{A}$. On obtient, comme précédemment

$$\frac{u_n}{n} \leqslant \mathrm{A}\left(1 - \frac{n}{r}\right) + \frac{\mathrm{M}}{n} \leqslant \frac{\mathrm{A}}{2},$$

pour n assez grand. ◁

On peut par exemple utiliser ce résultat pour montrer la convergence de la suite $\|\mathrm{A}^n\|^{1/n}$, où $\mathrm{A} \in \mathcal{M}_n(\mathbb{C})$ et où $\|\cdot\|$ désigne une norme sous-multiplicative sur $\mathcal{M}_n(\mathbb{C})$. En effet, on voit tout de suite que la suite $(\ln\|\mathrm{A}^n\|)_{n\geqslant 1}$ vérifie l'hypothèse de l'exercice. C'est alors un exercice classique de montrer que la limite de $\|\mathrm{A}^n\|^{1/n}$ est le rayon spectral de la matrice A, c'est-à-dire le plus grand module de ses valeurs propres.

Les deux exercices suivants utilisent le fait qu'une suite monotone converge si et seulement si elle est bornée.

2.4. Suites monotones

À toute suite croissante $(a_n)_{n\in\mathbb{N}}$ de réels telle que $a_0 = 1$, on associe la suite $(b_n)_{n\in\mathbb{N}}$ définie par

$$b_n = \sum_{k=1}^{n} \left(1 - \frac{a_{k-1}}{a_k}\right) \frac{1}{a_k}.$$

Montrer que $b_n \in [0,1]$. Étant donné $c \in [0,1[$, montrer l'existence d'une suite $(a_n)_{n\in\mathbb{N}}$ telle $\lim\limits_{n\to+\infty} b_n = c$.

(École polytechnique)

Solution.

La suite (a_n) étant croissante et positive, on a pour tout $k \in \mathbb{N}^*$, $\left(1 - \frac{a_{k-1}}{a_k}\right) \frac{1}{a_k} \geqslant 0$ et donc $b_n \geqslant 0$. D'autre part, on peut écrire par croissance de la suite (a_n),

$$\left(1 - \frac{a_{k-1}}{a_k}\right) \frac{1}{a_k} = \frac{a_k - a_{k-1}}{a_k^2} \leqslant \frac{a_k - a_{k-1}}{a_k a_{k-1}} \leqslant \frac{1}{a_{k-1}} - \frac{1}{a_k}.$$

En sommant ces inégalités pour $k = 1, \ldots, n$, on obtient

$$b_n \leqslant \sum_{k=1}^{n} \left(\frac{1}{a_{k-1}} - \frac{1}{a_k}\right) = \frac{1}{a_0} - \frac{1}{a_n} = 1 - \frac{1}{a_n} < 1.$$

Ainsi, la suite (b_n) qui est croissante et majorée par 1 converge.

Soit $q \in]1,+\infty[$ et la suite (a_n) définie par $a_n = q^n$. Elle a les propriétés voulues. Pour ce choix de (a_n), on obtient

$$b_n = \sum_{k=1}^{n} \left(1 - \frac{1}{q}\right) \frac{1}{q^k} = \left(1 - \frac{1}{q}\right) \left(\frac{1}{q}\right) \frac{1 - \frac{1}{q^n}}{1 - \frac{1}{q}} = \frac{1}{q} \left(1 - \frac{1}{q^n}\right).$$

On en déduit que $\lim\limits_{n\to+\infty} b_n = \frac{1}{q}$. Quand q décrit $]1,+\infty[$, $\frac{1}{q}$ décrit $]0,1[$. En particulier, pour tout $c \in]0,1[$, il existe $q = \frac{1}{c} > 1$ tel que $\lim\limits_{n\to+\infty} b_n = c$. Pour $c = 0$, la suite (a_n) constante égale à 1 convient. ◁

Outre la convergence monotone, l'exercice suivant utilise le théorème de Cesàro, dont nous rappelons l'énoncé. Si la suite (u_n) tend vers $\ell \in \overline{\mathbb{R}}$ alors la suite (v_n) de terme général $v_n = \frac{1}{n} \sum\limits_{k=1}^{n} u_k$ (moyenne de Cesàro de la suite des u_n) converge aussi vers ℓ. Ce résultat sera au centre des énoncés qui suivent.

2.5. Suites convexes bornées

On considère une suite de réels $(u_n)_{n\in\mathbb{N}}$ et on note pour $n \geqslant 0$, $\Delta u_n = u_n - u_{n+1}$ et $\Delta^2 u_n = \Delta u_n - \Delta u_{n+1}$. On suppose que la suite $(u_n)_{n\in\mathbb{N}}$ est bornée et que pour tout $n \in \mathbb{N}$, $\Delta^2 u_n \geqslant 0$.

1. Montrer que la suite $(u_n)_{n\in\mathbb{N}}$ est décroissante et que $n\Delta u_n$ converge vers 0.

2. Montrer que la série $\sum(n+1)\Delta^2 u_n$ converge.

(École polytechnique)

▷ **Solution.**

1. On note que la condition $\Delta^2 u_n \geqslant 0$ s'écrit $u_{n+1} \leqslant \frac{1}{2}(u_n + u_{n+2})$. Autrement dit, la suite (u_n) est convexe.

On a, pour tout $n \in \mathbb{N}$, $\Delta^2 u_n = \Delta u_n - \Delta u_{n+1} \geqslant 0$: la suite (Δu_n) est décroissante. De plus, elle est bornée, puisque (u_n) l'est. Elle converge donc. Si on note ℓ sa limite, alors on a, d'après le théorème de Cesàro, $\lim\limits_{n\to+\infty} \dfrac{\Delta u_0 + \Delta u_1 + \cdots + \Delta u_{n-1}}{n} = \ell$, *i.e.* $\lim\limits_{n\to+\infty} \dfrac{u_0 - u_n}{n} = \ell$. Si $\ell \neq 0$, la suite (u_n) n'est pas bornée, ce qui est contraire à l'hypothèse. On a donc $\ell = 0$. La suite (Δu_n), qui converge vers 0 en décroissant, est donc positive. On a ainsi, pour tout $n \in \mathbb{N}$, $\Delta u_n = u_n - u_{n+1} \geqslant 0$: la suite (u_n) décroît. Puisque de plus elle est bornée, elle converge. Notons L sa limite.

Pour montrer que la suite $(n\Delta u_n)$ converge, on commence par montrer que sa moyenne de Cesàro converge, ce qui est plus simple. En effet, quand on somme des termes successifs de cette suite, on obtient des simplifications puisque Δu_n est une différence entre deux termes consécutifs de la suite (u_n). On a, pour tout $n \in \mathbb{N}^*$,

$$\sum_{k=1}^{n} k\Delta u_k = \sum_{k=1}^{n} ku_k - \sum_{k=1}^{n} ku_{k+1} = \sum_{k=1}^{n} u_k - nu_{n+1} \text{ et donc}$$

$$\frac{1}{n}\sum_{k=1}^{n} k\Delta u_k = \frac{1}{n}\sum_{k=1}^{n} u_k - u_{n+1}.$$

Du théorème de Cesàro on déduit $\lim\limits_{n\to+\infty} \dfrac{1}{n}\sum\limits_{k=1}^{n} u_k = \mathrm{L}$ et donc $\lim\limits_{n\to+\infty} \dfrac{1}{n}\sum\limits_{k=1}^{n} k\Delta u_k = \mathrm{L} - \mathrm{L} = 0$. La suite (Δu_n) étant décroissante, on peut écrire, pour tout $n \in \mathbb{N}^*$,

$$\frac{1}{n}\sum_{k=1}^{n} k\Delta u_k \geqslant \frac{1}{n}\left(\sum_{k=1}^{n} k\right)\Delta u_n \geqslant \frac{n+1}{2}\Delta u_n \geqslant \frac{n}{2}\Delta u_n \geqslant 0.$$

On en déduit, par encadrement, que $\lim_{n\to+\infty} n\Delta u_n = 0$.

Le lecteur observera que cette dernière question est une formulation en termes de suites d'un résultat classique sur les séries (cf. exercice 3.10). En effet, la suite (Δu_n) est positive décroissante et la série $\sum \Delta u_n$ converge. Cela implique que $n\Delta u_n$ tend vers 0.

2. On a, pour tout $n \in \mathbb{N}$,

$$\begin{aligned}\sum_{k=0}^{n}(k+1)\Delta^2 u_k &= \sum_{k=0}^{n}(k+1)\Delta u_k - \sum_{k=1}^{n+1} k\Delta u_k \\ &= \sum_{k=0}^{n}\Delta u_k - (n+1)\Delta u_{n+1} \\ &= u_0 - u_{n+1} - (n+1)\Delta u_{n+1}.\end{aligned}$$

On sait que $\lim_{n\to+\infty} u_{n+1} = \mathrm{L}$ et que $\lim_{n\to+\infty}(n+1)\Delta u_{n+1} = 0$, d'après la question précédente. On en déduit que la série $\sum(k+1)\Delta^2 u_k$ converge et que

$$\boxed{\sum_{k=0}^{+\infty}(k+1)\Delta^2 u_k = u_0 - \mathrm{L}.} \quad \lhd$$

Une suite dont les moyennes de Cesàro convergent peut très bien diverger, comme le prouve le contre-exemple classique $u_n = (-1)^n$. L'exercice suivant montre que pour qu'une suite positive bornée $(u_n)_{n\in\mathbb{N}}$ tende vers 0 au sens de Cesàro, il faut et suffit que $(u_n)_{n\in\mathbb{N}}$ tende vers 0 sur une partie de $\mathbb{N}$ de densité 1.

2.6. Caractérisation de la convergence au sens de Cesàro

Soit $(a_n)_{n\geqslant 0}$ une suite réelle positive, majorée. Montrer qu'il y a équivalence entre

(i) $\mathrm{S}_n = \dfrac{1}{n}\sum_{k=0}^{n} a_k$ tend vers 0 ;

(ii) il existe une partie $\mathrm{A} \subset \mathbb{N}$ de densité nulle (c'est-à-dire telle que $\lim_{n\to+\infty} \dfrac{\mathrm{Card}(\mathrm{A}\cap[\![0,n]\!])}{n} = 0$) telle que $\lim_{\substack{n\to+\infty \\ n\notin \mathrm{A}}} a_n = 0$.

(École polytechnique)

Solution.

• Supposons d'abord (ii) réalisé. Soit $\varepsilon > 0$. Comme $\lim_{\substack{n\to+\infty \\ n\notin \mathrm{A}}} a_n = 0$, il existe $\mathrm{N} \in \mathbb{N}^*$ tel que pour $n \geqslant \mathrm{N}$ et $n \notin \mathrm{A}$, $a_n \leqslant \varepsilon$. Pour $n \geqslant \mathrm{N}$, on

coupe alors la moyenne de Cesàro en trois :

$$S_n = \frac{1}{n}\sum_{k=0}^{n} a_k = \frac{1}{n}\sum_{k=0}^{N-1} a_k + \frac{1}{n}\sum_{\substack{N\leqslant k\leqslant n\\ k\notin A}} a_k + \frac{1}{n}\sum_{\substack{N\leqslant k\leqslant n\\ k\in A}} a_k.$$

La somme du milieu est majorée par ε. Celle de gauche tend clairement vers 0 lorsque n tend vers l'infini. Enfin, si on note $M = \sup_{k\in\mathbb{N}} a_k$ (qui existe puisque la suite est majorée), celle de droite est inférieure à $\dfrac{M\,\mathrm{Card}(A\cap[\![N,n]\!])}{n}$ et *a fortiori* à $\dfrac{M\,\mathrm{Card}(A\cap[\![0,n]\!])}{n}$, quantité qui tend également vers 0 lorsque n tend vers l'infini. Il existe donc $N' \geqslant N$ tel que pour tout n supérieur à N', on ait $S_n \leqslant 2\varepsilon$. L'assertion (i) est démontrée.

• Supposons maintenant que (i) est vérifié et que la suite (a_n) n'est pas de limite nulle (sinon $A = \emptyset$ convient). Posons $\alpha_n = \sup\{S_p,\ p \geqslant n\}$. La suite (α_n) est une suite décroissante tendant vers 0 de réels strictement positifs. Considérons $A = \{p \in \mathbb{N},\ a_p \geqslant \sqrt{\alpha_p}\}$. Il est clair que $\lim_{n\notin A} a_n = 0$ (car si $n \notin A$, $a_n < \sqrt{\alpha_n}$). Il ne reste plus qu'à prouver que A est de densité nulle. On a

$$\mathrm{Card}(A\cap[\![0,n]\!]) = \sum_{k\in A\cap[\![0,n]\!]} 1 \leqslant \sum_{k\in A\cap[\![0,n]\!]} \frac{a_k}{\sqrt{\alpha_k}} \leqslant \sum_{k=0}^{n} \frac{a_k}{\sqrt{\alpha_k}},$$

de sorte que

$$\frac{1}{n}\,\mathrm{Card}(A\cap[\![0,n]\!]) \leqslant \frac{1}{n}\sum_{k=0}^{n} \frac{a_k}{\sqrt{\alpha_k}} \leqslant \frac{1}{n}\sum_{k=0}^{n} \frac{a_k}{\sqrt{\alpha_n}} = \frac{S_n}{\sqrt{\alpha_n}} \leqslant \sqrt{\alpha_n},$$

car $S_n \leqslant \alpha_n$. D'où le résultat. ◁

L'exercice qui suit porte aussi sur des suites qui convergent au sens de Cesàro, mais cette fois dans $\mathbb{C}$. Notons que le théorème de Cesàro subsiste dans $\mathbb{C}$.

2.7. Suites convergentes au sens de Cesàro telles que $u_n^6 \to 1$

Soit E l'ensemble des suites (u_n) de nombres complexes vérifiant $\lim_{n\to+\infty} u_n^6 = 1$ et telles que la suite de terme général $v_n = \dfrac{1}{n}\sum_{k=1}^{n} u_k$ converge. On considère l'application φ de E dans $\mathbb{C}$ qui à la suite (u_n) associe la limite ℓ de la suite (v_n). Déterminer $\varphi(E)$.

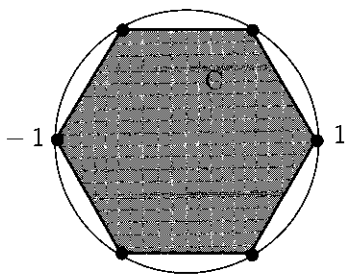
(École polytechnique)

Solution.

Notons U_6 l'ensemble des racines 6-ièmes de 1 et posons, pour $1 \leqslant k \leqslant 6$, $\omega_k = e^{\frac{ik\pi}{3}}$. La façon la plus simple de réaliser la condition $\lim_{n\to+\infty} u_n^6 = 1$ est que la suite (u_n^6) soit constante égale à 1, c'est-à-dire que, pour tout $n \in \mathbb{N}^*$, u_n soit élément de U_6. Notons F le sous-ensemble de E formé des suites ayant cette propriété et déterminons pour commencer $\varphi(F)$. Si (u_n) est dans F, on obtient, en notant, pour $n \in \mathbb{N}^*$ et $1 \leqslant i \leqslant 6$, $m_i(n)$ le nombre de termes u_k $(1 \leqslant k \leqslant n)$ de la suite (u_n) égaux à ω_i,

$$v_n = \frac{1}{n}\sum_{i=1}^{6} m_i(n)\omega_i = \sum_{i=1}^{n} \frac{m_i(n)}{n}\omega_i.$$

On en déduit que chaque terme de la suite (v_n) s'écrit comme barycentre à coefficients positifs de $(\omega_1, \ldots, \omega_6)$, c'est-à-dire appartient à l'enveloppe convexe C de U_6. L'ensemble C étant fermé, il en résulte que la limite ℓ de la suite (v_n) appartient à C.



Démontrons réciproquement que $C \subset \varphi(F)$. Soit ℓ un élément de C qui s'écrit $\ell = \sum_{i=1}^{6} \lambda_i\omega_i$, où $\lambda_1, \ldots, \lambda_6$ sont des réels positifs de somme égale à 1. Montrons qu'on peut trouver une suite (u_n) de F telle que, pour $1 \leqslant i \leqslant 6$, on ait, avec les notations précédentes, $\lim_{n\to+\infty} \frac{m_i(n)}{n} = \lambda_i$. Posons $\alpha_i(n) = E(\lambda_i n)$ si $1 \leqslant i \leqslant 5$ et $\alpha_6(n) = n - \sum_{i=1}^{5} \alpha_i(n)$ pour tout $n \in \mathbb{N}^*$. L'entier $\alpha_6(n)$ est positif, car

$$\sum_{i=1}^{5} \alpha_i(n) \leqslant \sum_{i=1}^{5} \lambda_i n \leqslant n.$$

Il est clair que $\lim_{n\to+\infty} \frac{\alpha_i(n)}{n} = \lim_{n\to+\infty} \frac{E(\lambda_i n)}{n} = \lambda_i$ pour $1 \leqslant i \leqslant 5$. Ceci entraîne $\lim_{n\to+\infty} \frac{\alpha_6(n)}{n} = 1 - \sum_{i=1}^{5} \lambda_i = \lambda_6$. On va choisir (u_n) pour

que $m_i(n) = \alpha_i(n)$, pour $1 \leqslant i \leqslant 6$ et $n \in \mathbb{N}^*$. On a, par construction, pour $1 \leqslant i \leqslant 5$ et $n \geqslant 2$,

$$\lambda_i(n-1) - 1 < \alpha_i(n-1) \leqslant \lambda_i(n-1) \quad \text{et} \quad \lambda_i n - 1 < \alpha_i(n) \leqslant \lambda_i n.$$

On en déduit que

$$-1 \leqslant -1 + \lambda_i < \alpha_i(n) - \alpha_i(n-1) < \lambda_i + 1 \leqslant 2$$

et donc que $\alpha_i(n) = \alpha_i(n-1)$ ou $\alpha_i(n-1) + 1$. Puisque $\sum_{i=1}^{6} \alpha_i(n) = n$, pour tout $n \in \mathbb{N}^*$, on a, pour tout $n \geqslant 2$, $\alpha_i(n) - \alpha_i(n-1) = 1$ pour l'un des entiers i ($1 \leqslant i \leqslant 6$) et 0 pour les autres. On note $f(n)$ cet entier i et $f(1)$ l'entier i tel que $\alpha_i(1) \neq 0$. On définit ensuite la suite (u_n) par $u_n = \omega_{f(n)}$. On a, par construction, $m_i(1) = \alpha_i(1)$ et, pour $n \geqslant 2$, $\alpha_i(n) - \alpha_i(n-1) = m_i(n) - m_i(n-1)$ et donc pour tout $n \in \mathbb{N}^*$, $m_i(n) = \alpha_i(n)$. On en déduit que $v_n = \sum_{i=1}^{6} \frac{\alpha_i(n)}{n} \omega_i$. Finalement on obtient que la suite (v_n) converge et que sa limite est $\sum_{i=1}^{6} \lambda_i \omega_i = \ell$. Autrement dit, la suite (u_n) appartient à F et sa limite ℓ appartient à $\varphi(\mathrm{F})$. On a bien $\varphi(\mathrm{F}) = \mathrm{C}$.

Nous allons enfin montrer que $\varphi(\mathrm{E}) = \varphi(\mathrm{F}) = \mathrm{C}$. Considérons une suite (u_n) de E. Pour n assez grand, tout terme de la suite (u_n) est proche d'un élément de U_6. En effet, la suite (u_n^6) converge vers 1. Pour tout $\varepsilon \in]0,1[$, il existe n_0 tel que $|u_n^6 - 1| \leqslant \varepsilon^6$ si $n \geqslant n_0$. Puisque $|u_n^6 - 1| = \prod_{i=1}^{6} |u_n - \omega_i|$, on en déduit qu'il existe $i \in [\![1,6]\!]$ tel que $|u_n - \omega_i| \leqslant \varepsilon$. Les images des ω_i dans le plan complexe formant un hexagone, la distance entre deux sommets est $\geqslant 1$. Si $\varepsilon \leqslant \frac{1}{3}$, les disques de centre ω_i et de rayon ε sont donc disjoints. En particulier, un entier i tel que $|u_n - \omega_i| \leqslant \varepsilon$ est unique. Prenons $\varepsilon = \frac{1}{3}$ et pour $n \geqslant n_0$, notons $f(n)$ un tel entier i, et prolongeons f en une application de $\mathbb{N}^*$ dans $[\![1,6]\!]$, en posant $f(n) = 1$ si $n < n_0$. Considérons la suite (u'_n) de terme général $\omega_{f(n)}$ et (v'_n) la moyenne de Cesàro associée. Pour $\varepsilon \in \left]0, \frac{1}{3}\right[$, il existe un entier $n_1 \geqslant n_0$ tel que, pour $n \geqslant n_1$, il existe un entier i tel que $|u_n - \omega_i| \leqslant \varepsilon$. Nous avons montré précédemment que nécessairement, $\omega_i = u'_n$. On a alors, pour $n \geqslant n_1$,

$$\begin{aligned} |v_n - v'_n| &\leqslant \frac{1}{n} \sum_{k=1}^{n} |u_n - u'_n| \leqslant \frac{1}{n} \sum_{k=1}^{n_1-1} |u_n - u'_n| + \varepsilon \frac{n - n_1}{n} \\ &\leqslant \frac{1}{n} \sum_{k=1}^{n_1-1} |u_n - u'_n| + \varepsilon. \end{aligned}$$

On aura donc, pour n assez grand, $|v_n - v'_n| \leqslant 2\varepsilon$. On en déduit que la suite $(v_n - v'_n)$ converge vers 0. La suite (v_n) convergeant vers $\varphi((u_n))$, il en est de même pour la suite (v'_n). Ceci démontre que (u'_n) appartient à F et que $\varphi((u_n)) = \varphi((u'_n))$. On a bien $\varphi(\mathrm{E}) = \varphi(\mathrm{F})$.

Conclusion. L'ensemble $\varphi(\mathrm{E})$ est l'enveloppe convexe de l'ensemble des racines 6-ièmes de l'unité, c'est-à-dire l'intérieur d'un hexagone régulier. ◁

Il est clair que le nombre 6 ne joue aucun rôle particulier et que si on considère les suites (u_n) telles que (u_n^p) converge vers 1, on trouve que $\varphi(\mathrm{E})$ est l'intérieur d'un polygone régulier à p côtés.

Dans l'exercice suivant il est encore question de suites dont la moyenne de Cesàro converge, cette fois avec des systèmes dynamiques discrets. La seconde question de l'exercice fournit des exemples de suites divergentes dont la moyenne de Cesàro converge.

2.8. Moyenne de Cesàro d'une suite récurrente

Soit $f : \mathbb{R} \to \mathbb{R}$ continue. Pour $a \in \mathbb{R}$, on définit $v_0(a) = a$ et pour tout $n \in \mathbb{N}$, $v_{n+1}(a) = f(v_n(a))$. On pose

$$u_n(a) = \frac{v_0(a) + v_1(a) + \cdots + v_n(a)}{n+1}.$$

1. On suppose qu'il existe $a \in \mathbb{R}$ tel que $(u_n(a))_{n\in\mathbb{N}}$ soit bornée. Montrer que f admet un point fixe.

2. Trouver un exemple de fonction continue $f : \mathbb{R} \to \mathbb{R}$ ayant un point fixe unique a et telle que pour tout $x \neq a$, $(u_n(x))_{n\in\mathbb{N}}$ converge vers une limite distincte de a.

(**École polytechnique**)

Solution.

1. Supposons par l'absurde que f n'ait pas de point fixe. Comme elle est continue, on a soit $f(x) > x$ pour tout x, soit $f(x) < x$ pour tout x. Plaçons-nous par exemple dans le premier cas. La suite $(v_n(a))$ est strictement croissante, et diverge nécessairement vers $+\infty$ (car si elle convergeait, sa limite serait un point fixe de f). Le théorème de Cesàro nous permet d'affirmer que la suite $(u_n(a))$ tend alors aussi vers $+\infty$ et ne peut donc être bornée, ce qui fournit la contradiction souhaitée.

2. Considérons la fonction f définie par $f(x) = -2x$ si $x \leqslant 0$ et $f(x) = -\dfrac{x}{2}$ si $x \geqslant 0$. Elle est continue sur $\mathbb{R}$ et son seul point fixe est 0. La fonction $f \circ f$ est l'identité de sorte que pour tout $x > 0$ la

suite $(v_n(x))$ vérifie $v_{2k}(x) = x$ et $v_{2k+1}(x) = -\dfrac{x}{2}$ pour tout entier naturel k. Il est alors aisé de voir que la transformée de Cesàro $(u_n(x))$ converge vers la moyenne de ces deux valeurs qui vaut $\dfrac{x}{4}\cdot$ De même, si $x < 0$ la suite $(u_n(x))$ converge vers $-\dfrac{x}{2}\cdot$ Cette fonction répond donc à la question. ◁

Plus généralement, toute fonction involutive continue, ayant 0 *comme seul point fixe et telle que* $f(x) \neq -x$ *pour* $x \neq 0$, *convient.*

Le théorème de Hardy énonce des conditions pour que la convergence de la moyenne de Cesàro entraîne la convergence de la suite initiale. Autrement dit, il étudie la réciproque du théorème de Cesàro. On peut le formuler en termes de séries : les suites dont on étudie ici la moyenne de Cesàro sont des sommes partielles de séries.

2.9. Théorème taubérien de Hardy

Soit $(u_n)_{n\geqslant 1}$ une suite de $\mathbb{C}$. On pose, pour $n \geqslant 1$,

$$\mathrm{S}_n = \sum_{k=1}^{n} u_k \quad \text{et} \quad \sigma_n = \frac{1}{n}\sum_{k=1}^{n} \mathrm{S}_k$$

1. On suppose que les u_n sont des réels positifs. Montrer qu'il y a équivalence entre :

(i) la suite $(\mathrm{S}_n)_{n\geqslant 1}$ converge ;

(ii) la suite $(\sigma_n)_{n\geqslant 1}$ converge.

2. On revient au cas général. Montrer que $u_n = O\left(\dfrac{1}{n}\right)$ est une condition suffisante pour que la convergence de la suite $(\sigma_n)_{n\geqslant 1}$ entraîne celle de la suite $(\mathrm{S}_n)_{n\geqslant 1}$.

(École polytechnique)

▷ **Solution.**

1. Si $(\mathrm{S}_n)_{n\geqslant 1}$ converge vers un réel ℓ alors, d'après le théorème de Cesàro, les moyennes σ_n convergent vers ℓ. Si (S_n) ne converge pas, elle diverge vers $+\infty$ puisqu'elle est croissante. Toujours d'après le théorème de Cesàro, $\lim\limits_{n\to+\infty} \sigma_n = +\infty$ et (σ_n) ne converge pas. L'équivalence est prouvée.

2. Supposons que $u_n = O\left(\dfrac{1}{n}\right)$ et que (σ_n) converge vers $\ell \in \mathbb{C}$. Nous contrôlons la moyenne des S_n. Il faut donc relier celle-ci à S_n. Afin d'y parvenir, on va comparer S_n à une moyenne des S_k pour k compris entre n et nx, où x est un réel strictement plus grand que 1. On pourra

encadrer cette moyenne, grâce à l'hypothèse de convergence de (σ_n). On imagine que pour avoir une expression proche de S_n, il faudra peu de termes dans la moyenne *i.e.* x proche de 1.

Posons pour $u > 0$

$$\sigma_u = \frac{1}{u} \sum_{1 \leqslant k \leqslant u} S_k = \frac{E(u)}{u} \sigma_{E(u)},$$

où E désigne la fonction partie entière. On a $\lim_{u \to +\infty} \sigma_u = \ell$. Pour $x > 1$, appelons $M_n(x)$ la moyenne des S_k pour $n < k \leqslant nx$

$$M_n(x) = \frac{1}{E(nx - n)} \sum_{n < k \leqslant nx} S_k$$

et étudions la limite de $M_n(x)$. On obtient

$$M_n(x) = \frac{nx\sigma_{nx} - n\sigma_n}{E(nx - n)} = \frac{n(x-1)}{E(n(x-1))} \frac{x\sigma_{nx} - \sigma_n}{(x-1)} \xrightarrow[n \to +\infty]{} \frac{\ell x - \ell}{x - 1} = \ell.$$

Estimons maintenant la différence entre la moyenne $M_n(x)$ et S_n. Comme on peut écrire $S_n = \frac{\sum_{n < k \leqslant nx} S_n}{E(n(x-1))}$, il vient

$$\begin{aligned} |S_n - M_n(x)| &= \left| \frac{1}{E(n(x-1))} \sum_{n < k \leqslant nx} (S_k - S_n) \right| \\ &\leqslant \frac{1}{E(n(x-1))} \sum_{n < k \leqslant nx} |S_k - S_n|. \end{aligned}$$

Par hypothèse, il existe $A \geqslant 0$ tel que si $n \geqslant 1$, $|u_n| \leqslant \frac{A}{n}$. Par conséquent si $k > n$, on a

$$\begin{aligned} |S_k - S_n| = |u_k + u_{k-1} + \cdots + u_{n+1}| &\leqslant A\left(\frac{1}{k} + \frac{1}{k-1} + \cdots + \frac{1}{n+1}\right) \\ &\leqslant A\left(\int_{k-1}^{k} \frac{dt}{t} + \int_{k-2}^{k-1} \frac{dt}{t} + \cdots + \int_{n}^{n+1} \frac{dt}{t}\right) = A \int_n^k \frac{dt}{t} \\ &\leqslant A \ln \frac{k}{n}. \end{aligned}$$

Il en résulte que

$$|S_n - M_n(x)| \leqslant \frac{A}{E(n(x-1))} \sum_{n < k \leqslant nx} \ln \frac{k}{n} \leqslant \frac{A}{E(n(x-1))} \sum_{n < k \leqslant nx} \ln \frac{nx}{n},$$

ce qui donne après simplifications

$$|S_n - M_n(x)| \leqslant \frac{A}{E(n(x-1))}\, E(n(x-1)) \ln x = A \ln x,$$

où le majorant de dépend pas de n. Donnons-nous $\varepsilon > 0$. Il existe $x_0 > 1$ tel que $A \ln x_0 \leqslant \varepsilon$. On a alors

$$|S_n - \ell| \leqslant |S_n - M_n(x_0)| + |M_n(x_0) - \ell| \leqslant \varepsilon + |M_n(x_0) - \ell| \leqslant 2\varepsilon$$

pour n assez grand, puisque $\lim\limits_{n\to+\infty} M_n(x_0) = \ell$. On conclut que

$$\boxed{\lim_{n\to+\infty} S_n = \ell.} \quad \triangleleft$$

Dans l'exercice suivant, il s'agit encore de comparer un terme d'une suite à une moyenne de termes d'indices consécutifs, en utilisant cette fois la monotonie de la suite.

2.10. Limites de tranches de Cauchy

Soit $(a_n)_{n\in\mathbb{N}}$ une suite décroissante de réels positifs. Pour $u \geqslant 0$ et $x > 1$, on pose $f_u(x) = \sum\limits_{u<n\leqslant ux} a_n$. Montrer l'équivalence des deux conditions suivantes :

(i) pour tout $x > 1$, $f_u(x)$ a une limite finie lorsque $u \to +\infty$;

(ii) la suite $(na_n)_{n\geqslant 0}$ converge.

(**École polytechnique**)

▷ **Solution.**

Remarquons pour commencer que la suite $(a_n)_{n\in\mathbb{N}}$ admet une limite $L \geqslant 0$ puisqu'elle est décroissante et minorée par 0. Si $L > 0$, (ii) n'est pas vérifiée. Par ailleurs, si N est un entier naturel tel que $a_n \geqslant \dfrac{L}{2}$ pour $n \geqslant N$, on a, pour $x > 1$ et $u \geqslant N$,

$$f_u(x) \geqslant \frac{L}{2}(ux - u) = \frac{uL(x-1)}{2}$$

et donc $\lim\limits_{u\to+\infty} f_u(x) = +\infty$. Ainsi, ni ($i$), ni ($ii$) ne sont vérifiées, ce qui montre l'équivalence dans le cas $L > 0$. Nous supposerons donc dans la suite que $(a_n)_{n\in\mathbb{N}}$ converge vers 0.

• Supposons que la suite (na_n) converge et notons ℓ sa limite. Fixons $x > 1$. Soit $\varepsilon > 0$. Il existe un rang $n_0 \in \mathbb{N}$ tel que $\ell - \varepsilon \leqslant na_n \leqslant \ell + \varepsilon$

pour $n \geqslant n_0$. Par conséquent, si $u \geqslant n_0$, on a

$$(\ell - \varepsilon) \sum_{n=\mathrm{E}(u)+1}^{\mathrm{E}(ux)} \frac{1}{n} \leqslant f_u(x) \leqslant (\ell + \varepsilon) \sum_{n=\mathrm{E}(u)+1}^{\mathrm{E}(ux)} \frac{1}{n}.$$

De $\displaystyle\sum_{n=\mathrm{E}(u)+1}^{\mathrm{E}(ux)} \frac{1}{n} \underset{u\to\infty}{\sim} \ln\left(\frac{\mathrm{E}(ux)}{\mathrm{E}(u)}\right) \xrightarrow[u\to+\infty]{} \ln x$, on déduit que, pour u assez grand, on a $(\ell - 2\varepsilon)\ln x \leqslant f_u(x) \leqslant (\ell + 2\varepsilon)\ln x$. Ce qui traduit le fait que $\lim\limits_{u\to+\infty} f_u(x) = \ell \ln x$.

• Supposons (i) et notons pour $x > 1$, $\ell(x) = \lim\limits_{u\to+\infty} f_u(x)$. Soit $\varepsilon > 0$. Comme $(a_n)_{n\geqslant 0}$ est décroissante, on peut écrire, pour $\mathrm{N} \geqslant 1$ et $x > 1$,

$$f_{\mathrm{N}}(x) = \sum_{\mathrm{N}<x\leqslant \mathrm{N}x} a_n \leqslant (\mathrm{E}(\mathrm{N}x) - \mathrm{N})a_{\mathrm{N}}$$

$$\text{et} \quad f_{\mathrm{N}/x}(x) = \sum_{\mathrm{N}/x<n\leqslant \mathrm{N}} a_n \geqslant (\mathrm{N} - \mathrm{E}\,(\mathrm{N}/x))a_{\mathrm{N}}.$$

On en déduit l'encadrement

$$\frac{f_{\mathrm{N}}(x)}{\dfrac{\mathrm{E}(\mathrm{N}x)}{\mathrm{N}} - 1} \leqslant \mathrm{N}a_{\mathrm{N}} \leqslant \frac{f_{\mathrm{N}/x}(x)}{1 - \dfrac{\mathrm{E}(\mathrm{N}/x)}{\mathrm{N}}}.$$

On a $\lim\limits_{\mathrm{N}\to+\infty} \dfrac{f_{\mathrm{N}}(x)}{\dfrac{\mathrm{E}(\mathrm{N}x)}{\mathrm{N}} - 1} = \dfrac{\ell(x)}{x-1}$ et $\lim\limits_{\mathrm{N}\to+\infty} \dfrac{f_{\mathrm{N}/x}(x)}{1 - \dfrac{\mathrm{E}(\mathrm{N}/x)}{\mathrm{N}}} = \dfrac{x\ell(x)}{x-1}$ $(*)$.

Supposons que $\lim\limits_{x\to 1^+} \dfrac{\ell(x)}{x-1}$ existe et vaille L. Alors, pour x assez proche de 1, on aura $\dfrac{\ell(x)}{x-1}$ et $\dfrac{x\ell(x)}{x-1}$ dans l'intervalle $[\mathrm{L} - \varepsilon, \mathrm{L} + \varepsilon]$. Des limites $(*)$ et de l'encadrement de $\mathrm{N}a_{\mathrm{N}}$ trouvé, il résulte que pour N assez grand, $\mathrm{N}a_{\mathrm{N}}$ est dans l'intervalle $[\mathrm{L} - 2\varepsilon, \mathrm{L} + 2\varepsilon]$. Autrement dit, on a $\lim\limits_{\mathrm{N}\to+\infty} \mathrm{N}a_{\mathrm{N}} = \mathrm{L}$.

Il reste donc à démontrer que $\lim\limits_{x\to 1^+} \dfrac{\ell(x)}{x-1}$ existe. La démonstration dans l'autre sens incite à penser que la fonction ℓ est un multiple de la fonction ln. La fonction ℓ est définie, positive et croissante sur $]1, +\infty[$. Pour $(x, y) \in]1, +\infty[^2$ et $u > 0$, on a

$$f_u(xy) = \sum_{u<n\leqslant uxy} a_n = \sum_{u<n\leqslant nu} a_n + \sum_{ux<n\leqslant uxy} a_n = f_u(x) + f_{ux}(y).$$

En faisant tendre u vers $+\infty$, il vient

$$\ell(xy) = \ell(x) + \ell(y).$$

C'est un résultat classique que les morphismes croissants du groupe $(\mathbb{R}, +)$ dans lui-même sont les applications $x \longmapsto ax$. On en déduit immédiatement, en composant par la fonction ln que les morphismes croissants de $(\mathbb{R}_+^*, \times)$ dans $(\mathbb{R}, +)$ sont les applications $x \longmapsto \mathrm{K} \ln x$.

Ici, ℓ n'est définie que sur $]1, +\infty[$. On la prolonge à $]0, +\infty[$ en posant

$$\begin{cases} \ell(x) = -\ell\left(\dfrac{1}{x}\right) & \text{si} \quad x \in]0, 1[\\ \ell(1) = 0. \end{cases}$$

On vérifie facilement qu'on obtient ainsi un morphisme croissant de $(\mathbb{R}_+^*, \times)$ dans $(\mathbb{R}, +)$. Il existe donc $\mathrm{K} \geqslant 0$ tel que pour tout $x > 0$, on ait $\ell(x) = \mathrm{K} \ln x$. Il s'ensuit que $\lim\limits_{x \to 1^+} \dfrac{\ell(x)}{x - 1} = \mathrm{K}$. On a donc prouvé l'implication. $\lhd$

Sont regroupés ci-après plusieurs exercices utilisant la notion de valeur d'adhérence. *Rappelons que, le corps* K *étant* $\mathbb{R}$ *ou* $\mathbb{C}$, *l'élément* ℓ *de* K *est valeur d'adhérence de la suite* $(u_n) \in \mathrm{K}^{\mathbb{N}}$ *s'il existe une sous-suite de* (u_n) *convergeant vers* ℓ *; autrement dit, s'il existe une application strictement croissante* φ *de* $\mathbb{N}$ *dans* $\mathbb{N}$ *(appelée* extraction*) telle que* $\lim\limits_{n \to +\infty} u_{\varphi(n)} = \ell$. *Il est parfois plus simple d'utiliser la caractérisation suivante : l est valeur d'adhérence de la suite* (u_n) *si, pour tout* $\varepsilon > 0$ *et tout* $\mathrm{N} \in \mathbb{N}$, *il existe* $n \geqslant \mathrm{N}$ *tel que* $|u_n - \ell| \leqslant \varepsilon$. *Autrement dit, dans tout voisinage de* ℓ, *il existe une infinité de termes de la suite.*

On peut démontrer que l'ensemble X_u *des valeurs d'adhérence de la suite* u *est égal à* $\bigcap\limits_{\mathrm{N} \in \mathbb{N}} \overline{\mathrm{X_N}}$, *où* $\mathrm{X_N} = \{u_p,\ p \geqslant \mathrm{N}\}$ *pour* $\mathrm{N} \in \mathbb{N}$. *Cela résulte des équivalences suivantes, où* $\mathcal{V}(\ell)$ *désigne l'ensemble des voisinages de* ℓ *:*

$$\begin{aligned} \ell \in \mathrm{X}_u &\iff \forall \mathrm{N} \in \mathbb{N}\ \ \forall \varepsilon > 0\ \ \exists n \geqslant \mathrm{N}\ \ |u_n - \ell| \leqslant \varepsilon \\ &\iff \forall \mathrm{N} \in \mathbb{N}\ \ \forall \mathrm{V} \in \mathcal{V}(\ell)\ \ \mathrm{V} \cap \mathrm{X_N} \neq \emptyset \\ &\iff \forall \mathrm{N} \in \mathbb{N}\ \ \ell \in \overline{\mathrm{X_N}}. \end{aligned}$$

Les ensembles $\overline{\mathrm{X_N}}$ *étant fermés, il en de même de* X_u. *Si, de plus, la suite* (u_n) *est bornée, les ensembles* $\overline{\mathrm{X_N}}$ *sont bornés et il en de même de* X_u *qui est donc compact.*

Le théorème de Bolzano-Weierstrass affirme que toute suite bornée de K *possède une valeur d'adhérence. L'ensemble des valeurs d'adhérence d'une suite bornée est donc un compact non vide. Le premier exercice donne un exemple non trivial.*

2.11. Valeurs d'adhérence d'une suite complexe

On pose $z_n = \prod_{k=1}^{n} \left(1 + \frac{i}{k}\right)$. Déterminer l'ensemble des valeurs d'adhérence de la suite $(z_n)_{n\in\mathbb{N}}$.

(École polytechnique)

Solution.

Notons, pour $n \in \mathbb{N}^*$, ρ_n le module de z_n. On a

$$\rho_n = \prod_{k=1}^{n} \sqrt{1 + \frac{1}{k^2}} = \exp\left(\frac{1}{2}\sum_{k=1}^{n} \ln\left(1 + \frac{1}{k^2}\right)\right)$$

et comme $\ln\left(1 + \frac{1}{k^2}\right) \sim \frac{1}{k^2}$ pour k tendant vers l'infini, le théorème de comparaison des séries à termes positifs assure la convergence de la série $\sum \ln\left(1 + \frac{1}{k^2}\right)$. Par continuité de l'exponentielle, on obtient

$$\lim_{n\to+\infty} \rho_n = \mathrm{R} = \exp\left(\frac{1}{2}\sum_{k=1}^{+\infty} \ln\left(1 + \frac{1}{k^2}\right)\right).$$

Notons pour $k \geqslant 1$, $\alpha_k = \arctan\left(\frac{1}{k}\right) \equiv \arg\left(1 + \frac{i}{k}\right) \pmod{2\pi}$. On a alors

$$z_n = \rho_n e^{i(\alpha_1+\cdots+\alpha_n)} = \rho_n e^{i\mathrm{S}_n},$$

où $\mathrm{S}_n = \alpha_1 + \cdots + \alpha_n$.

Les valeurs d'adhérence de la suite $(e^{i\mathrm{S}_n})$ sont de module 1. Montrons qu'on obtient le cercle unité U tout entier. Considérons $\alpha \in \mathbb{R}$ et montrons que $e^{i\alpha}$ est une valeur d'adhérence de $(e^{i\mathrm{S}_n})$. Il faut montrer pour cela que, pour tout $\varepsilon > 0$ et tout entier $k_0 \in \mathbb{N}$, il existe un entier naturel $n \geqslant k_0$ tel que $|e^{i\alpha} - e^{i\mathrm{S}_n}| \leqslant \varepsilon$.

Comme $\alpha_k \underset{k\to\infty}{\sim} \frac{1}{k}$, il existe $k_1 \geqslant k_0$ tel que $0 \leqslant \alpha_k \leqslant \varepsilon$ si $k \geqslant k_1$. Soit $\mathrm{N} \in \mathbb{N}$ tel que $2\mathrm{N}\pi + \alpha \geqslant \mathrm{S}_{k_1}$. D'après le théorème de comparaison des séries à termes positifs, S_n diverge vers $+\infty$. Il existe donc $n \geqslant k_1$ tel que $\mathrm{S}_n > 2\mathrm{N}\pi + \alpha$. Considérons n_0, le plus petit des entiers $n \geqslant k_1$ vérifiant $\mathrm{S}_n > 2\mathrm{N}\pi + \alpha$. Dans ces conditions, on a $\mathrm{S}_{n_0-1} \leqslant 2\mathrm{N}\pi + \alpha < \mathrm{S}_{n_0}$. Comme $0 \leqslant \mathrm{S}_{n_0} - \mathrm{S}_{n_0-1} = \alpha_{n_0} \leqslant \varepsilon$, on en déduit que $|\mathrm{S}_{n_0} - \alpha - 2\mathrm{N}\pi| \leqslant \varepsilon$ et, d'après l'inégalité des accroissements finis,

$$\left|e^{i\mathrm{S}_{n_0}} - e^{i\alpha}\right| = \left|e^{i\mathrm{S}_{n_0}} - e^{i(\alpha+2\mathrm{N}\pi)}\right| \leqslant |\mathrm{S}_{n_0} - \alpha - 2\mathrm{N}\pi| \leqslant \varepsilon.$$

L'ensemble des valeurs d'adhérence de $(e^{i\mathrm{S}_n})_{n\geqslant 1}$ est donc exactement U.

Conclusion. L'ensemble des valeurs d'adhérence de la suite $(z_n)_{n\geqslant 1}$ est le cercle de centre 0 et de rayon R. ◁

On peut démontrer plus généralement que si $\lim_{n\to+\infty} S_n = +\infty$ *et* $\lim_{n\to+\infty} S_{n+1} - S_n = 0$, *l'ensemble des valeurs d'adhérence de la suite* (e^{iS_n}) *est le cercle unité.*

Il est important de savoir qu'une suite bornée converge si et seulement si elle ne possède qu'une seule valeur d'adhérence. Il est clair que la condition est nécessaire. Pour montrer qu'elle est suffisante, on raisonne par l'absurde, en supposant que la suite bornée (u_n) *ne converge pas vers son unique valeur d'adhérence* ℓ. *Il existe donc un réel* $\varepsilon > 0$ *tel que l'on ait* $|u_n - \ell| \geqslant \varepsilon$ *pour une infinité d'indices. Autrement dit, on peut extraire de* (u_n) *une sous-suite* $(u_{\varphi(n)})$ *telle* $|u_{\varphi(n)} - \ell| \geqslant \varepsilon$, *pour tout n. La suite* $(u_{\varphi(n)})$, *qui est bornée, possède une valeur d'adhérence* ℓ' *qui vérifie* $|\ell' - \ell| \geqslant \varepsilon$ *et donc* $\ell' \neq \ell$. *C'est impossible car* ℓ' *est aussi une valeur d'adhérence de la suite* (u_n). *Les deux exercices suivants utilisent ce résultat qu'il faudrait bien entendu redémontrer lors de l'oral.*

2.12. Question de convergence

Soit $(u_n)_{n\geqslant 0}$ une suite réelle bornée telle que $u_n + \dfrac{u_{2n}}{2}$ converge. Montrer que $(u_n)_{n\geqslant 0}$ converge.

(**École normale supérieure, école polytechnique**)

▷ **Solution.**

Il nous suffit de prouver que la suite $(u_n)_{n\geqslant 0}$ admet une unique valeur d'adhérence. Soit a une telle valeur (la suite en a au moins une d'après le théorème de Bolzano-Weierstarss) et $\varphi : \mathbb{N} \to \mathbb{N}$ strictement croissante telle que $u_{\varphi(n)}$ tende vers a. Notons par ailleurs ℓ la limite de la suite $v_n = u_n + \dfrac{u_{2n}}{2}$. En considérant la suite $v_{\varphi(n)}$ on voit que $2(\ell - a)$ est encore une valeur d'adhérence de la suite $(u_n)_{n\geqslant 0}$. En itérant cela, on en déduit que tous les termes de la suite $(a_n)_{n\geqslant 0}$ définie par $a_0 = a$ et $a_{n+1} = 2(\ell - a_n)$ sont des valeurs d'adhérence. Cette suite arithmético-géométrique de raison -2 n'est pas bornée sauf si $a_0 = a$ est le point fixe de la fonction affine $x \mapsto 2(\ell - x)$. Comme l'ensemble des valeurs d'adhérence de $(u_n)_{n\geqslant 0}$ est borné, on a donc forcément $a = \dfrac{2\ell}{3}$.

Conclusion. La suite bornée $(u_n)_{n\geqslant 0}$ admet $\dfrac{2\ell}{3}$ pour unique valeur d'adhérence et converge donc vers cette valeur. ◁

2.13. Étude de convergence

1. Soit a_n et b_n deux suites réelles telles que $a_n + b_n \to 0$ et $e^{a_n} + e^{b_n} \to 2$. Montrer que les deux suites convergent.

2. Soient a_n, b_n, c_n trois suites réelles telles que $a_n + b_n + c_n \to 0$ et $e^{a_n} + e^{b_n} + e^{c_n} \to 3$. Que peut-on dire de la convergence de ces trois suites ?

(École normale supérieure)

Solution.

1. Posons $u_n = e^{a_n} + e^{b_n}$. Comme $e^{a_n} \leqslant u_n$ et $e^{b_n} \leqslant u_n$ on peut déjà dire que les suites (e^{a_n}) et (e^{b_n}) sont majorées. Il en est donc de même des suites (a_n) et (b_n). On en déduit que les deux suites sont bornées : en effet, en écrivant par exemple $a_n = (a_n + b_n) - b_n$, on montre que (a_n) est minorée comme somme de deux suites minorées.

Soit alors a une valeur d'adhérence de a_n et $\varphi : \mathbb{N} \to \mathbb{N}$ telle que $(a_{\varphi(n)})$ converge vers a. La suite $(b_{\varphi(n)})$ converge alors vers $-a$ et on a donc $e^a + e^{-a} = 2$. Or pour $x > 0$ on a toujours $x + \dfrac{1}{x} \geqslant 2$ avec égalité si et seulement si $x = 1$ (inégalité des moyennes). On en déduit que $e^a = 1$ donc que $a = 0$. Ainsi (a_n) admet 0 pour seule valeur d'adhérence et converge donc vers 0. Il en est de même pour (b_n).

2. On montre comme dans la question précédente que les trois suites sont bornées. Soit a une valeur d'adhérence de la suite (a_n) et $\varphi : \mathbb{N} \to \mathbb{N}$ telle que $(a_{\varphi(n)})$ converge vers a. Quitte à extraire une nous-suite convergente de la suite bornée $(b_{\varphi(n)})$ on peut directement supposer que $(b_{\varphi(n)})$ converge vers une valeur d'adhérence b de la suite b_n. L'hypothèse $a_n + b_n + c_n \to 0$ permet alors de dire que $(c_{\varphi(n)})$ tend vers $-a-b$. Par continuité de l'exponentielle on a donc $e^a + e^b + e^{-a-b} = 3$. On va prouver que cela implique $a = b = 0$ ce qui permettra de conclure comme dans la question précédente.

Pour cela on étudie $f(x, y) = x + y + \dfrac{1}{xy} - 3$ sur $(\mathbb{R}_+^*)^2$: il suffit de prouver que le minimum de f est 0 et qu'il n'est atteint qu'en $(1, 1)$. Supposons $y > 0$ fixé. La dérivée partielle par rapport à x vaut $1 - \dfrac{1}{x^2 y}$ et $x \mapsto f(x, y)$ est minimal lorsque $x = \dfrac{1}{\sqrt{y}}$. On se ramène donc à l'étude de la fonction d'une variable $g : y \mapsto y + \dfrac{2}{\sqrt{y}} - 3$. Comme $g'(y) = 1 - y^{-3/2}$, on voit tout de suite que g est minimale en 1 et seulement en ce point, et que le minimum vaut 0.

Conclusion. Les trois suites convergent vers 0. ◁

Nous avons vu précédemment que l'ensemble X_u *des valeurs d'adhérence d'une suite bornée* (u_n) *est un compact non vide. Dans le cas d'une suite réelle,* X_u *possède donc un plus petit élément* ℓ*, appelé* limite inférieure *de* (u_n)*, noté* $\liminf u_n$*, et un plus grand élément* L*, appelé* limite supérieure *de* (u_n)*, noté* $\limsup u_n$.

Si (u_n) *est une suite réelle bornée, la suite de terme général* $\sup_{p\geqslant n} u_p$ *est décroissante et bornée donc convergente. On peut démontrer que* $\limsup u_n = \lim\limits_{n\to+\infty} (\sup\limits_{p\geqslant n} u_p)$ *et de même que* $\liminf u_n = \lim\limits_{n\to+\infty} (\inf\limits_{p\geqslant n} u_p)$*, ce qui explique la notation.*

Dans l'exercice suivant on cherche donc à minorer la limite supérieure de la suite considérée.

2.14. Minoration de la limite supérieure

Soit (a_n) une suite réelle à termes strictement positifs. On pose $b_n = n\left(\dfrac{1+a_{n+1}}{a_n} - 1\right)$. Montrer que $\lim\limits_{n\to+\infty} \sup\limits_{k\geqslant n} b_k \geqslant 1$.

(**École normale supérieure**)

▷ **Solution.**

L'existence de la limite résulte de la décroissance de la suite $(\sup\limits_{k\geqslant n} b_k)_{n\in\mathbb{N}}$. On raisonne par l'absurde et on suppose que $\lim\limits_{n\to+\infty} \sup\limits_{k\geqslant n} b_k = \ell < 1$. Soit $\lambda \in]\ell, 1[$. Il existe $n_0 \in \mathbb{N}$ tel que, pour $n \geqslant n_0$, on ait $\sup\limits_{k\geqslant n} b_k \leqslant \lambda$, c'est-à-dire $b_k \leqslant \lambda$ pour tout $k \geqslant n_0$. On écrit cette inégalité $1 + a_{k+1} - a_k \leqslant \lambda \dfrac{a_k}{k}$. En sommant ces inégalités pour k variant de n_0 à n , on obtient, pour tout $n > n_0$,

$$n - n_0 + 1 + a_{n+1} - a_{n_0} \leqslant \lambda \sum_{k=n_0}^{n} \frac{a_k}{k}.$$

On pose $\mathrm{S}_n = \sum\limits_{k=n_0}^{n} \dfrac{a_k}{k}$. On obtient $a_{n+1} = (n+1)(\mathrm{S}_{n+1} - \mathrm{S}_n)$ et donc

$$\mathrm{S}_{n+1} \leqslant \frac{(n+1+\lambda)\mathrm{S}_n - n + n_0 - 1 + a_{n_0}}{n+1}.$$

La suite (S_n) est croissante. Si elle converge vers L, on obtient par passage à la limite $\mathrm{L} \leqslant \mathrm{L} - 1$. La suite (S_n) diverge donc vers $+\infty$. Comme $\lambda < 1$, il existe $n_1 \geqslant n_0$ tel que l'on ait $\lambda \mathrm{S}_n + n_0 - 1 + a_{n_0} \leqslant \mathrm{S}_n$ et donc $\mathrm{S}_{n+1} \leqslant \dfrac{(n+2)\mathrm{S}_n - n}{n+1}$ pour $n \geqslant n_1$. Cette inégalité peut s'écrire

$$\frac{S_n}{n+1} - \frac{S_{n+1}}{n+2} \geqslant \frac{n}{(n+1)(n+2)}.$$

La suite de terme général $\dfrac{S_n}{n+1}$ est donc décroissante. Comme elle est minorée par 0, elle converge. On en déduit que la série de terme général $\dfrac{S_n}{n+1} - \dfrac{S_{n+1}}{n+2}$ converge. C'est impossible car elle est supérieure à une série de terme général équivalent à $\dfrac{1}{n}$ et donc divergente. Nous obtenons la contradiction voulue ◁

Il a été démontré dans 1.13 que tout réel de $[-1,1]$ *est valeur d'adhérence de la suite* $(\sin n)$. *Ce résultat est utilisé dans l'exercice suivant. On notera aussi que si* ℓ *est une valeur d'adhérence de la suite* (u_n), *alors la suite* (u_n) *ne peut converger que vers* ℓ.

2.15. Suites et approximation diophantienne

1. *Théorème de Dirichlet.* Soit α un irrationnel. Montrer qu'il existe une infinité de couples $(p,q) \in \mathbb{Z} \times \mathbb{N}^*$ tels que

$$\left|\alpha - \frac{p}{q}\right| \leqslant \frac{1}{q^2}.$$

2. Étudier la convergence de la suite $\left(\dfrac{1}{n \sin n}\right)_{n \geqslant 1}$.

3. Montrer qu'il existe $\alpha > 0$ tel que, pour tout $s \geqslant 0$, la suite $\left(\dfrac{1}{n^s \sin(\pi \alpha n)}\right)_{n \geqslant 1}$ diverge.

(École normale supérieure)

▷ **Solution.**

1. Pour $N \in \mathbb{N}^*$, on considère les $N+1$ réels $k\alpha$ (où k décrit les entiers de 0 à N). Les $N+1$ différences $k\alpha - E(k\alpha)$ appartiennent à l'intervalle $[0,1[$ qui est réunion disjointe des N «tiroirs» $\left[\dfrac{i}{N}, \dfrac{i+1}{N}\right[$ $(0 \leqslant i \leqslant N-1)$. Ainsi, un de ces intervalles de longueur $\dfrac{1}{N}$ contient nécessairement deux de ces différences ; autrement dit, il existe $0 \leqslant k < \ell \leqslant N$ et $0 \leqslant i \leqslant N-1$ tels que

$$k\alpha - E(k\alpha) \text{ et } \ell\alpha - E(\ell\alpha) \text{ sont dans } \left[\frac{i}{N}, \frac{i+1}{N}\right[.$$

En posant $q = \ell - k \in [\![1, N]\!]$ et $p = E(\ell\alpha) - E(k\alpha)$ on obtient

$$|q\alpha - p| < \frac{1}{\mathrm{N}} \leqslant \frac{1}{q} \quad \text{et} \quad \left|\alpha - \frac{p}{q}\right| \leqslant \frac{1}{q^2}.$$

Il nous reste à vérifier qu'il y a une infinité de couples vérifiant l'inégalité. Si on suppose ces couples en nombre fini, on peut en trouver un pour lequel la distance $d = |q\alpha - p|$ est minimale. Comme α est irrationnel, d est strictement positif et on peut choisir N entier naturel avec $\frac{1}{\mathrm{N}} < d$. Le travail qui vient d'être fait permet de construire un couple (p', q') de $\mathbb{Z} \times \mathbb{N}^*$ tels que $|q'\alpha - p'| < \frac{1}{\mathrm{N}} < d$ et $\left|\alpha - \frac{p'}{q'}\right| \leqslant \frac{1}{q'^2}$, ce qui est contradictoire.

2. Il est classique que la suite $(\sin n)_{n\geqslant 0}$ est dense dans $[-1, 1]$ (c'est l'objet de l'exercice 1.13.) Aussi peut-on trouver une suite $(k_n)_{n\geqslant 0}$ d'entiers, strictement croissante, telle que la limite de $\sin k_n$ soit égale à 1. La suite extraite $\left(\frac{1}{k_n \sin k_n}\right)$ converge alors vers 0.

Pour prouver que la suite de terme général $u_n = \frac{1}{n \sin n}$ ne peut pas converger vers 0, on cherche une suite extraite pour laquelle u_n reste assez grand, c'est-à-dire $n \sin n$ petit et donc n proche de $\pi\mathbb{Z}$, ou encore $\frac{n}{\pi}$ proche de $\mathbb{Z}$. On pose $\alpha = \frac{1}{\pi}$, qui est irrationnel, de telle manière que $\sin n = \sin \pi\alpha n$. On contrôle la distance de $n\alpha$ aux entiers à l'aide du théorème de Dirichlet, objet de la question précédente.

Soit $\mathrm{E} = \left\{q \in \mathbb{N}^*, \exists p \in \mathbb{Z}, \left|\alpha - \frac{p}{q}\right| \leqslant \frac{1}{q^2}\right\}$. L'ensemble E est infini car non majoré. En effet, si E était majoré, il n'y aurait qu'un nombre fini de couples (p, q) satisfaisant à la relation $\left|\alpha - \frac{p}{q}\right| \leqslant \frac{1}{q^2}$ puisque, dans ces conditions,

$$|p| - |\alpha| q \leqslant \frac{1}{q} \leqslant 1 \quad \text{et} \quad |p| \leqslant 1 + |\alpha| q,$$

ce qui, pour $q \in \mathrm{E}$, conduit à un nombre fini de p. Il existe donc une suite $(q_n)_{n\in\mathbb{N}}$ strictement croissante constituée des éléments de E et une suite $(p_n)_{n\in\mathbb{N}}$ d'entiers relatifs telles que, pour tout entier n,

$$\left|\alpha - \frac{p_n}{q_n}\right| \leqslant \frac{1}{q_n^2}, \quad i.e. \quad |\alpha q_n - p_n| \leqslant \frac{1}{q_n}.$$

Ainsi, pour n entier naturel,

$$|\sin \pi\alpha q_n| = |\sin(\pi\alpha q_n - \pi p_n)| = |\sin \pi(\alpha q_n - p_n)| \leqslant \frac{\pi}{q_n},$$

si bien que $|u_{q_n}| = \frac{1}{q_n |\sin \alpha\pi q_n|} \geqslant \frac{1}{\pi}$. La suite $(u_n)_{n\geqslant 1}$ ne saurait converger vers 0. Comme 0 est une valeur d'adhérence, elle est donc divergente.

3. On va faire en sorte que la suite $(\sin \pi \alpha n)$ ait 0 comme valeur d'adhérence et, comme dans la question précédente, on va essayer de contrôler la vitesse de convergence afin d'estimer le produit $n^s \sin(\pi \alpha n)$. On pose $\alpha = \sum_{k=0}^{+\infty} \frac{1}{10^{k!}}$. On a alors

$$|\sin \pi \alpha 10^{n!}| = \left|\sin\left(\pi \sum_{k=n+1}^{+\infty} \frac{1}{10^{k!-n!}}\right)\right| \leqslant \pi \left(\sum_{k=n+1}^{+\infty} \frac{1}{10^{k!-n!}}\right)$$

$$\leqslant \pi \sum_{k=(n+1)!-n!}^{+\infty} \frac{1}{10^k} \leqslant \frac{\pi}{10^{(n+1)!-n!}} \frac{1}{1-\frac{1}{10}} \leqslant \mathrm{K} \frac{10^{n!}}{10^{(n+1)!}},$$

où $\mathrm{K} = \frac{10\pi}{9}$. Pour $s \geqslant 0$, on étudie $v_n = (10^{n!})^s |\sin \pi \alpha 10^{n!}|$. On a la majoration

$$v_n \leqslant \mathrm{K} \frac{(10^{n!})^{s+1}}{10^{(n+1)!}}$$

et, en passant au logarithme,

$$\ln v_n \leqslant \ln \mathrm{K} + n!(s+1)\ln 10 - (n+1)!\ln 10 \xrightarrow[n \to +\infty]{} -\infty.$$

On en déduit que (v_n) converge vers 0 et par conséquent,

$$\lim_{n \to +\infty} \left| \frac{1}{k_n^s \sin \pi \alpha k_n} \right| = +\infty \quad \text{pour } k_n = 10^{n!}.$$

Conclusion. Pour $\alpha = \sum_{k=0}^{+\infty} \frac{1}{10^{k!}}$ et $s \geqslant 0$, la suite $\left(\frac{1}{n^s \sin(\pi \alpha n)}\right)_{n \geqslant 1}$ diverge. ◁

On aura reconnu en α un nombre de Liouville (voir l'exercice 1.11 du présent tome ainsi que l'exercice 5.47 du tome 1 d'algèbre).

Nous commençons une série d'exercices étudiant des exemples concrets de suites.

2.16. Équivalent d'une suite d'entiers

Soit $(u_n)_{n \geqslant 1}$ définie par $u_1 = \mathbf{1}$, $u_2 = u_3 = 2$, $u_4 = u_5 = u_6 = 3$,... Elle prend une fois la valeur 1, deux fois la valeur 2, trois fois la valeur 3 et ainsi de suite. Donner une expression du terme général (à l'aide de la fonction partie entière) et un équivalent de u_n.

(École polytechnique)

▷ **Solution.**

Il est clair que la suite est croissante et diverge vers $+\infty$. En fait, on a $u_n = k$ lorsque $\dfrac{k(k-1)}{2} + 1 \leqslant n < \dfrac{k(k+1)}{2} + 1$. Sans expliciter le terme général, cet encadrement nous donne déjà un équivalent de u_n. En effet, on a

$$\frac{(u_n - 1)}{2u_n} + \frac{1}{u_n} \leqslant \frac{n}{u_n^2} < \frac{(u_n + 1)}{2u_n} + \frac{1}{u_n}$$

de sorte que $\dfrac{n}{u_n^2}$ tend vers $\dfrac{1}{2}$ et $u_n^2 \sim 2n$ soit $u_n \sim \sqrt{2n}$.

Cherchons maintenant une expression explicite de u_n. L'inégalité $\dfrac{k(k-1)}{2} + 1 \leqslant n$ équivaut $k^2 - k + 2 - 2n \leqslant 0$ et, k étant positif, elle est vérifiée si et seulement si $k \leqslant \dfrac{1 + \sqrt{8n-7}}{2}$. De la même manière on a $n < \dfrac{k(k+1)}{2} + 1$ si et seulement si $k + 1 > \dfrac{1 + \sqrt{8n-7}}{2}$. Cet encadrement prouve que

$$\boxed{u_n = \mathrm{E}\left(\frac{1 + \sqrt{8n-7}}{2}\right).}$$

Bien entendu on retrouve l'équivalent obtenu plus haut. ◁

2.17. Somme des puissances n-ièmes des n premiers entiers

Étudier la suite $u_n = \dfrac{1}{n^n} \sum_{k=1}^{n} k^n$.

(**École polytechnique**)

▷ **Solution.**

On a

$$u_n = \frac{1}{n^n} + \cdots + \frac{(n-1)^n}{n^n} + \frac{n^n}{n^n} = 1 + \left(\frac{n-1}{n}\right)^n + \left(\frac{n-2}{n}\right)^n + \cdots + \left(\frac{1}{n}\right)^n.$$

Comme $\left(1 - \dfrac{p}{n}\right)^n$ tend vers e^{-p} lorsque n tend vers l'infini, le second terme de cette somme tend vers $\dfrac{1}{e}$, le troisième vers $\dfrac{1}{e^2}$, ..., de sorte qu'on peut penser que la limite de (u_n) est $\sum_{k=0}^{+\infty} e^{-k} = \dfrac{e}{e-1}$.

Pour le montrer, on écrit $u_n = \sum_{k=0}^{n-1} \left(1 - \dfrac{k}{n}\right)^n$. En utilisant l'inégalité $\ln(1+x) \leqslant x$, valable pour tout réel $x > -1$, on obtient déjà

$$0 \leqslant u_n \leqslant \sum_{k=0}^{n-1} e^{-k} = \frac{1-e^{-n}}{1-e^{-1}} \leqslant \frac{e}{e-1}.$$

Minorons u_n maintenant. Soit $N \geqslant 0$. Pour $n \geqslant N$, on a

$$u_n \geqslant \sum_{k=0}^{N} \left(1-\frac{k}{n}\right)^n = v_n$$

et la limite de v_n est $\ell_N = \sum_{k=0}^{N} e^{-k} = \frac{e-e^{-N-1}}{e-1}$. La limite de ℓ_N est $\frac{e}{e-1}$ quand N tend vers l'infini. Soit $\varepsilon > 0$. On choisit N tel que $\ell_N \geqslant \frac{e}{e-1} - \varepsilon$. Comme v_n converge vers ℓ_N, pour n assez grand, $v_n \geqslant \ell_N - \varepsilon \geqslant \frac{e}{e-1} - 2\varepsilon$. Dans ces conditions, pour n assez grand, on a l'encadrement

$$\frac{e}{e-1} - 2\varepsilon \leqslant v_n \leqslant u_n \leqslant \frac{e}{e-1},$$

ce qui prouve que u_n converge vers $\frac{e}{e-1}$. ◁

Dans l'exercice suivant ont été regroupés deux énoncés posés séparément. Il fait la transition avec le prochain thème consacré aux suites récurrentes.

2.18. Racines itérées

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels positifs. On étudie la suite $(u_n)_{n\in\mathbb{N}}$ définie par

$$\forall n \in \mathbb{N}, \quad u_n = \sqrt{a_0 + \sqrt{a_1 + \cdots + \sqrt{a_n}}}$$

1. Étudier (u_n) lorsque (a_n) est constante égale à $a > 0$, puis lorsque $a_n = \lambda^{2^{n+1}}$, où $\lambda > 0$.

2. Montrer que la suite (u_n) converge si et seulement si la suite $(a_n^{1/2^n})_{n\geqslant 0}$ est bornée.

3. Déterminer $\lim_{n\to+\infty} \sqrt{1+2\sqrt{1+3\sqrt{1+\cdots+(n-1)\sqrt{1+n}}}}$.

(École polytechnique)

▷ **Solution.**

1. Supposons tout d'abord que (a_n) est constante égale à $a > 0$. On a alors $u_{n+1} = \sqrt{a + u_n}$ pour tout entier n, ce qui nous ramène à une suite récurrente classique. L'intervalle $\mathbb{R}_+$ est stable par l'application $f : x \longmapsto \sqrt{x + a}$. Celle-ci f est croissante sur $\mathbb{R}_+$ et y possède un unique point fixe $\ell = \dfrac{1 + \sqrt{1 + 4a}}{2}$. Sur l'intervalle $[0, \ell]$ on a $f(x) > x$. Comme $u_0 = \sqrt{a} < \ell$, la suite (u_n) est croissante, majorée par ℓ et converge donc nécessairement vers ℓ.

Supposons maintenant que $a_n = \lambda^{2^{n+1}}$. On a alors $u_0 = \sqrt{\lambda^2} = \lambda$, $u_1 = \sqrt{\lambda^2 + \sqrt{\lambda^4}} = \lambda\sqrt{1 + \sqrt{1}}$ et plus généralement,

$$u_n = \lambda\sqrt{1 + \sqrt{1 + \cdots + \sqrt{1}}}.$$

D'après le point précédent, (u_n) converge donc vers $\dfrac{1 + \sqrt{5}}{2}\lambda$.

2. Commençons par observer que la suite (u_n) est toujours croissante, car comme $a_n \leqslant a_n + \sqrt{a_{n+1}}$, la croissance de la fonction racine carrée conduit à $u_n \leqslant u_{n+1}$. La suite (u_n) converge donc si et seulement si elle est majorée. Notons aussi que si (a'_n) est une seconde suite de réels positifs qui vérifie $a_n \leqslant a'_n$ pour tout n, alors on a $u_n \leqslant u'_n$ pour tout n, (u'_n) étant la suite associée à (a'_n) de la même manière que (u_n) est associée à (a_n). Démontrons maintenant l'équivalence proposée.

Si dans l'expression de u_n on minore $a_0, a_1, \ldots, a_{n-1}$ par 0, il vient $u_n \geqslant a_n^{1/2^{n+1}}$. On a donc $a_n^{1/2^n} \leqslant u_n^2$ de sorte que si la suite (u_n) converge, la suite $(a_n^{1/2^n})_{n \geqslant 0}$ est bornée. Montrons la réciproque.

Il existe par hypothèse un réel $\lambda > 0$ tel que pour tout n, $a_n^{1/2^n} \leqslant \lambda$. On en déduit que $a_n \leqslant \lambda^{2^n}$ Les remarques qui précèdent, associées au résultat de la première question, permettent d'affirmer que (u_n) converge.

3. En faisant passer les entiers $2, 3, \ldots$ dans les racines carrées qui suivent on retrouve bien une suite du type précédent. En notant (u_n) la suite proposée, on écrit ainsi,

$$u_2 = \sqrt{1 + 2\sqrt{1 + 3}} = \sqrt{1 + \sqrt{2^2 + \sqrt{2^4.3^2}}},$$

$$u_3 = \sqrt{1 + 2\sqrt{1 + 3\sqrt{1 + 4}}} = \sqrt{1 + \sqrt{2^2 + \sqrt{2^4.3^2 + \sqrt{2^8.3^4.4^2}}}}.$$

Il s'agit donc de la suite (u_n) associée à la suite (a_n) définie par $a_0 = 1$ et $a_n = 2^{2^n} 3^{2^{n-1}} \ldots n^4 (n+1)^2$ pour tout $n \geqslant 1$. On utilise le critère de convergence de la question 2. On a pour tout $n \geqslant 2$,

$$\frac{\ln a_n}{2^n} = \sum_{k=2}^{n+1} \frac{\ln k}{2^{k-2}}.$$

Il s'agit de la somme partielle d'une série convergente. La suite $\dfrac{\ln a_n}{2^n}$ converge, et en passant à l'exponentielle, la suite $(a_n^{1/2^n})$ également. En particulier elle est bornée. D'après la question précédente, la limite recherchée est finie. On demande toutefois ici de déterminer sa valeur. Une étude numérique semble indiquer que la limite est 3. Pour le démontrer, on va majorer la quantité $3 - u_n$ en multipliant chaque expression par sa conjuguée pour éliminer les racines carrées. On a, puisque $u_n \geqslant 1$,

$$\begin{aligned} |3 - u_n| = \frac{|9 - u_n^2|}{3 + u_n} &\leqslant \frac{1}{4}|9 - u_n^2| \\ &\leqslant \frac{2}{4}\left|4 - \sqrt{1 + 3\sqrt{1 + 4\sqrt{\cdots + (n-1)\sqrt{1+n}}}}\right|. \end{aligned}$$

On recommence la même majoration. Il vient

$$|3 - u_n| \leqslant \frac{2 \cdot 3}{4 \cdot 5}\left|5 - \sqrt{1 + 4\sqrt{1 + 5\sqrt{\cdots + (n-1)\sqrt{1+n}}}}\right|.$$

On réitère le procédé pour obtenir finalement,

$$|3 - u_n| \leqslant \frac{2 \cdot 3 \cdots (n-1)}{4 \cdot 5 \cdots (n+1)}\left(n + 1 - \sqrt{1+n}\right) \leqslant \frac{6}{n},$$

ce qui prouve que (u_n) converge vers 3. ◁

Les exercices qui suivent sont consacrés à l'étude de systèmes dynamiques discrets, autrement dit de suites vérifiant une relation de récurrence de la forme $u_{n+1} = f(u_n)$. L'intervalle I *étant stable par f et u_0 appartenant à* I*, on s'intéresse donc à la suite $(f^n(u_0))_{n\in\mathbb{N}}$ des images de u_0 par les itérés de f, qui constituent ce qu'on appelle l'orbite du point u_0.*

On sait que dans le cas où f est continue, les limites possibles de la suite (u_n) sont les points fixes de f. L'exercice théorique suivant donne une condition nécessaire et suffisante de convergence d'un système dynamique lorsque f est définie sur un intervalle borné.

2.19. Un critère de convergence pour les systèmes dynamiques discrets

Soit $f : [0,1] \to [0,1]$ continue et $(u_n)_{n\in\mathbb{N}}$ une suite de $[0,1]$ tel que pour tout $n \in \mathbb{N}$, $f(u_n) = u_{n+1}$. Montrer que $(u_n)_{n\in\mathbb{N}}$ converge si, et seulement si, $\lim\limits_{n\to+\infty}(u_{n+1} - u_n) = 0$.

(École polytechnique)

▷ **Solution.**

Bien évidemment, si $(u_n)_{n\in\mathbb{N}}$ converge vers ℓ,

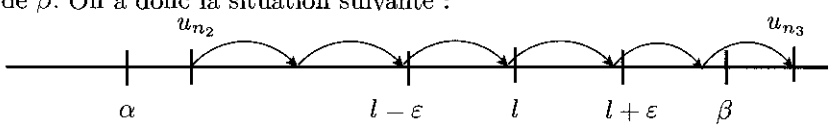
$$\lim_{n\to+\infty}(u_{n+1} - u_n) = \ell - \ell = 0.$$

Passons à la réciproque. Supposons que $\lim\limits_{n\to+\infty}(u_{n+1} - u_n) = 0$. La suite (u_n) étant bornée, il suffit de démontrer qu'elle possède une seule valeur d'adhérence. Raisonnons par l'absurde et supposons que (u_n) possède deux valeurs d'adhérence α et β $(\alpha < \beta)$.

Nous allons montrer qu'alors tout réel $\ell \in]\alpha, \beta[$ est aussi valeur d'adhérence. Soit $\varepsilon > 0$ et $N \geqslant 0$. Il faut démontrer l'existence de $n \geqslant N$ tel que $|u_n - \ell| \leqslant \varepsilon$. Quitte à diminuer ε, on peut supposer qu'on a $\alpha < \ell - \varepsilon < \ell + \varepsilon < \beta$.

Par hypothèse, il existe $n_0 \in \mathbb{N}$ tel que $|u_{n+1} - u_n| \leqslant \varepsilon$ pour tout $n \geqslant n_0$. Notons $n_1 = \max(N, n_0)$. Pour $n \geqslant n_1$, on passe de u_n à u_{n+1} en faisant des «sauts» dont la longueur ne peut excéder ε.

Puisque α est une valeur d'adhérence, il existe $n_2 \geqslant n_1$ tel que u_{n_2} soit dans le voisinage $]-\infty, \ell - \varepsilon[$ de α. De même, β étant une valeur d'adhérence, il existe $n_3 \geqslant n_2$ tel que u_{n_3} soit dans le voisinage $]\ell+\varepsilon, +\infty[$ de β. On a donc la situation suivante :



Pour aller de $u_{n_2} < \ell - \varepsilon$ à $u_{n_3} > \ell + \varepsilon$ en faisant des sauts inférieurs à ε, les u_n pour $n \in [\![n_2, n_3]\!]$ sont obligés de passer dans l'intervalle $[\ell - \varepsilon, \ell + \varepsilon]$. Considérons

$$n = \max\{p \in [\![n_2, n_3]\!], u_p < \ell - \varepsilon\}.$$

L'entier n est bien défini puisque $\{p \in [\![n_2, n_3]\!], u_p < \ell - \varepsilon\}$ est une partie non vide (elle contient n_2), majorée de $\mathbb{N}$ et $n < n_3$. On a donc $u_{n+1} \geqslant \ell - \varepsilon$, mais aussi

$$u_{n+1} \leqslant u_n + |u_{n+1} - u_n| \leqslant u_n + \varepsilon \leqslant \ell - \varepsilon + \varepsilon = \ell.$$

On a trouvé un entier $n + 1 \geqslant N$ tel que $u_{n+1} \in [\ell - \varepsilon, \ell]$. Ceci montre que ℓ est une valeur d'adhérence de $(u_n)_{n\in\mathbb{N}}$.

Montrons qu'alors tout élément $\ell \in [\alpha, \beta]$ est point fixe de f. Puisque ℓ est valeur d'adhérence de la suite, il existe une sous-suite $(u_{\varphi(n)})$ qui converge vers ℓ. Par continuité de f, $f(u_{\varphi(n)}) = u_{\varphi(n)+1} \xrightarrow[n\to+\infty]{} f(\ell)$. Puisque par hypothèse $u_{\varphi(n)+1} - u_{\varphi(n)} \xrightarrow[n\to+\infty]{} 0$, on a $f(\ell) = \ell$.

En particulier, puisque $\dfrac{\alpha+\beta}{2}$ est une valeur d'adhérence de $(u_n)_{n\in\mathbb{N}}$, il existe $n_0 \in \mathbb{N}$ tel que $u_{n_0} \in [\alpha, \beta]$. Mais alors on a $u_n = u_{n_0}$ pour tout $n \geqslant n_0$ et $(u_n)_{n\in\mathbb{N}}$ converge, contrairement à ce que nous avions supposé. ◁

La démonstration précédente permet de montrer que si (u_n) est une suite bornée vérifiant $\lim\limits_{n\to+\infty}(u_{n+1} - u_n) = 0$, alors l'ensemble X_u des valeurs d'adhérence de la suite (u_n) est un intervalle. Mais nous savons déjà que, pour toute suite bornée, X_u est compact. Nous en déduisons que X_u est un segment[2].

Tout point fixe ℓ de f est limite d'au moins une suite vérifiant la récurrence $u_{n+1} = f(u_n)$, à savoir la suite constante obtenue en prenant $u_0 = \ell$. Mais le cas intéressant est celui où il y a «beaucoup» de suites de limite ℓ : on dit par exemple que ℓ est un point fixe attractif s'il y a un intervalle ouvert I *contenant ℓ tel que (u_n) converge vers ℓ si $u_0 \in$ I. C'est le cas si f est de classe $\mathcal{C}^1$ et $|f'(\ell)| < 1$. On choisit $k \in]|f'(1)|, 1[$ et un intervalle* I *contenant ℓ sur lequel $|f'| \leqslant k$. La convergence de (u_n) pour $u_0 \in$ I résulte de l'inégalité des accroissements finis. Au contraire, si f est de classe $\mathcal{C}^1$ et si $|f'(\ell)| > 1$, la suite (u_n) ne converge vers ℓ que si la suite est stationnaire : le point ℓ est appelé point fixe répulsif (ce résultat est démontré dans l'exercice suivant).*

On peut s'intéresser à d'autres comportements asymptotiques que la convergence, par exemple au cas où la suite des termes pairs (u_{2n}) converge vers une limite ℓ_1 et la suite des termes impairs (u_{2n+1}) vers une limite ℓ_2. On a alors forcément $f(\ell_1) = \ell_2$ et $f(\ell_2) = \ell_1$, c'est-à-dire que ℓ_1 et ℓ_2 sont des points fixes de $f \circ f$. Si ℓ_1 et ℓ_2 sont distincts ils ne sont pas points fixes de f et on parle de points d'ordre 2 : la suite définie par $u_0 = \ell_1$ (ou $u_0 = \ell_2$) est une suite périodique de période 2.

On sait que dans le cas où f est croissante la suite (u_n) est monotone. L'exercice suivant étudie un exemple de système dynamique $u_{n+1} = f(u_n)$, où la fonction f est décroissante. La fonction $f \circ f$ est croissante et les deux sous-suites (u_{2n}) et (u_{2n+1}) sont monotones de sens contraires.

---

2. Ce résultat se généralise : l'ensemble des valeurs d'adhérence d'une suite $(u_n)_{n\in\mathbb{N}}$ d'un espace métrique compact (E, d) qui vérifie $\lim d(u_n, u_{n+1}) = 0$ est connexe : voir l'exercice 1-11 de (A.) CHAMBERT-LOIR, (S.) FERMIGIER, (V.) MAILLOT, *Exercices de mathématiques pour l'agrégation*, Analyse 1, Masson 1994.

2.20. Étude d'un système dynamique (1)

Soit $\lambda \in]0,1]$. Étudier la suite $(x_n)_{n\in\mathbb{N}}$ définie par $x_0 \in]0,1[$ et pour tout $n \in \mathbb{N}$, $x_{n+1} = 1 - \lambda x_n^2$.

(École polytechnique)

▷ **Solution.**

Considérons la fonction $f : x \longmapsto 1 - \lambda x^2$. L'intervalle $]0,1[$ étant stable par f, la suite (x_n) est à valeurs dans $]0,1[$. D'autre part, f est décroissante sur $]0,1[$. On en déduit que la fonction $f \circ f$ est croissante sur $]0,1[$. Les deux sous-suites $(x_{2n})_{n\in\mathbb{N}^*}$ et $(x_{2n+1})_{n\in\mathbb{N}}$ sont donc monotones. Comme elles sont bornées, elles sont toutes deux convergentes. Reste à savoir si elles ont la même limite.

Les suites $(x_{2n})_{n\in\mathbb{N}}$ et $(x_{2n+1})_{n\in\mathbb{N}}$ convergent vers un point fixe de $f \circ f$ dans $[0,1]$. Pour tout $x \in [0,1]$,

$$(f \circ f)(x) - x = -\lambda^3 x^4 + 2\lambda^2 x^2 - x + 1 - \lambda.$$

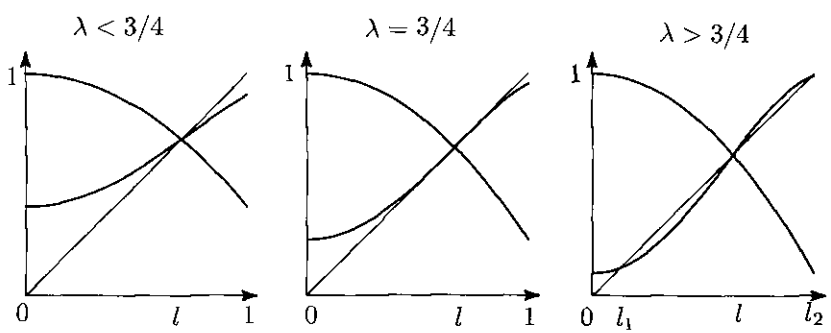
Cette expression peut être factorisée par $f(x) - x = -\lambda x^2 - x + 1$, car les points fixes de f sont des points fixes de $f \circ f$. On trouve

$$(f \circ f)(x) - x = (-\lambda x^2 - x + 1)(\lambda^2 x^2 - \lambda x + 1 - \lambda).$$

L'équation $-\lambda x^2 - x + 1 = 0$ a une seule solution dans $[0,1]$, qui est l'unique point fixe de f dans $[0,1]$,

$$\ell = \frac{\sqrt{1+4\lambda}-1}{2\lambda} = \frac{2}{\sqrt{1+4\lambda}+1}.$$

Quant au discriminant de $\lambda^2 x^2 - \lambda x + 1 - \lambda$, il est égal à $\lambda^2(-3+4\lambda)$. Son signe dépend de la place de λ par rapport à $\frac{3}{4}$. On va donc distinguer trois cas. La figure ci-après donne l'allure des graphes de f et $f \circ f$ dans chacun des trois cas :



• Premier cas : $\lambda \in \left]0, \frac{3}{4}\right[$. L'équation $\lambda^2 x^2 - \lambda x + 1 - \lambda = 0$ n'a pas de solution ; le seul point fixe de $f \circ f$ dans $[0,1]$ est ℓ. Les suites $(x_{2n})_{n\in\mathbb{N}}$ et $(x_{2n+1})_{n\in\mathbb{N}}$ convergent vers ℓ. La suite $(x_n)_{n\in\mathbb{N}}$ converge vers ℓ.

• Second cas : $\lambda = \frac{3}{4}$. L'équation $\lambda^2 x^2 - \lambda x + 1 - \lambda = 0$ a une seule solution $\frac{1}{2\lambda} = \frac{2}{3}$, qui est égale à ℓ. La conclusion est la même.

• Troisième cas : $\lambda \in \left]\frac{3}{4}, 1\right]$. L'équation $\lambda^2 x^2 - \lambda x + 1 - \lambda = 0$ a deux solutions

$$\ell_1 = \frac{1 - \sqrt{4\lambda - 3}}{2\lambda} \quad \text{et} \quad \ell_2 = \frac{1 + \sqrt{4\lambda - 3}}{2\lambda}.$$

Montrons que $(x_n)_{n\in\mathbb{N}}$ ne peut pas converger vers ℓ, à moins d'être stationnaire. On a $|f'(\ell)| = |-2\lambda\ell| = |\sqrt{1+4\lambda} - 1| = \sqrt{1+4\lambda} - 1 > 1$, car $\lambda > \frac{3}{4}$. On dit que ℓ est un point fixe répulsif de f. Soit $k \in]1, |f'(\ell)|[$. Par continuité de f', il existe un voisinage V de ℓ tel que si $x \in \mathrm{V}$, alors $|f'(x)| > k$. Supposons que la suite $(x_n)_{n\in\mathbb{N}}$ converge vers ℓ. Il existe un entier n_0 tel que $x_n \in \mathrm{V}$ si $n \geqslant n_0$. De l'inégalité des accroissements finis, on déduit que $|f(x_n) - f(\ell)| \geqslant k|x_n - \ell|$, c'est-à-dire $|x_{n+1} - \ell| \geqslant k|x_n - \ell|$ pour $n \geqslant n_0$, puis $|x_n - \ell| \geqslant k^{n-n_0}|x_{n_0} - \ell|$ pour $n \geqslant n_0$. Si $x_{n_0} \neq \ell$, alors $\lim\limits_{n\to+\infty} k^{n-n_0}|x_{n_0} - \ell| = +\infty$, ce qui contredit l'hypothèse $(x_n)_{n\in\mathbb{N}}$ converge. On a donc $x_{n_0} = \ell$ et $(x_n)_{n\in\mathbb{N}}$ est stationnaire. (Notons que cette démonstration est valable pour un point fixe répulsif d'une suite récurrente quelconque.)

Mais, pour tout $n \in \mathbb{N}$, on a

$$x_{n+1} = \ell \iff 1 - \lambda x_n^2 = 1 - \lambda\ell^2 \iff x_n^2 = \ell^2 \iff x_n = \ell,$$

car $x_n > 0$. Donc $(x_n)_{n\in\mathbb{N}}$ n'est stationnaire que si $x_0 = \ell$. La suite $(x_n)_{n\in\mathbb{N}}$ ne converge vers ℓ que si $x_0 = \ell$.

Si $x_0 \neq \ell$, $(x_n)_{n\in\mathbb{N}}$ ne converge pas. Les suites $(x_{2n})_{n\in\mathbb{N}}$ et $(x_{2n+1})_{n\in\mathbb{N}}$ ne peuvent pas converger vers ℓ (sinon l'autre sous-suite convergerait vers $f(\ell) = \ell$ et la suite $(x_n)_{n\in\mathbb{N}}$ convergerait elle aussi). Elles ne peuvent pas avoir la même limite, pour les mêmes raisons. On en déduit que l'une converge vers ℓ_1 et l'autre vers ℓ_2, ce qui montre que $f(\ell_1) = \ell_2$ et $f(\ell_2) = \ell_1$.

Notons qu'on a $\ell_1 < \ell < \ell_2$. En effet, l'inégalité $\ell_1 < \ell$ équivaut à $\sqrt{1+4\lambda} + \sqrt{4\lambda - 3} > 2$. Cette dernière inégalité est réalisée car $\lambda > \frac{3}{4}$ et donc $\ell_1 < \ell$. La fonction f étant strictement décroissante, on en déduit $f(\ell_1) > f(\ell)$, c'est-à-dire que $\ell_2 > \ell$.

Si $x_0 < \ell$ alors $x_1 > \ell$ et pour tout entier n, $x_{2n} < \ell$ et $x_{2n+1} > \ell$, car $[0, \ell[$ et $]\ell, 1]$ sont stables par $f \circ f$; $(x_{2n})_{n\in\mathbb{N}}$ converge vers ℓ_1 et $(x_{2n+1})_{n\in\mathbb{N}}$ converge vers ℓ_2. Si $x_0 > \ell$, c'est le contraire. ◁

En $\lambda = \dfrac{3}{4}$, *on a un changement de dynamique appelé* bifurcation.

La définition donnée précédemment d'un point d'ordre 2 *se généralise aisément. Si* α *est un point fixe de* f^p *qui n'est pas point fixe de* f^k *pour* $k < p$, *on dit que* α *est un point* périodique d'ordre p, *ou* p-périodique. *La suite définie par* $u_0 = \alpha$ *et* $u_{n+1} = f(u_n)$ *est périodique de période* p. *L'orbite de* α, *c'est-à-dire* $\{\alpha, f(\alpha), \dots, f^{p-1}(\alpha)\}$, *est appelée* cycle d'ordre p. *On peut alors regarder le comportement de la suite si* u_0 *est choisi proche de* α. *Dans le cas où* u_{np+k} *tend vers* $f^k(\alpha)$ *pour tout* $k \in [\![0, p-1]\!]$ *on dit que l'orbite de* α *est attractive. Il suffit pour cela que* $|(f^p)'(\alpha)| < 1$.

L'étude d'une suite définie par la relation de récurrence $u_{n+1} = 2 - u_n^2$ *prolonge celle qui a été faite dans l'exercice 2.20. En effet, si on pose* $x_n = \dfrac{u_n}{2}$, *on obtient* $x_{n+1} = 1 - 2x_n^2$. *Il s'agit donc d'étudier le cas* $\lambda = 2$.

2.21. Étude d'un système dynamique (2)

Étudier la suite $(u_n)_{n\in\mathbb{N}}$ définie par $u_0 \in \mathbb{R}$ et la relation de récurrence $u_{n+1} = 2 - u_n^2$.

(École polytechnique)

▷ **Solution.**

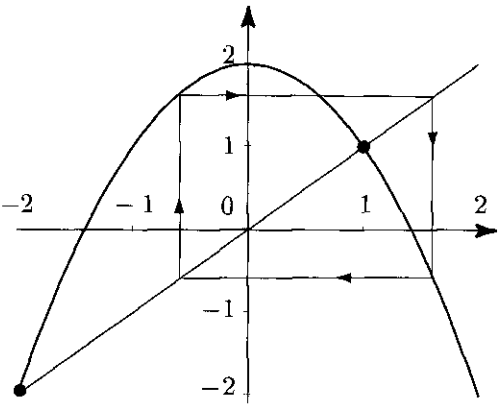
Une étude rapide de la fonction $f : x \longmapsto 2 - x^2$ montre qu'elle possède deux points fixes -2 et 1 et que

$$f([-2, 2]) = [-2, 2], \quad f(]-\infty, -2[) = f(]2, +\infty[) =]-\infty, -2[.$$

Enfin, l'étude de $f(x) - x = -(x-1)(x+2)$ montre que $f(x) < x$ si $x \in]-\infty, -2[\cup]1, +\infty[$ et $f(x) > x$ si $x \in]-2, 1[$.

Si u_0 appartient à $]-\infty, -2[$ ou à $]2, +\infty[$, tous les u_n, pour $n \geqslant 1$, sont dans $]-\infty, -2[$. On a alors $u_{n+1} - u_n = f(u_n) - u_n < 0$. La suite (u_n) décroît et ne peut pas converger, puisque f ne possède pas de points fixes dans $]-\infty, -2[$. La suite (u_n) diverge donc vers $-\infty$.

Si u_0 est dans $[-2, 2]$, tous les termes de la suite sont dans $[-2, 2]$. On remarque qu'on a $|f'(-2)| = 4$ et $|f'(1)| = 2$. Les deux points fixes de f étant répulsifs, la suite ne peut converger ni vers -2, ni vers 1, à moins d'être stationnaire. Elle diverge donc si elle n'est pas stationnaire.



Mais on peut calculer le terme général de la suite u_n. Si on écrit $u_n = -2\cos\theta_n$ (ce qui se révèle plus agréable que $u_n = 2\cos\theta_n$), on a alors $u_{n+1} = 2 - 4\cos^2\theta_n = -2\cos 2\theta_n$. Si on pose donc $u_0 = -2\cos\theta_0$, on obtient, pour tout $n \in \mathbb{N}$, $u_n = -2\cos(2^n\theta_0)$. On peut donc prendre $\theta_n = 2^n\theta_0$. Ceci permet de déterminer les valeurs de θ_0 pour lesquelles (u_n) est stationnaire. Il faut qu'il existe n tel que $u_n = -2$ ou 1, c'est-à-dire $\cos\theta_n = 1$ ou $-\frac{1}{2}$. On trouve $\theta_0 = 2k\pi\left(\frac{1}{2}\right)^n$ ou $\theta_0 = \left(\pm\frac{2\pi}{3} + 2k\pi\right)\left(\frac{1}{2}\right)^n$ $(k \in \mathbb{Z}, n \in \mathbb{N})$.

Supposons que la suite (u_n) ne soit pas stationnaire. La fonction f étant croissante sur $[-2, 0]$, les termes de la suite ne peuvent pas être tous dans cet intervalle à partir d'un certain rang, sinon la suite serait monotone et bornée et convergerait.

La fonction f est décroissante sur $[0, 2]$. Si tous les termes de la suite sont dans $[0, 2]$ à partir d'un certain rang, les deux sous-suites (u_{2n}) et (u_{2n+1}) sont monotones et bornées. Elles convergent vers des limites distinctes, et distinctes de -2 et 1 (si u_{2n} converge vers un point fixe ℓ de f, $u_{2n+1} = f(u_{2n})$ converge vers $f(\ell) = \ell$), points fixes de $f \circ f$. Cherchons les points fixes de $f \circ f$. On obtient, pour tout $x \in [-2, 2]$,

$$f\circ f(x) - x = 2 - (2-x^2)^2 - x = -x^4 + 4x^2 - x - 2 = (-x^2 + 2 - x)(x^2 - x - 1),$$

car $f \circ f(x) - x$ est factorisable par $f(x) - x$. Outre -2 et 1, les points fixes de $f \circ f$ sont $\frac{1+\sqrt{5}}{2}$ et $\frac{1-\sqrt{5}}{2}$. On remarque que $f\left(\frac{1+\sqrt{5}}{2}\right) = \frac{1-\sqrt{5}}{2}$ et que $f\left(\frac{1-\sqrt{5}}{2}\right) = \frac{1+\sqrt{5}}{2}$. Déterminons si ces points fixes de $f \circ f$ sont attractifs ou répulsifs. On obtient

$$(f\circ f)'\left(\frac{1+\sqrt{5}}{2}\right) = (f\circ f)'\left(\frac{1-\sqrt{5}}{2}\right) = f'\left(\frac{1+\sqrt{5}}{2}\right) f'\left(\frac{1-\sqrt{5}}{2}\right) = -4.$$

Ces points sont donc répulsifs. Les deux sous-suites (u_{2n}) et (u_{2n+1}) ne peuvent donc pas converger si elles ne sont pas stationnaires. L'expression de u_n permet de déterminer les valeurs de u_0 pour lesquelles ceci est réalisé. Ce cas étant écarté, les termes de la suite ne peuvent être tous dans l'intervalle $[0,2]$ à partir d'un certain rang. Ils sont alternativement dans $[-2,0]$ et $[0,2]$.

On peut démontrer plus généralement que, pour tout $p \in \mathbb{N}$, f possède des points d'ordre p pour tout p. On peut les déterminer en remarquant que si $x = -2\cos\theta$, alors $f^p(x) = -2\cos(2^p\theta)$ et $f^p(x) = x$ si $(2^p-1)\theta \equiv 0[2\pi]$ ou $(2^p+1)\theta \equiv 0[2\pi]$. On en déduit que $x = -2\cos\dfrac{2\pi}{2^p+1}$ convient. On peut montrer que tous les points de ces cycles sont répulsifs. En effet, si $x = -2\cos\theta_0$ est un tel point, on a

$$(f^p)'(x) = \prod_{k=0}^{p-1} f'(f^k(x)) = \prod_{k=0}^{p-1}(-2f^k(x)) = (-2)^p \prod_{k=0}^{p-1}(-2\cos(2^k\theta_0))$$
$$= 4^p \prod_{k=0}^{p-1}\cos(2^k\theta_0) = 4^p \prod_{k=0}^{p-1}\frac{\sin(2^{k+1}\theta_0)}{2\sin(2^k\theta_0)} = 2^p\frac{\sin(2^p\theta_0)}{\sin\theta_0}.$$

Puisque $\cos(2^p\theta_0) = \cos(\theta_0)$, on en déduit que $|(f^p)'(x)| = 2^p$. Aucune sous-suite de (u_n) ne peut converger vers un tel point d'ordre p, si elle n'est stationnaire. $\lhd$

Le spectaculaire théorème de Sarkowski, qui fait l'objet de l'exercice suivant, montre que lorsque qu'une application f admet un point d'ordre 3, elle a des points d'ordre p pour tout entier $p \geqslant 1$. Cela pourra être constaté sur l'exemple de l'exercice 2.23.

2.22. Théorème de Sarkowski (1964)

Soit I un segment de $\mathbb{R}$ et $f : \mathrm{I} \to \mathrm{I}$ une application continue.

1. Si K est un segment inclus dans $f(\mathrm{I})$, montrer qu'il existe un segment L inclus dans I tel que $\mathrm{K} = f(\mathrm{L})$.

2. On suppose qu'il existe n segments $\mathrm{I}_0, \mathrm{I}_1, \ldots, \mathrm{I}_{n-1}$ inclus dans I tels que $\mathrm{I}_0 \subset f(\mathrm{I}_{n-1})$ et $\mathrm{I}_{k+1} \subset f(\mathrm{I}_k)$ pour $0 \leqslant k \leqslant n-2$. Montrer que $f^n = f \circ \cdots \circ f$ a un point fixe x_0 tel que $f^k(x_0) \in \mathrm{I}_k$ pour tout $k \in [\![0, n-1]\!]$.

3. Si $x \in \mathrm{I}$ vérifie $f^n(x) = x$ et $f^k(x) \neq x$ pour $k \in [\![1, n-1]\!]$ on dit que x est un point n-périodique. Montrer que s'il existe un point 3-périodique, alors il existe un point n-périodique pour tout $n \in \mathbb{N}^*$.

(École normale supérieure)

Solution.

1. Posons $K = [\alpha, \beta]$. Comme $K \subset f(I)$, il existe $(a, b) \in I^2$ tel que $\alpha = f(a)$ et $\beta = f(b)$. Si $\alpha = \beta$, alors $K = \{\alpha\}$ et le singleton $L = \{a\}$ convient. On suppose désormais $\alpha \neq \beta$ et donc $a \neq b$.

• Supposons $a < b$. L'idée est de prendre dans $[a, b]$ un antécédent u de α et un antécédent v de β, tels qu'entre u et v il n'y ait plus d'autre antécédent de α ni de β. Considérons $A = \{x \in [a, b], f(x) = \beta\}$. C'est un fermé non vide (il contient b) et minoré par a. Prenons v le plus petit élément de A. On a $f(v) = \beta$ et $f(t) < \beta$ pour tout $t \in [a, v[$ en vertu du théorème des valeurs intermédiaires (car $f(a) = \alpha < \beta$). Considérons alors $B = \{x \in [a, v], f(x) = \alpha\}$. Par continuité de f, c'est un fermé non vide et majoré par v. Prenons u le plus grand élément de B. On a $u < v$ et $f([u, v]) = [\alpha, \beta]$. Le segment $L = [u, v]$ répond à la question.

• Le cas $a > b$ se traite de la même manière en considérant cette fois $u = \max\{x \in [b, a], f(x) = \beta\}$, puis $v = \min\{x \in [u, a], f(x) = \alpha\}$.

2. Pour avoir des idées, il peut être bon de commencer par regarder les premières valeurs de n.

Commençons par le cas $n = 1$ (notons que cette question figure aussi dans l'exercice 4.8). On dispose, par hypothèse, d'un segment $I_0 = [a, b]$ tel que $I_0 \subset f(I_0)$. En particulier, il existe α et β dans I_0 tels que $f(\alpha) = a$ et $f(\beta) = b$. La fonction $g(x) = f(x) - x$ prend alors des valeurs de signes opposés en α et β et, comme elle est continue, s'annule par le théorème des valeurs intermédiaires. On en déduit l'existence d'un point fixe de f dans l'intervalle I_0.

Regardons maintenant le cas $n = 2$. On a $I_0 \subset f(I_1)$ et $I_1 \subset f(I_0)$. En particulier, $I_0 \subset f(I_1) \subset f^2(I_0)$. Par le cas $n = 1$, le fait que $I_0 \subset f^2(I_0)$ implique l'existence d'un point fixe $x_0 \in I_0$ pour f^2. Mais il n'y a pas de raison *a priori* pour que $f(x_0)$ appartienne à I_1 (on a seulement l'inclusion $I_1 \subset f(I_0)$). On va raffiner un petit peu le choix du point fixe. En fait, d'après la question précédente, il existe un segment $J_1 \subset I_0$ tel que $f(J_1) = I_1$. On a alors $J_1 \subset I_0 \subset f(I_1) = f^2(J_1)$. Le même argument que précédemment, assure l'existence d'un point fixe x_0 de f^2 dans J_1. Et celui-ci vérifie bien que $f(x_0) \in I_1$.

Dans le cas général, on applique la même démarche. Comme on a $f(I_1) \subset I_0$, on peut choisir un segment $J_1 \subset I_0$ tel que $f(J_1) = I_1$. On a alors $I_2 \subset f(I_1) = f^2(J_1)$. On choisit $J_2 \subset J_1$ tel que $f^2(J_2) = I_2$. Et de proche en proche, on construit ainsi une suite finie de segments $J_{n-1} \subset \cdots \subset J_2 \subset J_1 \subset I_0$ telle que $f^k(J_k) = I_k$ pour tout $k \in [\![1, n-1]\!]$. On a enfin, $I_0 \subset f(I_{n-1}) = f^n(J_{n-1})$ de sorte qu'il existe $J_n \subset J_{n-1}$ tel que $f^n(J_n) = I_0$. Comme $J_n \subset f^n(J_n)$, f^n admet un point fixe x_0 dans J_n. Par construction des intervalles J_k, $f^k(x_0) \in I_k$ pour tout $k \in [\![0, n-1]\!]$.

3. Pour simplifier les notations, si I_1, I_2 sont deux segments tels que

$I_1 \subset f(I_2)$, on écrira $I_1 \to I_2$. La question précédente a donc montré que si on a un cycle $I_0 \to I_1 \to \cdots \to I_{n-1} \to I_0$, f^n admet un point fixe x_0 dans I_0 vérifiant $f^k(x_0) \in I_k$ pour tout $k \in [\![0, n-1]\!]$.

Par hypothèse, f admet un point 3-périodique a. Posons $b = f(a)$ et $c = f(b) = f^2(a)$. Les points b et c sont aussi 3-périodiques et, quitte à remplacer a par b ou c, on suppose que $a = \min(a, b, c)$. Deux éventualités se présentent selon la disposition de b et c.

• Supposons tout d'abord $a < b < c$ et posons $I_0 = [a, b]$ et $I_1 = [b, c]$. Comme $f(a) = b$ et $f(b) = c$ on a $I_1 \subset f(I_0)$, c'est-à-dire $I_1 \to I_0$. De la même manière, on a aussi $I_0 \to I_1$ et $I_1 \to I_1$. Cette dernière inclusion montre déjà que f admet un point fixe dans I_1. De même, le cycle $I_0 \to I_1 \to I_0$ montre que f^2 admet un point fixe x_0 dans I_0 tel que $f(x_0) \in I_1$. Comme x_0 ne peut pas être égal à b, $x_0 \notin I_1$ et $f(x_0) \neq x_0$. Ainsi, x_0 est un point 2-périodique. Soit maintenant $n \geqslant 4$. On écrit le cycle $I_0 \to I_1 \to I_1 \to \cdots \to I_1 \to I_0$, où l'intervalle I_1 figure $n-1$ fois. D'après la question précédente, f^n admet un point fixe x dans I_0 tel que $f^k(x) \in I_1$ pour $k < n$. Comme précédemment, x ne peut pas être égal à b et est donc un point n-périodique.

• Regardons maintenant le cas où $a < c < b$. On pose $I_0 = [a, c]$ et $I_1 = [c, b]$. On a cette fois $I_1 \to I_0$, $I_0 \to I_0$ et $I_0 \to I_1$. On peut donc reprendre les idées précédentes en échangeant I_0 et I_1.

On a donc montré dans tous les cas que f admet des points n-périodiques pour tout $n \geqslant 1$. ◁

2.23. La fonction «tente»

On note $T : [0,1] \to [0,1]$ l'application définie par $T(x) = 2x$ pour $x \in \left[0, \frac{1}{2}\right]$ et $T(x) = 2 - 2x$ pour $x \in \left[\frac{1}{2}, 1\right]$ et, pour $n \in \mathbb{N}^*$, T^n la n-ième itérée de T.

1. Tracer les graphes de T, T^2, T^3.

2. Combien T^n a-t-elle de points fixes ?

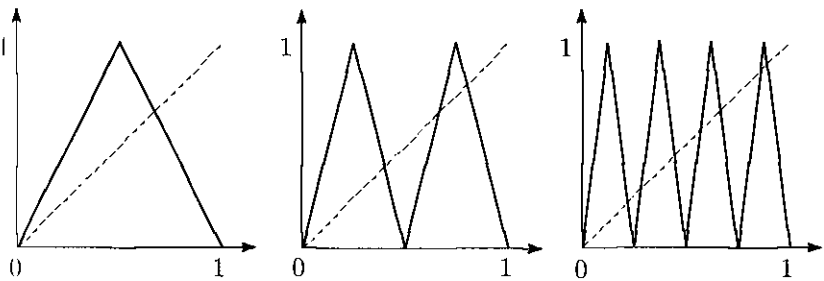
3. Soit $(x, y) \in [0,1]^2$ et $\varepsilon > 0$. Montrer qu'il existe $z \in [0,1]$ et $n \in \mathbb{N}^*$ tels que $|z - x| \leqslant \varepsilon$ et $|T^n(z) - y| \leqslant \varepsilon$.

(École polytechnique)

▷ **Solution.**

1. Voir ci-après les graphes demandés. Celui de T s'obtient immédiatement et explique le nom donné à la fonction T. Lorsque x parcourt l'intervalle $\left[0, \frac{1}{2}\right]$, $y = T(x)$ décrit tout le segment $[0,1]$ (deux fois plus vite). Il en résulte que le graphe de T^2 sur $\left[0, \frac{1}{2}\right]$ est aussi un pic mais de

base $\frac{1}{2}$. Quand x parcourt $\left[\frac{1}{2}, 1\right]$, $y = \mathrm{T}(x)$ décrit à nouveau l'intervalle $[0, 1]$ (en décroissant) ce qui donne un second pic pour le graphe de T^2. De la même manière, on voit que le graphe de T^3 est formé de 4 pics de base $\frac{1}{4}$.



2. Le graphe de T^n est constitué de 2^{n-1} pics de base $\frac{1}{2^{n-1}}$. Plus précisément, on montre facilement, par récurrence sur $n \in \mathbb{N}^*$, que si $0 \leqslant k \leqslant 2^{n-1} - 1$, on a

$$\begin{cases} \mathrm{T}^n(x) = 2^n(x - 2k) & \text{si } x \in \left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right] \\ \mathrm{T}^n(x) = 2^n(2k + 2 - x) & \text{si } x \in \left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right]. \end{cases}$$

Soit g_n la fonction $x \longmapsto \mathrm{T}^n(x) - x$. Sur chaque intervalle $\left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right]$ la fonction g_n est une fonction affine de coefficient directeur $2^n - 1 > 0$; elle est croissante. On a, de plus, $g_n\left(\frac{2k}{2^n}\right) = -\frac{2k}{2^n} \leqslant 0$ et $g_n\left(\frac{2k+1}{2^n}\right) = 1 - \frac{2k+1}{2^n} \geqslant \frac{1}{2^n} > 0$. On en déduit que g_n s'annule une fois sur chaque intervalle $\left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right[$. On montre de même que g_n s'annule une fois sur chaque intervalle $\left]\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right[$. Au total g_n s'annule 2^n fois et T^n possède donc 2^n points fixes.

Le plus petit point fixe non nul de T^n *est un point* n*-périodique.*

3. Soit $x \in [0, 1]$. Pour tout $n \in \mathbb{N}^*$, il existe un entier k compris entre 0 et $2^{n-1} - 1$ tel que x appartienne à un intervalle de la forme $\left[\frac{2k}{2^n}, \frac{2k+1}{2^n}\right]$ ou $\left[\frac{2k+1}{2^n}, \frac{2k+2}{2^n}\right]$. Notons I cet intervalle. L'application T^n réalise une bijection de I sur $[0, 1]$. Si $y \in [0, 1]$, il existe $z \in \mathrm{I}$ tel que $\mathrm{T}^n(z) = y$. On a alors $|\mathrm{T}^n(z) - y| = 0 \leqslant \varepsilon$ et $|z - x| \leqslant \frac{1}{2^n} \leqslant \varepsilon$ si n est choisi assez grand. ◁

La dernière question montre que T *est topologiquement transitive : dans le voisinage de tout point* x*, on peut trouver des points* z *dont l'orbite va couper tout ouvert fixé à l'avance. On observe aussi que l'ensemble des points périodiques de* T *est dense dans* $[0,1]$*. Ces deux conditions permettent de qualifier le système dynamique associé à* T *de chaotique. Dans la définition donnée par Devaney*[3] *figurait une troisième condition, la sensibilité aux conditions initiales qu'on peut formuler ainsi pour* $f : \mathrm{I} \to \mathrm{I}$ *: il existe* $\delta > 0$ *tel que, pour tout* $x \in \mathrm{I}$ *et tout* $\varepsilon > 0$*, il existe* $y \in \mathrm{I}$ *et* $n \in \mathbb{N}$ *tels que* $|x - y| < \varepsilon$ *et* $|f^n(x) - f^n(y)| > \delta$*. Il a été démontré en* 1992 *que, pour une fonction continue, cette condition est une conséquence des deux premières*[4]*.*

Dans l'exercice suivant, on cherche un équivalent du terme général d'une suite tendant vers 0*. La méthode est très classique et sera utilisée dans plusieurs des exercices qui suivent. Elle utilise le théorème de Cesàro sous la forme particulière appelée* lemme de l'escalier *: si* $\lim\limits_{n\to+\infty} u_{n+1} - u_n = \ell$*, alors* $\lim\limits_{n\to+\infty} \dfrac{u_n}{n} = \ell$ *(on applique le théorème de Cesàro à la suite* $(u_{n+1} - u_n)_{n\in\mathbb{N}}$*).*

2.24. Un système dynamique discret et son analogue continu

Soit $f : \mathbb{R} \longmapsto \mathbb{R}$ de classe $\mathcal{C}^1$ telle que $f(0) = f(1) = 0$, avec $f'(0) \in]-1, 0[$ et $-x < f(x) < 0$ pour tout $x \in]0, 1[$.

1. On considère la suite (x_n) définie par $x_0 \in]0, 1[$ et, pour tout $n \in \mathbb{N}$, $x_{n+1} - x_n = f(x_n)$.

a. Étudier la suite (x_n).

b. Montrer que, pour n assez grand, il existe un unique entier naturel $\varphi(n)$ tel que $x_{\varphi(n)+1} \leqslant \dfrac{1}{n} < x_{\varphi(n)}$.

c. Démontrer qu'il existe $\mathrm{C} \neq 0$ tel que $\varphi(n) \underset{n\to\infty}{\sim} \mathrm{C} \ln n$.

2. On passe au cas continu. On considère la solution de l'équation différentielle $x' = f(x)$ prenant en 0 la valeur $x(0) \in]0, 1[$.

Montrer que x est définie sur $[0, +\infty[$ et que, pour n assez grand, il existe un unique $t_n \geqslant 0$ tel que $x(t_n) = \dfrac{1}{n}$. Déterminer un équivalent de t_n.

(École normale supérieure)

3. DEVANEY (R.L.), *An introduction to chaotic dynamical systems*, Addison-Wesley, 2e édition, 1989.

4. On en trouvera la démonstration dans HOLMGREN (R.A), *A first course in discrete dynamical systems*, Springer, 1996, p. 81-83.

Solution.

1. a. Si $x_n \in]0,1[$, alors on a $0 < x_n + f(x_n) < x_n$, c'est-à-dire $0 < x_{n+1} < x_n < 1$. Une récurrence immédiate montre donc que la suite (x_n) est à termes dans $]0,1[$ et strictement décroissante. On en déduit qu'elle converge vers $\ell \in [0,1[$. La limite ℓ vérifie $\ell = \ell + f(\ell)$ et donc $f(\ell) = 0$. Par hypothèse, cela implique $\ell = 0$. La suite (x_n) converge vers 0.

b. Si n vérifie $\dfrac{1}{n} < x_0$, l'ensemble des entiers k tels que $\dfrac{1}{n} < x_k$ est non vide et majoré, puisque $\lim\limits_{k\to+\infty} x_k = 0$. Si on note $\varphi(n)$ le plus grand élément de cet ensemble, on a bien $x_{\varphi(n)+1} \leqslant \dfrac{1}{n} < x_{\varphi(n)}$. L'entier $\varphi(n)$ vérifiant ces inégalités est unique car, la suite (x_n) étant strictement décroissante, les intervalles $[x_{k+1}, x_k[$ $(k \in \mathbb{N})$ sont disjoints.

c. La suite $(\varphi(n))$ tend vers $+\infty$, puisque $\dfrac{1}{n}$ tend vers 0. Cherchons, pour commencer un équivalent de x_n. On a $\dfrac{x_{n+1}}{x_n} = 1 + \dfrac{f(x_n)}{x_n}$ pour $n \in \mathbb{N}$. On en déduit, puisque $f(0) = 0$ et que (x_n) converge vers 0, que $\lim\limits_{n\to+\infty} \dfrac{x_{n+1}}{x_n} = 1 + f'(0)$. Cette limite étant strictement positive, on peut écrire $\lim\limits_{n\to+\infty} \ln x_{n+1} - \ln x_n = \ln(1 + f'(0)) \neq 0$, puisque $f'(0) \neq 0$. Le théorème de Cesàro conduit à

$$\ln x_n \underset{n\to\infty}{\sim} \ln x_n - \ln x_0 \underset{n\to\infty}{\sim} n \ln(1 + f'(0)).$$

De la double inégalité $\ln(x_{\varphi(n)+1}) \leqslant -\ln n < \ln(x_{\varphi(n)})$, on déduit que $-\ln n \underset{n\to\infty}{\sim} \varphi(n) \ln(1 + f'(0))$, puisque les termes extrêmes de ces inégalités sont équivalents à $\varphi(n) \ln(1 + f'(0))$, et donc

$$\boxed{\varphi(n) \underset{n\to\infty}{\sim} -\frac{\ln n}{\ln(1 + f'(0))}.}$$

2. La fonction $(t,x) \longmapsto f(x)$ étant de classe $\mathcal{C}^1$ sur $\mathbb{R}\times]0,1[$, le théorème de Cauchy-Lipschitz s'applique et il existe une solution maximale unique x définie sur un intervalle ouvert contenant 0 et prenant en 0 la valeur $x(0)$.

Montrons pour commencer que x prend ses valeurs dans $]0,1[$. Démontrons d'abord qu'elle ne prend ni la valeur 0, ni la valeur 1. S'il existe t tel que $x(t) = 0$, alors la fonction constante $t \longmapsto 0$, qui est solution de l'équation différentielle, puisque $f(0) = 0$, coïncide avec x en t. D'après le théorème de Cauchy-Lipschitz, ces deux solutions sont égales. C'est faux, car $x(0) \neq 0$. La démonstration est la même pour la valeur 1. Puisque $x(0) \in]0,1[$, la fonction x prend ses valeurs dans $]0,1[$, car sinon, d'après le théorème des valeurs intermédiaires, elle prendrait la valeur 0 ou la valeur 1.

Notons que, pour tout $t \in \mathrm{I}$, $x'(t) = f(x(t)) < 0$, car $x(t) \in]0,1[$. La fonction x est donc strictement décroissante.

Montrons maintenant que l'ensemble de définition de cette fonction contient $[0,+\infty[$. Raisonnons par l'absurde et supposons que cet ensemble de définition est $\mathrm{I} =]a,b[$, avec $-\infty \leqslant a < 0$ et $b > 0$. Étant décroissante et minorée, x possède une limite finie ℓ appartenant à $[0,1[$ en b. À son tour, $x' = f(x)$ possède une limite finie $f(\ell)$ en b. On peut prolonger la solution de l'équation différentielle en b, ce qui contredit le fait qu'une solution maximale est définie sur un intervalle ouvert. La fonction est donc définie sur $[0,+\infty[$.

La fonction x, strictement décroissante et minorée, possède une limite finie $\ell \in [0,1[$ en $+\infty$ et x' a pour limite $f(\ell)$. Si ℓ n'est pas nul, $f(\ell)$ n'est pas nul également. La fonction x' n'est donc pas intégrable sur $[0,+\infty[$, ce qui contredit le fait que x a une limite finie en $+\infty$. On a donc $\ell = 0$. Ainsi la fonction x réalise une bijection de $[0,+\infty[$ sur $]0,x(0)]$. On en déduit que, dès que $\dfrac{1}{n}$ est inférieur à $x(0)$, il existe un unique $t_n \geqslant 0$ tel que $x(t_n) = \dfrac{1}{n}$. Puisque $\dfrac{1}{n}$ tend vers 0 quand n tend vers $+\infty$, t_n tend vers $+\infty$.

Quand t tend vers $+\infty$, $x(t)$ tend vers 0. Cherchons un équivalent de $x(t)$. Comme $f(0) = 0$ et $f'(0) \neq 0$, on a

$$x'(t) = f(x(t)) \underset{t\to+\infty}{\sim} f'(0)x(t),$$

c'est-à-dire que $(\ln \circ x)'(t) \underset{t\to+\infty}{\sim} f'(0)$. La fonction $t \longmapsto f'(0)$ est négative et non-intégrable sur $[0,+\infty[$; on obtient, par intégration des relations d'équivalence, $\ln(x(t)) \underset{t\to+\infty}{\sim} tf'(0)$. En particulier, on a $-\ln n \underset{n\to+\infty}{\sim} t_n f'(0)$ et donc

$$\boxed{t_n \underset{n\to+\infty}{\sim} -\frac{\ln n}{f'(0)}.} \quad \lhd$$

Explicitons le rapport entre $x_{n+1} = x_n + f(x_n)$ et l'équation différentielle $x' = f(x)$. On suppose connu $x(t_0)$ et on fixe un pas h. La méthode d'Euler de résolution de l'équation différentielle consiste à confondre la courbe de x sur l'intervalle $[t_0, t_0 + h]$ avec sa tangente au point d'abscisse t_0. La valeur approchée de $x(t_0 + h)$ obtenue est $x(t_0) + hx'(t_0) = x(t_0) + hf(x(t_0))$. On peut réitérer le procédé à partir de $t_0 + h$, puis $t_0 + 2h, \ldots, t_0 + nh$. On prend $h = 1$ et $t_0 = 0$. Si on note x_n la valeur approchée de $x(n)$ ainsi obtenue, on a, pour tout $n \in \mathbb{N}$, $x_{n+1} = x_n + f(x_n)$.

2.25. Équivalent d'une suite définie par récurrence

Soit $c > 0$ et $f : [0, c] \to [0, c]$ une fonction continue, admettant en 0 un développement asymptotique de la forme

$$f(x) = x - ax^\alpha + o(x^\alpha) \text{ où } a > 0 \text{ et } \alpha > 1.$$

1. Montrer que, pour u_0 assez petit, la suite (u_n) définie par $u_{n+1} = f(u_n)$ pour tout $n \in \mathbb{N}$, converge vers 0.

2. Déterminer alors un équivalent de u_n.

3. Traiter l'exemple de $x \mapsto \sin x$ et $x \longmapsto \ln(1+x)$.

(École polytechnique)

▷ **Solution.**

1. Par continuité de f on a $f(0) = 0$ et vu le développement asymptotique, il existe $\eta > 0$ tel que $f(x) < x$ sur $]0, \eta]$. Si on choisit u_0 dans cet intervalle, la suite (u_n) est décroissante et positive, donc converge vers 0 qui est le seul point fixe de f dans $[0, \eta]$.

2. Déterminons β tel que la suite $(u_{n+1}^\beta - u_n^\beta)$ ait une limite non nulle. On a, au voisinage de 0, puisque $\alpha > 1$,

$$\begin{aligned}(f(x))^\beta - x^\beta &= (x - ax^\alpha + o(x^\alpha))^\beta - x^\beta \\ &= x^\beta((1 - ax^{\alpha-1} + o(x^{\alpha-1}))^\beta - 1) \\ &= x^\beta(-a\beta x^{\alpha-1} + o(x^{\alpha-1})) \underset{x\to 0}{\sim} -a\beta x^{\alpha+\beta-1}.\end{aligned}$$

Si on prend $\beta = 1 - \alpha$, on obtient $\lim\limits_{x\to 0}(f(x))^\beta - x^\beta = a(\alpha - 1)$. En particulier, puisque la suite (u_n) converge vers 0, on a

$$\lim_{n\to+\infty} u_{n+1}^{1-\alpha} - u_n^{1-\alpha} = a(\alpha - 1).$$

Le théorème de Cesàro montre alors que

$$u_n^{1-\alpha} \underset{n\to+\infty}{\sim} u_n^{1-\alpha} - u_0^{1-\alpha} \underset{n\to+\infty}{\sim} na(\alpha - 1)$$

d'où l'on déduit finalement

$$u_n \underset{n\to+\infty}{\sim} (na(\alpha - 1))^{\frac{1}{1-\alpha}}.$$

3. Considérons le cas $f(x) = \sin x$. On se place sur l'intervalle $\left[0, \frac{\pi}{2}\right]$. Il est stable par la fonction sinus. Pour $x > 0$, on a $\sin x < x$. On en déduit que si $u_0 \in \left[0, \frac{\pi}{2}\right]$, la suite (u_n) décroît et converge vers l'unique point fixe de la fonction sinus, 0. Le développement limité en 0

$$\sin x = x - \frac{x^3}{6} + o(x^3) \text{ (cas } a = \frac{1}{6} \text{ et } \alpha = 3)$$

conduit à

$$u_n \underset{n\to\infty}{\sim} \left(\frac{n}{3}\right)^{-\frac{1}{2}} = \sqrt{\frac{3}{n}}.$$

Traitons, de même, le cas $f : x \longmapsto \ln(1+x)$. L'intervalle $\mathbb{R}_+$ est stable par f. Pour tout $x > 0$, on a $\ln(1+x) < x$. On en déduit que si $u_0 \in \mathbb{R}_+$, la suite (u_n) décroît et converge vers 0, unique point fixe de f. Cette fois, le développement limité en 0

$$\ln(1+x) = x - \frac{x^2}{2} + o(x^2) \ \text{(cas } a = \frac{1}{2} \ \text{ et } \ \alpha = 2)$$

donne

$$u_n \underset{n\to\infty}{\sim} \frac{2}{n}. \quad \lhd$$

Le résultat précédent fait penser à la formule des accroissements finis. En effet, l'hypothèse vérifiée par la suite (u_n) est

$$u_{n+1} - u_n = f(u_n) - u_n \underset{n\to\infty}{\sim} au_n^{\alpha},$$

c'est-à-dire

$$(u_{n+1} - u_n)u_n^{-\alpha} \underset{n\to\infty}{\sim} a.$$

La fonction $x \longmapsto x^{-\alpha}$ étant la dérivée à une constante près de la fonction $x \longmapsto x^{1-\alpha}$, il est naturel d'étudier $u_{n+1}^{1-\alpha} - u_n^{1-\alpha}$. Nous retrouverons cette méthode dans plusieurs exercices, y compris dans des cas où u_n tend vers $+\infty$.

L'exercice suivant où l'on cherche encore un équivalent d'une suite définie par récurrence illustre la remarque précédente.

2.26. Développement asymptotique d'une suite récurrente (1)

On considère la suite $(u_n)_{n\in\mathbb{N}}$ définie par $u_0 > 0$ et, pour $n \geqslant 0$, la relation $u_{n+1} = u_n + \dfrac{1}{u_n}$.

1. Donner un équivalent de u_n lorsque n tend vers l'infini.

2. On prend $u_0 = 5$. Montrer que $u_{1000} \in [45; 45{,}1]$.

(École polytechnique)

Solution.

1. Il est clair que l'on a $u_n \geqslant 1$ pour $n \geqslant 1$ et que la suite (u_n) est croissante. On en déduit que $(u_n)_{n\in\mathbb{N}}$ diverge vers $+\infty$ ou converge vers un réel $\ell \geqslant 1$. Dans ce dernier cas, vue la continuité de f, le réel ℓ serait point fixe de f et on aurait $\ell = \ell + \dfrac{1}{\ell}$, *i.e.* $\dfrac{1}{\ell} = 0$. C'est impossible. On en déduit que $\lim\limits_{n\to+\infty} u_n = +\infty$. Pour déterminer un équivalent de (u_n), on part de l'égalité $u_{n+1}^2 = u_n^2 + 2 + \dfrac{1}{u_n^2}$. Puisque $\lim\limits_{n\to+\infty} \dfrac{1}{u_n^2} = 0$, on a

$$u_{n+1}^2 - u_n^2 = 2 + \frac{1}{u_n^2} \underset{n\to\infty}{\sim} 2.$$

On peut écrire, d'après le théorème de Cesàro,

$$\sum_{k=1}^{n} (u_k^2 - u_{k-1}^2) \underset{n\to\infty}{\sim} 2n,$$

ce qui donne

$$u_n^2 \underset{n\to\infty}{\sim} u_n^2 - u_1^2 \underset{n\to\infty}{\sim} 2n.$$

Nous obtenons donc

$$\boxed{u_n \underset{n\to\infty}{\sim} \sqrt{2n}.}$$

2. Pour démontrer l'encadrement demandé, on va préciser le développement asymptotique de $u_{n+1}^2 - u_n^2$. En sommant la relation $u_{k+1}^2 - u_k^2 = 2 + \dfrac{1}{u_k^2}$ pour les entiers k de 0 à $n-1$, on obtient

$$u_n^2 = u_0^2 + 2n + \sum_{k=0}^{n-1} \frac{1}{u_k^2}.$$

On va déduire de cette égalité un encadrement de u_n. On a déjà l'inégalité $u_n^2 > u_0^2 + 2n$, ce qui conduit à $u_{1000}^2 > 25 + 2000 = 2025 = 45^2$ et donc à la minoration voulue. Par ailleurs, cette inégalité nous permet de majorer les termes $\dfrac{1}{u_k^2}$ par $\dfrac{1}{u_0^2 + 2k}$. On obtient alors

$$u_n^2 < u_0^2 + 2n + \sum_{k=0}^{n-1} \frac{1}{u_0^2 + 2k} \leqslant u_0^2 + 2n + \sum_{k=1}^{2n} \frac{1}{k} \leqslant u_0^2 + 2n + \ln(2n),$$

puisqu'il est bien connu que la somme partielle de la série harmonique $H_n = 1 + \dfrac{1}{2} + \cdots + \dfrac{1}{n}$ est majorée par $\ln n$. En prenant $n = 1000$, on obtient

$$u_{1000} < \sqrt{2025 + \ln 2000} \simeq 45{,}08\ldots < 45{,}1,$$

ce qui établit la deuxième partie de l'inégalité demandée. ◁

Il s'agit à nouveau, dans l'exercice suivant, de calculer un développement asymptotique d'une suite définie par une relation de récurrence. On retiendra la méthode qui consiste à construire successivement les termes du développement asymptotique. Partant de $u_{n+1} - u_n \underset{n\to\infty}{\sim} a_n$, *on en déduit, en sommant, un équivalent de* u_n, *puis on cherche un équivalent de* $u_{n+1} - u_n - a_n$ *et on somme de nouveau, et ainsi de suite. Bien entendu le seul théorème de Cesàro ne suffit plus. On utilise alors le théorème de sommation des équivalents : si* (a_n) *et* (b_n) *sont deux suites équivalentes à termes positifs telles que* $\sum a_n$ *diverge, alors* $\sum_{k=0}^{n} a_k \underset{n\to\infty}{\sim} \sum_{k=0}^{n} b_k$ *(voir aussi pages 150 à 155).*

2.27. Développement asymptotique d'une suite récurrente (2)

Soit $(u_n)_{n\in\mathbb{N}}$ définie par $u_0 \in \mathbb{R}$ et la relation $u_{n+1} = u_n + e^{-u_n}$ pour tout $n \in \mathbb{N}$. Donner les deux premiers termes du développement asymptotique de u_n.

(École polytechnique)

▷ **Solution.**

La suite $(u_n)_{n\in\mathbb{N}}$ est croissante. Si elle convergeait vers ℓ, on aurait par continuité de l'exponentielle, $\lim\limits_{n\to+\infty} e^{-u_n} = e^{-\ell} = 0$, ce qui est impossible. La suite $(u_n)_{n\in\mathbb{N}}$ tend donc vers $+\infty$.

Posons, pour tout entier naturel n, $v_n = e^{u_n}$. On obtient

$$v_{n+1} = e^{u_{n+1}} = e^{u_n + \frac{1}{v_n}} = v_n e^{\frac{1}{v_n}}.$$

La suite $(v_n)_{n\in\mathbb{N}}$ est à termes strictement positifs, croît et diverge vers $+\infty$.

Selon une méthode classique, on cherche successivement les termes du développement asymptotique de la suite $(v_n)_{n\in\mathbb{N}}$. On part de

$$v_{n+1} = v_n\left(1 + \frac{1}{v_n} + \frac{1}{2v_n^2} + o\left(\frac{1}{v_n^2}\right)\right) = v_n + 1 + \frac{1}{2v_n} + o\left(\frac{1}{v_n}\right).$$

On en déduit que $\lim\limits_{n\to+\infty} (v_{n+1} - v_n) = 1$. Le théorème de Cesàro nous donne

$$v_{n+1} - v_0 \underset{n\to+\infty}{\sim} n \text{ et donc } v_n \underset{n\to+\infty}{\sim} n.$$

Pour obtenir le deuxième terme du développement asymptotique de $(v_n)_{n\in\mathbb{N}}$, on reprend l'expression de $v_{n+1} - v_n$ qui conduit à

$$v_{n+1} - v_n - 1 \underset{n\to+\infty}{\sim} \frac{1}{2v_n} \underset{n\to+\infty}{\sim} \frac{1}{2n}.$$

En sommant de nouveau, on obtient

$$v_n - v_1 - (n-1) \underset{n\to+\infty}{\sim} \sum_{k=1}^{n} \frac{1}{2k} \underset{n\to+\infty}{\sim} \frac{1}{2} \ln n,$$

ce qui peut encore s'écrire $v_n = n + \frac{1}{2} \ln n + o(\ln n)$. On a alors

$$u_n = \ln v_n = \ln n + \ln\left(1 + \frac{1}{2}\frac{\ln n}{n} + o\left(\frac{\ln n}{n}\right)\right)$$

et finalement

$$\boxed{u_n = \ln n + \frac{\ln n}{2n} + o\left(\frac{\ln n}{n}\right).} \quad \triangleleft$$

Les suites étudiées dans les exercices suivants vérifient des relations de récurrence plus compliquées que les systèmes dynamiques : x_{n+1} s'exprime en fonction de x_n et de n. On peut les rapprocher des équations différentielles $x' = f(x,t)$.

Le premier est assez facile car il est possible d'expliciter u_n en fonction de n.

2.28. Détermination d'une suite récurrente

Existe-t-il une suite $(u_n)_{n\geqslant 1}$ de réels telle que, pour tout $n \in \mathbb{N}^*$, on ait $u_n u_{n+1} = n$ et $\lim\limits_{n\to+\infty} \dfrac{u_{n+1}}{u_n} = 1$?

(École polytechnique)

Solution.

Il est clair que la donnée de $u_1 \neq 0$ et de la relation $u_n u_{n+1} = n$, pour tout $n \in \mathbb{N}^*$, définit la suite (u_n) de manière unique. Le problème est donc de choisir u_1 pour que $\lim\limits_{n\to+\infty} \dfrac{u_{n+1}}{u_n} = 1$.

On a pour $n \in \mathbb{N}^*$, $u_{2n} = \dfrac{2n-1}{u_{2n-1}} = \dfrac{2n-1}{\left(\dfrac{2n-2}{u_{2n-2}}\right)} = \dfrac{2n-1}{2n-2} u_{2n-2}$. Par une récurrence immédiate, on obtient

$$u_{2n} = \frac{2n-1}{2n-2} \cdot \frac{2n-3}{2n-4} \cdots \frac{5}{4} \cdot \frac{3}{2} \cdot u_2.$$

En multipliant par $(2n-2)\cdot(2n-4)\cdots 4 \cdot 2$ le numérateur et le

dénominateur, on trouve

$$u_{2n} = \frac{(2n-1)!}{[2(n-1)]^2[2(n-2)]^2\dots[2.2]^2[2.1]^2}u_2 = \frac{(2n-1)!}{2^{2(n-1)}(n-1)!^2}u_2$$
$$= \frac{2n-1}{4^{n-1}}\mathrm{C}^{n-1}_{2(n-1)}u_2,$$

avec $u_2 = \dfrac{1}{u_1}\cdot$ On en déduit $u_{2n+2} = \dfrac{2n+1}{4^n}\mathrm{C}^n_{2n}u_2$. On obtient de même

$$u_{2n+1} = \frac{2n}{2n-1}u_{2n-1} = \frac{2n}{2n-1}\cdot\frac{2n-2}{2n-3}\cdots\frac{2}{1}u_1 = \frac{2^{2n}n!^2}{(2n)!}u_1 = \frac{4^n}{\mathrm{C}^n_{2n}}u_1.$$

Avec la formule de Stirling qui donne $n! \underset{n\to\infty}{\sim} \sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n$, on obtient au voisinage de l'infini,

$$\frac{u_{2n+2}}{u_{2n+1}} = \frac{(2n+1)(\mathrm{C}^n_{2n})^2}{4^{2n}}\frac{u_2}{u_1} \sim (2n+1)\frac{\left(\sqrt{4\pi n}\left(\dfrac{2n}{e}\right)^{2n}\right)^2}{4^{2n}\left(\sqrt{2\pi n}\left(\dfrac{n}{e}\right)^n\right)^4}\frac{1}{u_1^2}$$
$$\sim \frac{4\pi n(2n+1)}{(2\pi n)^2}\frac{1}{u_1^2} \sim \frac{2}{\pi u_1^2},$$
$$\text{et}\quad \frac{u_{2n+1}}{u_{2n}} = \frac{2n+2}{2n+1}\cdot\frac{u_{2n+1}}{u_{2n+2}} \sim \frac{\pi u_1^2}{2}.$$

La condition $\lim\limits_{n\to+\infty}\dfrac{u_{n+1}}{u_n} = 1$ équivaut à $\lim\limits_{n\to+\infty}\dfrac{u_{2n+2}}{u_{2n+1}} = 1$ et $\lim\limits_{n\to+\infty}\dfrac{u_{2n+1}}{u_{2n}} = 1$, ce qui est réalisé si et seulement $u_1 = \pm\sqrt{\dfrac{2}{\pi}}\cdot$ Il existe donc deux suites ayant les propriétés voulues. Elles sont opposées. ◁

2.29. Étude de $x_{n+1} = 1 + \dfrac{n}{x_n}$

Étudier la suite $(x_n)_{n\in\mathbb{N}^*}$ définie par $x_1 = 1$ et $x_{n+1} = 1 + \dfrac{n}{x_n}$ pour tout $n \geqslant 1$.

(École polytechnique)

▷ **Solution.**

La suite est à termes strictement positifs. Le calcul des premiers termes, $x_2 = 2$, $x_3 = 2$, $x_4 = 2{,}5$, $x_5 = 2{,}6\ldots$ semble montrer qu'elle est croissante. On étudie l'inégalité $x_{n+1} \geqslant x_n$. On a les équivalences suivantes :

$$x_{n+1} \geqslant x_n \iff 1 + \frac{n}{x_n} \geqslant x_n \iff x_n^2 - x_n - n \leqslant 0.$$

Le trinôme $x^2 - x - n$ possède deux racines $\dfrac{1 - \sqrt{1+4n}}{2}$ et $\dfrac{1 + \sqrt{1+4n}}{2}$. Compte tenu de $x_n > 0$, on obtient

$$x_{n+1} \geqslant x_n \iff x_n \leqslant \frac{1 + \sqrt{1+4n}}{2}.$$

On remarque alors que

$$\begin{aligned} x_n \leqslant \frac{1 + \sqrt{1+4n}}{2} &\iff x_{n+1} \geqslant 1 + \frac{2n}{1 + \sqrt{1+4n}} \\ &\iff x_{n+1} \geqslant \frac{1 + \sqrt{1+4n}}{2}. \end{aligned}$$

Il semble plausible de démontrer que pour tout n de $\mathbb{N}^*$, on a

$$\frac{1 + \sqrt{4n-3}}{2} \leqslant x_n \leqslant \frac{1 + \sqrt{1+4n}}{2}.$$

Montrons-le par récurrence. C'est vrai pour $n = 1$, car $x_1 = 1$ et si l'on suppose que $\dfrac{1 + \sqrt{4n-3}}{2} \leqslant x_n \leqslant \dfrac{1 + \sqrt{1+4n}}{2}$, alors on a, d'après ce qui précède, $x_{n+1} \geqslant \dfrac{1 + \sqrt{1+4n}}{2}$. On a d'autre part

$$x_{n+1} \leqslant 1 + \frac{2n}{\sqrt{4n-3}+1} \leqslant 1 + \frac{2n(\sqrt{4n-3}-1)}{4(n-1)} \leqslant \frac{n\sqrt{4n-3}+n-2}{2(n-1)}.$$

Pour montrer que $x_{n+1} \leqslant \dfrac{1 + \sqrt{5+4n}}{2}$, il suffit de démontrer que

$$\frac{n\sqrt{4n-3}+n-2}{2(n-1)} \leqslant \frac{1 + \sqrt{5+4n}}{2} \quad \text{i.e. } n\sqrt{4n-3} \leqslant (n-1)\sqrt{4n+5} + 1.$$

En élevant au carré et en simplifiant par $n - 1$, on montre que cette inégalité équivaut à $3 \leqslant \sqrt{4n+5}$, ce qui est vérifié pour $n \geqslant 1$. La récurrence est terminée. On a donc, pour tout $n \geqslant 1$,

$$\frac{1 + \sqrt{4n-3}}{2} \leqslant x_n \leqslant \frac{1 + \sqrt{1+4n}}{2}.$$

On en déduit que la suite $(x_n)_{n \in \mathbb{N}^*}$ est croissante, que $\lim\limits_{n \to +\infty} x_n = +\infty$ et que $\lim\limits_{n \to \infty} x_n - \sqrt{n} = \dfrac{1}{2}$. En effet, on a

$$\frac{\sqrt{4n+1}}{2} - \sqrt{n} = \sqrt{n + \frac{1}{2}} - \sqrt{n} = \frac{\frac{1}{2}}{\sqrt{n} + \sqrt{n + \frac{1}{2}}} \xrightarrow[n \to +\infty]{} 0$$

et de même $\lim_{n\to\infty} \dfrac{\sqrt{4n-3}}{2} - \sqrt{n} = 0$. On en déduit que $x_n \underset{n\to\infty}{\sim} \sqrt{n}$. Plus précisément, on a le développement asymptotique

$$\boxed{x_n = \sqrt{n} + \frac{1}{2} + o(1).} \quad \lhd$$

L'énoncé suivant provient des Olympiades de 1985.

2.30. Un exercice d'Olympiades

On définit la suite $(x_n)_{n\geqslant 1}$ par la donnée de x_1 et la relation $x_{n+1} = x_n\left(x_n + \dfrac{1}{n}\right)$ pour tout $n \geqslant 1$. Démontrer qu'il existe un unique x_1 tel que, pour tout $n \in \mathbb{N}^*$, on ait $0 < x_n < x_{n+1} < 1$.

(École polytechnique)

▷ **Solution.**

Soit S l'ensemble des suites définies par $x_1 > 0$ et la relation de récurrence considérée. Toutes les suites de S sont à termes strictement positifs. Pour commencer, étudions les différents comportements possibles d'une suite (x_n) de S.

S'il existe $p \geqslant 1$ tel que $x_p \leqslant x_{p-1}$, alors on a

$$x_{p+1} = x_p\left(x_p + \frac{1}{p}\right) < x_{p-1}\left(x_{p-1} + \frac{1}{p-1}\right) = x_p$$

et par une récurrence immédiate, la suite (x_n) décroît strictement à partir du rang p. La suite (x_n) est donc soit strictement décroissante à partir d'un certain rang, soit strictement croissante. Notons que s'il existe $p \geqslant 1$ tel que $x_p \geqslant 1$, alors $x_{p+1} > x_p \geqslant 1$ et la suite est croissante.

Si la suite (x_n) décroît à partir d'un certain rang, elle converge. Sa limite ℓ vérifie $\ell = \ell^2$: c'est 0 ou 1. D'après la remarque qui précède, ça ne peut pas être 1. Elle converge vers 0. Si la suite (x_n) est croissante, elle converge vers 1 (si tous ses termes sont dans $]0,1[$) ou diverge vers $+\infty$.

Soit I l'ensemble des réels a tels que la suite (x_n) de S définie par $x_1 = a$ soit croissante. L'ensemble I n'est pas vide, car il contient 1. Son complémentaire dans $\mathbb{R}_+^*$ n'est pas vide non plus ; en effet, il contient 0,2 car si $x_1 = 0{,}2$ alors $x_2 = 0{,}24$ et $x_3 = 0{,}1776 < x_2$ et la suite décroît à partir du rang 2. Soit u et v deux éléments de S tels que $u_1 = a$ et $v_1 = b$, avec $a < b$. Par une récurrence immédiate, on a, pour tout $n \in \mathbb{N}^*$, $u_n < v_n$, puis $\dfrac{u_{n+1}}{u_n} < \dfrac{v_{n+1}}{v_n}$. On en déduit que si $a \in$ I,

alors $b \in I$. Considérons $\alpha = \inf I$. Il résulte de ce qui précède que $\alpha > 0$ et que I contient $]\alpha, +\infty[$. On a donc $I =]\alpha, +\infty[$ ou $[\alpha, +\infty[$.

Soit a et b deux éléments de I (avec $a < b$), u et v les suites correspondantes de S. Montrons que u et v ne peuvent pas tendre toutes les deux vers 1. On a en effet, pour tout $n \in \mathbb{N}^*$, $u_n < v_n$ et $v_{n+1} - u_{n+1} = (v_n - u_n)(v_n + u_n + \frac{1}{n})$. Si les suites (u_n) et (v_n) convergeaient toutes deux vers 1, on aurait $\lim\limits_{n\to+\infty} \frac{v_{n+1} - u_{n+1}}{v_n - u_n} = 2 > 1$, ce qui implique que $\lim\limits_{n\to+\infty} v_n - u_n = +\infty$ et une contradiction.

La seule valeur de x_1 qui puisse avoir la propriété voulue est α. Reste à montrer qu'elle convient effectivement, c'est-à-dire que la suite de S de premier terme α est croissante et majorée par 1. Pour cela, on va utiliser la définition de α.

Partons de la remarque suivante. L'application $f_p : x_1 \longmapsto x_p$ est continue sur $\mathbb{R}_+^*$ pour tout $p \in \mathbb{N}^*$ (on le montre facilement par récurrence).

Soit (x_n) la suite de S définie par $x_1 = \alpha$. Supposons que (x_n) n'est pas croissante. Soit $p \in \mathbb{N}^*$ tel que $x_{p+1} < x_p$ et donc $x_p + \frac{1}{p} < 1$. Par continuité de f_p, il existe $y_1 > x_1$ tel que la suite (y_n) de S de premier terme y_1 vérifie $y_p + \frac{1}{p} < 1$. On a donc $y_{p+1} < y_p$; la suite (y_n) décroît à partir du rang p et $y_1 \notin I$. C'est faux car I contient $]\alpha, +\infty[$. On en déduit que $\alpha \in I$.

Supposons que (x_n) n'est pas majorée par 1. Il existe alors $p \in \mathbb{N}^*$ tel que $x_p > 1$. Toujours par continuité de f_p, il existe $y_1 < x_1$ tel que la suite (y_n) de S de premier terme y_1 vérifie $y_p > 1$. L'existence de p tel que $y_p > 1$ implique que (y_n) est croissante et donc $y_1 \in I$, d'où une nouvelle contradiction. On conclut que la suite (x_n) est croissante et majorée par 1.

Conclusion. Si $x_1 = \alpha$, la suite est croissante et majorée par 1 ; elle converge vers 1. Si $x_1 \in]0, \alpha[$, la suite est décroissante à partir d'un certain rang; elle converge vers 0. Si $x_1 > \alpha$, la suite est croissante et diverge vers $+\infty$. ◁

2.31. Équivalent d'une suite récurrente

Soit $(u_n)_{n\geqslant 0}$ une suite réelle vérifiant $u_{n+1} = |u_n - n|$ pour tout n. Déterminer un équivalent de u_n.

(École normale supérieure)

▷ **Solution.**

Bien entendu la valeur absolue nous gêne pour expliciter la suite et il

est naturel de s'intéresser au signe de $u_n - n$. Il n'est pas possible que cette quantité soit toujours positive. En effet, on aurait alors $u_{n+1} = u_n - n$ pour tout n et la suite serait décroissante. Cela contredit clairement le fait que $u_n \geqslant n$ pour tout n, qui implique que (u_n) diverge vers $+\infty$. Il existe donc toujours au moins un indice $\mathrm{N} \geqslant 1$ tel que $u_\mathrm{N} \leqslant \mathrm{N}$. Mais on a alors $u_{\mathrm{N}+1} = \mathrm{N} - u_\mathrm{N} \leqslant \mathrm{N} < \mathrm{N}+1$ et cela se propage clairement par récurrence : on a donc $u_n \leqslant n$ pour tout $n \geqslant \mathrm{N}$. La relation de récurrence se transforme donc à partir du rang N en $u_{n+1} = -u_n + n$. Cela permet d'expliciter la suite : si on pose $v_n = (-1)^n u_n$ on a $v_{n+1} - v_n = (-1)^{n+1} n$ pour tout $n \geqslant \mathrm{N}$. En sommant cette égalité entre le rang N et un rang n plus grand, il vient

$$\forall n \geqslant \mathrm{N}, \quad v_n = v_\mathrm{N} + \sum_{k=\mathrm{N}}^{n-1} (-1)^{k+1} k,$$

autrement dit

$$u_n = (-1)^n v_n = (-1)^n v_\mathrm{N} + \sum_{k=\mathrm{N}}^{n-1} (-1)^{n+k+1} k = \mathrm{C}_n + \sum_{k=1}^{n-1} (-1)^{n-1-k} k$$

où (C_n) est une suite bornée. La somme alternée

$$\sum_{k=1}^{n-1} (-1)^{n-1-k} k = (n-1) - (n-2) + (n-3) - (n-4) + \cdots \pm 1$$

est égale à $\dfrac{n-1}{2}$ si n est impair et à $\dfrac{n}{2}$ si n est pair (regrouper les termes deux par deux). On en déduit donc que $\boxed{u_n \sim \dfrac{n}{2}}$. ◁

Les deux exercices qui suivent concernent des suites récurrentes de la forme $u_{n+1} = f_n(u_n)$ que l'on peut considérer comme des perturbations de suites récurrentes. Le premier concerne une suite arithmético-géométrique $u_{n+1} = au_n + b$ (a non nul). Rappelons qu'une telle suite se ramène par une simple translation à l'étude d'une suite géométrique de raison a. En particulier, une telle suite converge (vers le point fixe de $x \mapsto ax+b$) lorsque $|a| < 1$. On peut se demander ce qu'il en est d'une suite vérifiant une relation de la forme $u_{n+1} = au_n + b + \varepsilon_n$, où (ε_n) tend vers 0. C'est ce que propose l'exercice suivant dans le cas particulier $a = -\dfrac{1}{2}$: on y montre que la suite converge toujours vers le point fixe.

2.32. Suite arithmético-géométrique perturbée

Soit $(x_n)_{n\geqslant 0}$ une suite de nombre réels. On pose, pour $n \geqslant 1$, $y_n = x_{n-1} + 2x_n$. Montrer que (x_n) converge si et seulement si (y_n) converge.

(École polytechnique)

Solution.

• Il est clair que si (x_n) converge vers ℓ, alors (y_n) converge vers 3ℓ. Pour la réciproque, on observe que l'application qui à (x_n) associe (y_n) est linéaire et qu'il suffit donc d'établir le résultat lorsque (y_n) tend vers 0 (dans le cas général où (y_n) converge vers ℓ, on regarde $x'_n = x_n - \dfrac{\ell}{3}$; la suite $y'_n = x'_{n-1} + 2x'_n$ converge vers 0 donc x'_n converge aussi vers 0 et x_n tend vers $\dfrac{\ell}{3}$).

• On suppose donc que (y_n) tend vers 0 et on se propose de montrer que (x_n) tend vers 0. Comme $x_n = -\dfrac{1}{2}x_{n-1} + \dfrac{1}{2}y_n$ on peut voir la suite (x_n) comme une suite géométrique perturbée. On commence par exprimer x_n en fonction des y_k. On a

$$\begin{aligned} y_n &= x_{n-1} + 2x_n, \\ y_{n-1} &= x_{n-2} + 2x_{n-1}, \\ &\vdots \\ y_1 &= x_0 + 2x_1. \end{aligned}$$

On multiplie la première ligne par $\dfrac{1}{2}$, la seconde par $-\dfrac{1}{4}$, la troisième par $\dfrac{1}{8}$, ... et on additionne le tout. Il vient

$$\sum_{k=1}^{n} (-1)^{n-k} \frac{y_k}{2^{n+1-k}} = x_n + \frac{(-1)^{n+1}}{2^n} x_0.$$

Il suffit de prouver que le premier membre tend vers 0. On a

$$\left| \sum_{k=1}^{n} (-1)^{n-k} \frac{y_k}{2^{n+1-k}} \right| \leqslant \frac{1}{2^{n+1}} \sum_{k=1}^{n} 2^k |y_k|.$$

Soit $\varepsilon > 0$ et n_0 tel que $|y_n| \leqslant \varepsilon$ pour $n \geqslant n_0$. On obtient, pour $n \geqslant n_0$,

$$\frac{1}{2^{n+1}} \sum_{k=1}^{n} 2^k |y_k| \leqslant \frac{1}{2^{n+1}} \sum_{k=1}^{n_0-1} 2^k |y_k| + \frac{\varepsilon}{2^{n+1}} \sum_{k=n_0}^{n} 2^k \leqslant \frac{1}{2^{n+1}} \sum_{k=1}^{n_0-1} 2^k |y_k| + \varepsilon$$

et ceci est $\leqslant 2\varepsilon$ pour n assez grand. D'où le résultat. $\lhd$

Pour prouver que $\dfrac{1}{2^{n+1}}\sum_{k=1}^{n} 2^k|y_k|$ *tend vers* 0, *on peut aussi invoquer le théorème de sommation des relations de négligeabilité. Puisque la suite* (y_n) *converge vers* 0, *on a* $2^n y_n \underset{n\to+\infty}{=} o(2^n)$. *Comme* $\sum 2^n$ *diverge, on en déduit que* $\sum_{k=1}^{n} y_k 2^k = o\left(\sum_{k=1}^{n} 2^k\right)$, *ce qui donne le résultat voulu car* $\sum_{k=1}^{n} 2^k = 2^{n+1} - 1 \sim 2^{n+1}$.

Voici un autre énoncé posé à l'oral de l'école polytechnique, qui se ramène au résultat de l'exercice : étudier la convergence d'une suite définie par $u_1 > 0$ *et* $u_{n+1} = \dfrac{n+1}{2n}\sqrt{u_n}$. *Si on pose* $v_n = \ln u_n$ *on a aisément que* $v_{n+1} = \dfrac{1}{2}v_n - \ln 2 + \varepsilon_n$ *où* ε_n *tend vers* 0 . *On montre alors que* (v_n) *converge vers* $-2\ln 2$ *et donc que* (u_n) *converge vers* $\dfrac{1}{4}$.

L'énoncé suivant donne un dernier exemple un peu plus compliqué.

2.33. Système dynamique perturbé

Soit $(u_n)_{n\in\mathbb{N}}$ une suite bornée de réels telle que $u_{n+1} - u_n - u_n^2$ tende vers 0. Montrer que $(u_n)_{n\in\mathbb{N}}$ converge vers 0.

(École polytechnique)

$\rhd$ **Solution.**

Notons que si la suite (u_n) converge, sa limite est nécessairement 0. Posons $v_n = u_{n+1} - u_n - u_n^2$. Soit $\varepsilon > 0$ et $n_0 \in \mathbb{N}$ tel que $|v_n| \leqslant \varepsilon^2$ pour tout $n \geqslant n_0$.

On a $u_{n+1} - u_n = u_n^2 + v_n \geqslant u_n^2 - \varepsilon^2$ pour tout $n \geqslant n_0$. S'il existe $n \geqslant n_0$ tel que $u_n \geqslant \varepsilon$, on a $u_{n+1} \geqslant u_n \geqslant \varepsilon$ et par une récurrence immédiate, la suite est croissante à partir du rang n et minorée par ε. Comme elle est bornée, elle converge vers $\ell \geqslant \varepsilon$. C'est impossible car elle ne peut converger que vers 0. On a donc $u_n < \varepsilon$ pour tout $n \geqslant n_0$.

On va montrer que $u_n > -\varepsilon$ à partir d'un certain rang. Si, pour tout $n \geqslant n_0$, on a $u_n \leqslant -\varepsilon$, on a, pour tout $n \geqslant n_0$,

$$u_{n+1} - u_n = u_n^2 + v_n \geqslant u_n^2 - \varepsilon^2 \geqslant 0.$$

La suite $(u_n)_{n\geqslant n_0}$ est croissante et majorée par $-\varepsilon$. Elle converge donc vers $\ell \leqslant -\varepsilon$. C'est impossible. Ainsi il existe $n_1 \geqslant n_0$ tel que $u_{n_1} > -\varepsilon$.

Si $n \geqslant n_0$ et $u_n \geqslant -\varepsilon$, on a alors $u_{n+1} \geqslant u_n + u_n^2 - \varepsilon^2$. La fonction $f : x \longmapsto x + x^2$ est croissante sur $\left[-\dfrac{1}{2}, +\infty\right[$, donc si $\varepsilon < \dfrac{1}{2}$, $u_n \geqslant -\varepsilon$

implique $f(u_n) \geqslant f(-\varepsilon)$ et donc $u_{n+1} \geqslant f(-\varepsilon) - \varepsilon^2 \geqslant -\varepsilon$. Pour un tel choix de ε, on a par une récurrence immédiate $u_n \geqslant -\varepsilon$ pour tout $n \geqslant n_1$.

Pour tout $\varepsilon \in \left]0, \frac{1}{2}\right[$, il existe n_1 tel que $|u_n| < \varepsilon$ pour $n \geqslant n_1$, donc (u_n) converge vers 0. ◁

Viennent maintenant des suites vérifiant des relations de récurrence linéaires.

2.34. Suite récurrente linéaire d'ordre 2

On définit une suite de polynômes : $P_0 = 1$, $P_1 = X$ et pour $n \geqslant 1$, $(q+1)XP_n = qP_{n+1} + P_{n-1}$ où q est un réel non nul fixé. Déterminer l'ensemble des réels z tels que la suite $(P_n(z))_{n\in\mathbb{N}}$ soit bornée.

(École polytechnique)

Solution.

Puisque q n'est pas nul, la relation donnée définit bien une suite de polynômes (P_n). Pour tout réel z, la suite $(P_n(z))$ vérifie une relation de récurrence linéaire d'ordre deux à coefficients constants

$$P_{n+1}(z) = \frac{(q+1)z}{q} P_n(z) - \frac{1}{q} P_{n-1}(z)$$

dont l'équation caractéristique est $q\lambda^2 - (q+1)z\lambda + 1 = 0$.

Examinons d'abord le cas où le discriminant $\Delta = (q+1)^2 z^2 - 4q$ n'est pas nul. L'équation possède deux solutions distinctes, réelles ou complexes conjuguées, λ_1 et λ_2 et il existe des constantes a et b telles que, pour tout $n \in \mathbb{N}$, $P_n(z) = a\lambda_1^n + b\lambda_2^n$. Les conditions $P_0(z) = 1$ et $P_1(z) = z$ conduisent à $a = \dfrac{\lambda_2 - z}{\lambda_2 - \lambda_1}$ et $b = \dfrac{\lambda_1 - z}{\lambda_1 - \lambda_2}$ et donc

$$P_n(z) = \frac{\lambda_2 - z}{\lambda_2 - \lambda_1} \lambda_1^n + \frac{\lambda_1 - z}{\lambda_1 - \lambda_2} \lambda_2^n.$$

Notons que $z = \lambda_1$ ou λ_2, c'est-à-dire que z est solution de l'équation caractéristique, si $z = \pm 1$. Si on a par exemple $z = \lambda_1$, alors $P_n(z) = \lambda_1^n = z^n = \pm 1$ et la suite (P_n) est bornée. Les valeurs $z = -1$ et $z = 1$ conviennent. Désormais, on écarte ces cas et on suppose donc a et b non nuls.

La suite $(P_n(z))$ est alors bornée si et seulement si $|\lambda_1| \leqslant 1$ et $|\lambda_2| \leqslant 1$. La condition est clairement suffisante. Elle est aussi nécessaire si $|\lambda_1|$ et $|\lambda_2|$ sont distincts, car alors on a $|P_n(z)| \underset{n\to\infty}{\sim} |a||\lambda_1|^n$ si $|\lambda_1| > |\lambda_2|$.

Si enfin, on a $|\lambda_1| = |\lambda_2| > 1$ et que la suite $(P_n(z))$ est bornée alors, en posant $\dfrac{\lambda_1}{\lambda_2} = e^{i\theta}$, on obtient $\lim\limits_{n\to\infty} a + be^{in\theta} = \lim\limits_{n\to\infty} \dfrac{P_n(z)}{\lambda_1^n} = 0$. En écrivant

$$a(1 - e^{i\theta}) = a + be^{i(n+1)\theta} - e^{i\theta}(a + be^{in\theta})$$

et en faisant tendre n vers l'infini, on obtient $a(1 - e^{i\theta}) = 0$, ce qui est contradictoire avec les hypothèses puisque $a \neq 1$ et $\lambda_1 \neq \lambda_2$.

Distinguons maintenant selon le signe de Δ.

• La condition $\Delta < 0$ nécessite $q > 0$ et s'écrit $|z| < \dfrac{2\sqrt{q}}{q+1}$. Dans ce cas, les racines λ_1 et λ_2 sont complexes conjuguées et on a donc $|\lambda_1| = |\lambda_2| = \sqrt{\lambda_1\lambda_2} = \dfrac{1}{\sqrt{q}}$. La suite $(P_n(z))$ est bornée si $q \geqslant 1$.

• Si $\Delta > 0$, c'est-à-dire si $(q+1)^2z^2 > 4q$, l'équation caractéristique a deux solutions réelles. La suite est bornée si elles sont comprises entre -1 et 1. En notant f le trinôme $qX^2 - (q+1)zX + 1$, il faut $qf(1) \geqslant 0$ et $qf(-1) \geqslant 0$ (pour que -1 et 1 soient à l'extérieur des racines) et $-1 \leqslant \dfrac{\lambda_1 + \lambda_2}{2} = \dfrac{(q+1)z}{2q} \leqslant 1$ (pour que -1 et 1 soient de part et d'autre des racines). Les deux premières conditions s'écrivent $q(q+1)(1-z) \geqslant 0$ et $q(q+1)(1+z) \geqslant 0$, soit $-q(q+1) \leqslant q(q+1)z \leqslant q(q+1)$, ce qui nécessite $q(q+1) \geqslant 0$. Si $q = -1$, toutes les conditions sont réalisées pour z quelconque. Nous écartons ce cas désormais. Les conditions s'écrivent alors $-1 \leqslant z \leqslant 1$ et $-\dfrac{2q}{q+1} \leqslant z \leqslant \dfrac{2q}{q+1}$.

Deux cas sont à distinguer.

⋆ Si $q < -1$, la condition $\Delta > 0$ est réalisée. Par ailleurs, on a $\dfrac{2q}{q+1} > 1$, donc on trouve une seule condition $|z| \leqslant 1$.

⋆ Si $q > 0$, la condition $\Delta > 0$ s'écrit $|z| > \dfrac{2\sqrt{q}}{q+1}$. Compte tenu de la condition $|z| \leqslant \dfrac{2q}{q+1}$, il faut $\sqrt{q} < q$ et donc $q > 1$. On a alors de nouveau $\dfrac{2q}{q+1} \geqslant 1$, donc la condition sur z est $\dfrac{2\sqrt{q}}{q+1} < |z| \leqslant 1$.

Reste à traiter le cas où $\Delta = 0$, soit $(q+1)^2z^2 = 4q$ (ce qui implique $q \geqslant 0$). L'équation caractéristique a une seule solution $\lambda = \dfrac{(q+1)z}{2q}$. Il existe des constantes a et b telles que $P_n(z) = (a + bn)\lambda^n$. Les conditions initiales donnent $b = 1$, $a = z - 1$. Le cas $z = 1$ étant écarté, il faut $|\lambda| = \dfrac{1}{\sqrt{q}} < 1$, soit $q > 1$ et $|z| < \dfrac{2\sqrt{q}}{q+1}$. Il ne reste plus qu'à rassembler les différents cas pour conclure.

Conclusion. Si $q \in]-\infty, -1[\cup [1, +\infty[$, la suite $(P_n(z))$ est bornée si $|z| \leqslant 1$. Si $q = -1$, on peut prendre z quelconque. Si enfin $q \in]-1, 1[$, la suite $(P_n(z))$ n'est bornée que pour $z = -1$ et $z = 1$. ◁

Bien que l'énoncé suivant concerne a priori une équation fonctionnelle, les suites récurrentes linéaires d'ordre 2 sont au cœur de la solution.

2.35. Une équation fonctionnelle

Déterminer les applications $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$ vérifiant

$$\forall x > 0,\ f(f(x)) = 6x - f(x).$$

(**École normale supérieure**)

Solution.

Analyse. Soit f une solution. Fixons $x > 0$ et posons $u_n = f^n(x)$ où $f^n = f \circ f \circ \cdots \circ f$ est la composée de f n fois avec elle-même. Par hypothèse, on a la relation de récurrence linéaire $u_{n+2} = 6u_n - u_{n+1}$. L'équation caractéristique associée est $X^2 + X - 6 = 0$ et ses racines sont 2 et -3. Il existe donc deux constantes A, B telles que $u_n = A2^n + B(-3)^n$ pour tout n. On trouve A et B avec les conditions initiales $u_0 = x$ et $u_1 = f(x)$. Il vient alors,

$$\forall n \in \mathbb{N},\quad f^n(x) = u_n = \frac{3x + f(x)}{5} 2^n + \frac{2x - f(x)}{5} (-3)^n.$$

Or $f^n(x)$ doit être strictement positif pour tout n. Cela impose $f(x) = 2x$ (regarder u_{2n} et u_{2n+1} et faire tendre n vers l'infini).

Synthèse. La fonction $x \mapsto 2x$ est bien solution du problème posé. ◁

Voici encore une récurrence linéaire d'ordre 2. Les coefficients ne sont plus constants, mais les méthodes employées sont similaires. On utilise le fait que l'ensemble des solutions est encore un espace vectoriel de dimension 2.

2.36. Suites vérifiant $u_{n+2} = (n+1)u_{n+1} + u_n$

Soit E l'ensemble des suites réelles (u_n) telles que, pour tout entier $n \in \mathbb{N}$,

$$u_{n+2} = (n+1)u_{n+1} + u_n.$$

1. Montrer que E est un espace vectoriel de dimension 2.

2. On considère les deux suites (a_n) et (b_n) de E définies par $a_0 = 1$, $a_1 = 0$, $b_0 = 0$, $b_1 = 1$. Montrer que les suites (a_n) et (b_n) tendent vers $+\infty$.

3. On pose $w_n = a_{n+1}b_n - a_n b_{n+1}$. Expliciter w_n.

4. Soit $c_n = \dfrac{a_n}{b_n}$, pour $n \geqslant 1$. Montrer que la suite (c_n) converge.

5. Montrer qu'il existe un unique réel r tel que $(a_n + rb_n)$ converge vers 0.

(École polytechnique)

▷ **Solution.**

1. Il est clair que E est un sous-espace vectoriel de l'espace vectoriel des suites réelles. L'application $\varphi : (u_n) \in \mathrm{E} \longmapsto (u_0, u_1) \in \mathbb{R}^2$ est linéaire et clairement bijective car un élément de E est déterminé de manière unique par la donnée de u_0 et u_1, réels quelconques. L'espace vectoriel E est donc isomorphe à $\mathbb{R}^2$.

2. On a $a_n > 0$ et $b_n > 0$ pour $n \geqslant 2$, comme le montre une récurrence immédiate. On en déduit que les suites (a_n) et (b_n) sont strictement croissantes à partir du rang 2. Comme elles sont à valeurs dans $\mathbb{N}$, elles tendent vers $+\infty$.

3. On a, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} w_{n+1} &= a_{n+2}b_{n+1} - a_{n+1}b_{n+2} \\ &= ((n+1)a_{n+1} + a_n)b_{n+1} - a_{n+1}((n+1)b_{n+1} + b_n) = -w_n. \end{aligned}$$

Puisque $w_0 = -1$, on a $w_n = (-1)^{n+1}$, pour tout $n \in \mathbb{N}$.

4. On a, pour tout $n \in \mathbb{N}$,

$$c_{n+1} - c_n = \frac{a_{n+1}}{b_{n+1}} - \frac{a_n}{b_n} = \frac{w_n}{b_n b_{n+1}} = \frac{(-1)^{n+1}}{b_n b_{n+1}}.$$

La série de terme général $c_{n+1} - c_n$ est alternée et vérifie le critère spécial de convergence des séries alternées, car la suite (b_n) est à termes strictement positifs à partir du rang 1 et croît vers $+\infty$. De la convergence de $\sum (c_{n+1} - c_n)$, on déduit la convergence de la suite (c_n). On note ℓ la limite de cette suite.

5. Si la suite $(a_n + rb_n)$ converge vers 0, alors *a fortiori* $\left(\dfrac{a_n}{b_n} + r\right)_{n\in\mathbb{N}}$ converge vers 0, puisque (b_n) tend vers $+\infty$ et donc $r = -\ell$. Montrons, réciproquement, que cette valeur convient. On a, pour tout $n \in \mathbb{N}^*$,

$$c_n - \ell = \lim_{p\to+\infty}(c_n - c_p) = \lim_{p\to+\infty} \sum_{k=n}^{p-1} (c_k - c_{k+1}) = \sum_{k=n}^{+\infty} (c_k - c_{k+1}).$$

Puisque cette série vérifie le critère spécial de convergence des séries alternées, cette somme est majorée en valeur absolue par la valeur absolue du premier terme. On a donc, pour $n \geqslant 1$,

$$|c_n - \ell| \leqslant |c_n - c_{n+1}| \leqslant \frac{1}{b_n b_{n+1}}, \text{ soit } |a_n - \ell b_n| \leqslant \frac{1}{b_{n+1}}.$$

On en déduit que $(a_n - \ell b_n)$ converge vers 0. ◁

Les exercices ci-après concernent les suites récurrentes linéaires d'ordre quelconque.

2.37. Suites à récurrence linéaire (1)

1. Soit E l'ensemble des suites $u \in \mathbb{C}^{\mathbb{N}}$ telles qu'il existe $d \in \mathbb{N}^*$ et $(a_i)_{0 \leqslant i < d} \in \mathbb{C}^d$ tels que pour tout $n \in \mathbb{N}$, $u_{n+d} = \sum_{i=0}^{d-1} a_i u_{n+i}$. Montrer que E est stable par addition et multiplication.

2. Soit $\mathcal{E}$ le sous-ensemble de $\mathbb{C}^{\mathbb{N}}$ formé des suites périodiques. Montrer que $\mathcal{E}$ est un sous-espace vectoriel de $\mathbb{C}^{\mathbb{N}}$ et en donner une base.

3. Soit P un polynôme de $\mathbb{C}[X]$. Peut-on expliciter une récurrence linéaire vérifiée par la suite $(P(n))_{n \geqslant 0}$?

(École polytechnique)

Solution.

1. Lorsque $d = 2$, nous savons quelle est l'expression de u_n en fonction de n et des racines de $X^2 - a_1 X - a_0$. Le lemme suivant étend le résultat à d quelconque.

Lemme. *Soit $(a_i)_{0 \leqslant i < d} \in \mathbb{C}^d$. On note*

$$P = X^d - a_{d-1}X^{d-1} - \cdots - a_1 X - a_0 = \prod_{i=1}^{r} (X - \lambda_i)^{k_i},$$

où $\lambda_1, \ldots, \lambda_r \in \mathbb{C}$ sont deux à deux distincts et les k_i dans $\mathbb{N}^*$.

On suppose $a_0 \neq 0$. Alors, l'ensemble des suites vérifiant pour tout $n \in \mathbb{N}$, $u_{n+d} = \sum_{i=0}^{d-1} a_i u_{n+i}$ est constitué des suites s'écrivant

$$\forall n \in \mathbb{N}, \quad u_n = Q_1(n)\lambda_1^n + \cdots + Q_r(n)\lambda_r^n$$

avec pour tout i, $Q_i \in \mathbb{C}[X]$ et $\deg Q_i < k_i$.

Si $a_0 = 0$ et $\lambda_r = 0$, l'ensemble est constitué des suites s'écrivant

$$\forall n \geqslant k_r, \quad u_n = Q_1(n)\lambda_1^n + \cdots + Q_{r-1}(n)\lambda_{r-1}^n$$

avec pour tout i, $Q_i \in \mathbb{C}[X]$ et $\deg Q_i < k_i$, les $k_r - 1$ premiers termes étant quelconques.

Démonstration.

Notons T l'opérateur de translation sur $\mathbb{C}^{\mathbb{N}}$: si $u = (u_n)_{n\in\mathbb{N}}$ est une suite de $\mathbb{C}$, $\mathrm{T}(u) = (u_{n+1})_{n\geqslant 0}$. Par définition de P, $\mathrm{P}(\mathrm{T})(u)$ est la suite de terme général $u_{n+d} - \sum_{i=0}^{d-1} a_i u_{n+i}$, si bien que $\operatorname{Ker}\mathrm{P}(\mathrm{T})$ n'est autre que l'ensemble des suites $(u_n)_{n\in\mathbb{N}}$ vérifiant la relation de récurrence linéaire : pour tout $n \in \mathbb{N}$, $u_{n+d} = \sum_{i=0}^{d-1} a_i u_{n+i}$. Le lemme de décomposition des noyaux nous donne que $\operatorname{Ker}\mathrm{P}(\mathrm{T})$ est la somme directe des $\operatorname{Ker}(\mathrm{T} - \lambda_i \operatorname{Id})^{k_i}$. On est donc amené à préciser $\operatorname{Ker}(\mathrm{T} - \lambda \operatorname{Id})^k$ pour $k \in \mathbb{N}^*$ et $\lambda \in \mathbb{C}$.

Traitons le cas λ non nul. Montrons que les éléments de $\operatorname{Ker}(\mathrm{T}-\lambda \operatorname{Id})^k$ sont les suites $u = (u_n)_{n\in\mathbb{N}}$ telles que pour $n \in \mathbb{N}$, $u_n = \mathrm{Q}(n)\lambda^n$ avec $\mathrm{Q} \in \mathbb{C}[\mathrm{X}]$ de degré strictement inférieur à k. On remarque que si u_n est de cette forme, $u_{n+1} - \lambda u_n = (\mathrm{Q}(n+1) - \mathrm{Q}(n))\lambda^{n+1} = \mathrm{D}(\mathrm{Q})(n)\lambda^{n+1}$ où D est l'endomorphisme de $\mathbb{C}[\mathrm{X}]$ défini par $\mathrm{A} \longmapsto \mathrm{A}(\mathrm{X}+1) - \mathrm{A}(\mathrm{X})$. Par itération, $(\mathrm{T} - \lambda \operatorname{Id})^k(u)$ est la suite de terme général $\mathrm{D}^k(\mathrm{Q})(n)\lambda^{n+k}$, ce qui vaut 0 si $k > \deg \mathrm{Q}$ car D diminue le degré d'une unité. Réciproquement, montrons par récurrence sur $k \geqslant 1$ qu'une suite $u = (u_n)_{n\in\mathbb{N}}$ de $\operatorname{Ker}(\mathrm{T} - \lambda \operatorname{Id})^k$ est de la forme $(\mathrm{Q}(n)\lambda^n)_{n\in\mathbb{N}}$ avec $\mathrm{Q} \in \mathbb{C}[\mathrm{X}]$ et $\deg \mathrm{Q} < k$. Pour $k = 1$, la suite vérifie pour tout $n \in \mathbb{N}$, $u_{n+1} - \lambda u_n = 0$ ce qui donne $u_n = \lambda^n u_0$: c'est bien la forme attendue. Supposons $k \geqslant 2$. Soit $u = (u_n)_{n\in\mathbb{N}} \in \operatorname{Ker}(\mathrm{T} - \lambda \operatorname{Id})^k$. Alors $\mathrm{T}(u) - \lambda u$ est dans $\operatorname{Ker}(\mathrm{T} - \lambda \operatorname{Id})^{k-1}$ et elle s'écrit par hypothèse de récurrence $(\mathrm{R}(n)\lambda^n)_{n\in\mathbb{N}}$ avec $\mathrm{R} \in \mathbb{C}[\mathrm{X}]$ et $\deg \mathrm{R} < k-1$. Comme D diminue le degré d'une unité, on a classiquement l'égalité $\mathrm{D}(\mathbb{C}_{k-1}[\mathrm{X}]) = \mathbb{C}_{k-2}[\mathrm{X}]$, si bien qu'il existe $\mathrm{Q} \in \mathbb{C}[\mathrm{X}]$ avec $\deg \mathrm{Q} < k$ tel que $\mathrm{Q}(\mathrm{X}+1) - \mathrm{Q}(\mathrm{X}) = \dfrac{\mathrm{R}(\mathrm{X})}{\lambda}$. Ainsi, pour tout $n \in \mathbb{N}$, $u_{n+1} - \lambda u_n = (\mathrm{Q}(n+1) - \mathrm{Q}(n))\lambda^{n+1}$. La suite $v_n = \dfrac{u_n}{\lambda^n} - \mathrm{Q}(n)$ vérifie $v_{n+1} = v_n$, autrement dit, elle est constante. Ainsi, $u_n = (\mathrm{Q}(n) + v_0)\lambda^n$ et la suite est bien de la forme voulue.

Le cas $\lambda = 0$ est à part : le noyau $\operatorname{Ker}\mathrm{T}^k$ est l'ensemble des suites nulles à partir du rang k.

Comme $\operatorname{Ker}\mathrm{P}(\mathrm{T})$ est somme directe des sous-espaces propres, on obtient bien la forme annoncée par le lemme. ◇

Reprenons les notations du lemme et posons pour $\lambda \in \mathbb{C}^*$ et $\ell \in \mathbb{N}$, $e_{\lambda,\ell} = (n^\ell \lambda^n)_{n\in\mathbb{N}}$. À $\lambda \neq 0$ fixé, la famille des suites $e_{\lambda,\ell}$ pour $\ell \in \mathbb{N}$ est libre car si $\alpha_0, \ldots, \alpha_p \in \mathbb{C}$ avec $\alpha_p \neq 0$, la limite de

$$\frac{\alpha_0\lambda^n + \alpha_1 n\lambda^n + \cdots + \alpha_p n^p \lambda^n}{n^p \lambda^n}$$

est α_p et la combinaison linéaire $\alpha_0\lambda^n + \alpha_1 n\lambda^n + \cdots + \alpha_p n^p\lambda^n$ ne peut être nulle pour tout $n \in \mathbb{N}$. On en déduit que les suites $e_{\lambda,0} = (\lambda^n)$,

$e_{\lambda,1} = (n\lambda^n), \ldots, e_{\lambda,k-1} = (n^{k-1}\lambda^n)$ forment une base de $\mathrm{Ker}(\mathrm{T} - \lambda\,\mathrm{Id})^k$. En notant $e_{0,\ell} = (\delta_{n,\ell})_{n\in\mathbb{N}}$, les k suites $(e_{0,0}, e_{0,1}, \ldots, e_{0,k-1})$ constituent une base de $\mathrm{Ker}\,\mathrm{T}^k$. Par recollement des bases, puisque la somme des sous-espaces propres est directe et fait $\mathrm{Ker}\,\mathrm{P}(\mathrm{T})$, on est en mesure de donner une base de $\mathrm{Ker}\,\mathrm{P}(\mathrm{T})$ à savoir les $e_{\lambda_i,\ell}$ avec $1 \leqslant i \leqslant r$ et $0 \leqslant \ell \leqslant k_i - 1$. On remarque que quel que soit $\lambda \in \mathbb{C}$, $\dim \mathrm{Ker}(\mathrm{T} - \lambda\,\mathrm{Id})^k = k$ et donc la dimension de $\mathrm{Ker}\,\mathrm{P}(\mathrm{T})$ est le degré de P.

L'ensemble E apparaît alors comme la réunion des noyaux de $\mathrm{P}(\mathrm{T})$ avec $\mathrm{P} \in \mathbb{C}[\mathrm{X}]$ non nul. On en déduit que E est contenu dans le sous-espace E′ engendré par les suites $e_{\lambda,\ell}$ avec $\ell \in \mathbb{N}$ et $\lambda \in \mathbb{C}$. Réciproquement, une suite de E′ s'écrit comme combinaison linéaire d'éléments de la forme $e_{\mu_1,\ell_1}, \ldots, e_{\mu_s,\ell_s}$ avec les $\mu_i \in \mathbb{C}$ (pas forcément deux à deux distincts) et $\ell_i \in \mathbb{N}$. D'après le lemme, elle est dans $\mathrm{Ker}\,\mathrm{Q}(\mathrm{T})$ avec $\mathrm{Q} = (\mathrm{X} - \mu_1)^{\ell_1} \cdots (\mathrm{X} - \mu_s)^{\ell_s}$ et elle vérifie donc une relation de récurrence linéaire d'ordre $\deg \mathrm{Q}$: c'est un élément de E, autrement dit E est le sous-espace de $\mathbb{C}^{\mathbb{N}}$ qui admet pour base les suites $e_{\lambda,\ell}$ avec $\lambda \in \mathbb{C}$ et $\ell \in \mathbb{N}$ (cette famille est effectivement libre).

Il est alors clair que E est stable par addition. Pour montrer la stabilité pour la multiplication, par multilinéarité du produit, il suffit de vérifier que le produit de deux éléments de la base des $e_{\lambda,\ell}$ est encore dans E, ce qui est immédiat.

2. La suite nulle est 1-périodique. Si u et u' sont deux suites de $\mathbb{C}^{\mathbb{N}}$ de périodes respectives d et d' et $\lambda \in \mathbb{C}$, λu est périodique de période d et $u+u'$ est périodique de période dd' : l'ensemble $\mathcal{E}$ est bien un sous-espace de $\mathbb{C}^{\mathbb{N}}$. Il apparaît en reprenant les notations de la question précédente que u est périodique de période d si, et seulement si,

$$u \in \mathrm{Ker}(\mathrm{T}^d - \mathrm{Id}) = \bigoplus_{z\in \mathrm{U}_d} \mathrm{Ker}(\mathrm{T} - z\,\mathrm{Id}),$$

où U_d désigne le groupe des racines d-ièmes de l'unité. Ainsi, $\mathcal{E}$ est la réunion pour $d \in \mathbb{N}^*$ des noyaux $\mathrm{Ker}(\mathrm{T}^d - \mathrm{I})$ qui admettent comme base la famille constituée des suites $(z^n)_{n\in\mathbb{N}}$ où $z \in \mathrm{U}_d$. On en déduit que les suites $\varepsilon_z = (z^n)_{n\in\mathbb{N}}$ pour z racine d-ième de l'unité ($d \in \mathbb{N}^*$) engendrent $\mathcal{E}$. Comme $\varepsilon_z \in \mathrm{Ker}(\mathrm{T} - z\,\mathrm{Id})$, la famille des ε_z est libre : elle constitue une base de $\mathcal{E}$.

3. Si P est de degré d, alors P est dans le sous-espace F de $\mathbb{C}^{\mathbb{N}}$ engendré par les suites $(1), (n), \ldots, (n^d)$ qui est d'après le travail fait à la première question le noyau de $(\mathrm{T} - \mathrm{Id})^{d+1} = \sum_{k=0}^{d+1} (-1)^{d+1-k} \mathrm{C}_{d+1}^k \mathrm{T}^k$. On en déduit que pour tout $n \in \mathbb{N}$,

$$\boxed{\mathrm{P}(n+d+1) = \sum_{k=0}^{d} (-1)^{d-k} \mathrm{C}_{d+1}^k \mathrm{P}(n+k)}. \quad \lhd$$

2.38. Suites à récurrence linéaire (2)

Soit $(a_1, \dots, a_p) \in \mathbb{C}^p$ avec $a_1 + \cdots + a_p = 1$. On suppose que la suite $(u_n)_{n\in\mathbb{N}}$ définie par ses p premiers termes $u_0, \dots, u_{p-1}$ et la relation de récurrence $u_{n+p} = a_p u_{n+p-1} + \cdots + a_1 u_n$ converge. Déterminer la limite de $(u_n)_{n\in\mathbb{N}}$.

(École polytechnique)

▷ **Solution.**

Comme dans l'exercice précédent, on note T l'opérateur de translation sur $\mathbb{C}^{\mathbb{N}}$: si $u = (u_n)_{n\in\mathbb{N}}$ est une suite de $\mathbb{C}$, $\mathrm{T}(u) = (u_{n+1})_{n\geqslant 0}$. En posant $\mathrm{P} = \mathrm{X}^p - a_p\mathrm{X}^{p-1} - \cdots - a_1$, la suite $u = (u_n)_{n\in\mathbb{N}}$ est un élément du noyau de $\mathrm{P}(\mathrm{T})$. L'hypothèse sur les a_i nous assure que $\mathrm{P}(1) = 0$ et P s'écrit donc $\mathrm{P} = (\mathrm{X}-1)^r\mathrm{Q}$ avec $r \geqslant 1$ et $\mathrm{Q} \in \mathbb{C}[\mathrm{X}]$, $\mathrm{Q}(1) \neq 0$. Il s'avère que la suite $\mathrm{Q}(\mathrm{T})(u) \in \mathrm{Ker}(\mathrm{T} - \mathrm{Id})^r$ et qu'il s'agit donc d'un polynôme d'après le lemme donnant la forme des suites à récurrence linéaire, démontré à l'occasion de l'exercice précédent. Or, u étant supposée convergente, il en est de même pour $\mathrm{Q}(\mathrm{T})(u)$ par théorème d'opérations ce qui entraîne que $\mathrm{Q}(\mathrm{T})(u)$ est une constante : en écrivant $\mathrm{Q} = \alpha_{p-1}\mathrm{X}^{d-1} + \cdots + \alpha_1\mathrm{X} + \alpha_0$ ($\deg \mathrm{Q} \leqslant p-1$), pour tout $n \in \mathbb{N}$, on a

$$\alpha_{p-1}u_{n+p-1} + \cdots + \alpha_1 u_{n+1} + \alpha_0 u_n = \alpha_{p-1}u_{p-1} + \cdots + \alpha_1 u_1 + \alpha_0 u_0.$$

En notant ℓ la limite de $(u_n)_{n\in\mathbb{N}}$, et en faisant tendre n vers l'infini, il vient $\mathrm{Q}(1)\ell = \alpha_{p-1}u_{p-1} + \cdots + \alpha_1 u_1 + \alpha_0 u_0$ ce qui donne

$$\boxed{\ell = \frac{\alpha_{p-1}u_{p-1} + \cdots + \alpha_1 u_1 + \alpha_0 u_0}{\mathrm{Q}(1)}}.$$

Remarquons que d'après la formule de Taylor, $\mathrm{Q}(1) = \dfrac{\mathrm{P}^{(r)}(1)}{r!}$. En particulier, si $r = 1$, on a $\mathrm{Q}(1) = \mathrm{P}'(1)$ et la factorisation suivante

$$\mathrm{P} = (\mathrm{X}-1)\Big(\mathrm{X}^{p-1} + (a_{p-1} + \cdots + a_1)\mathrm{X}^{p-2} + \cdots + (a_2 + a_1)\mathrm{X} + a_1\Big),$$

si bien que

$$\ell = \frac{u_{p-1} + (a_{p-1} + \cdots + a_1)u_{p-2} + \cdots + (a_2 + a_1)u_1 + a_1 u_0}{p - (p-1)a_p - \cdots - a_2}. \quad \triangleleft$$

Il y a une autre solution utilisant les séries entières qui a le mérite de ne pas utiliser le théorème de structure de l'espace des suites à récurrence linéaire : on multiplie la relation de récurrence par x^{n+d} et on somme

de $n = 0$ à l'infini. La série $\sum u_n x^n$ a un rayon de convergence $\geqslant 1$ (exactement 1 si $\ell \neq 0$). En notant f la somme de cette série, on a alors pour $0 < |x| < 1$, $(1 - a_p x - \cdots - a_1 x^p) f(x) = x^p \mathrm{P}\left(\dfrac{1}{x}\right) f(x) = g(x)$, où g est une fonction polynôme. On obtient ℓ en faisant tendre x vers 1, car $\lim\limits_{x \to 1} (1 - x) f(x) = \ell$ (cf exercice 3.22 du tome 2 d'analyse). En particulier si 1 est racine simple de P, on a $\ell = \dfrac{g(1)}{\mathrm{P}'(1)}$.

Les exercice suivants concernent des suites récurrentes diverses qui n'entrent pas dans les catégories précédentes.

2.39. Étude d'une suite récurrente

On considère la suite u définie par $u_0 = x \geqslant 0$ et la relation de récurrence $u_{n+1} = \sqrt{1 + (u_0 + u_1 + \cdots + u_n)^2}$. Déterminer la limite de $\dfrac{2^n}{u_n}$. On pourra poser $\sin \theta_n = \dfrac{1}{u_n}$.

(École polytechnique)

Solution.

On observe que la suite u est bien définie et strictement croissante. Il est assez naturel de comparer la suite à 2^n. En effet, si on modifie la relation de récurrence en enlevant le 1, c'est-à-dire si on regarde la suite v définie par $v_0 = x$ et pour tout $n \in \mathbb{N}$, $v_{n+1} = v_0 + v_1 + \cdots + v_n$, on constate que $v_n = 2^{n-1} x$ pour tout $n \geqslant 1$. Ce travail prouve déjà que $u_n \to +\infty$, puisqu'une récurrence très facile montre que $u_n \geqslant v_n$ pour tout n.

Pour $n \geqslant 1$, on a $u_n \geqslant 1$. Suivons l'indication de l'énoncé, en considérant l'unique réel $\theta_n \in]0, \dfrac{\pi}{2}]$ tel que $\sin \theta_n = \dfrac{1}{u_n}$. On a alors

$$(u_0 + u_1 + \cdots + u_n)^2 = u_{n+1}^2 - 1 = \frac{1}{\sin^2 \theta_{n+1}} - 1 = \cot^2 \theta_{n+1}.$$

Il en résulte que

$$\begin{aligned} \cot \theta_{n+1} &= u_0 + u_1 + \cdots + u_n = \cot \theta_n + u_n \\ &= \cot \theta_n + \frac{1}{\sin \theta_n} = \frac{\cos \theta_n + 1}{\sin \theta_n} = \cot \frac{\theta_n}{2}, \end{aligned}$$

puis que, pour $n \geqslant 1$, $\theta_{n+1} = \dfrac{\theta_n}{2}$ et donc enfin $\theta_n = \dfrac{\theta_1}{2^{n-1}}$. Ceci permet de conclure que

$$\frac{2^n}{u_n} = 2^n \sin \frac{\theta_1}{2^{n-1}} \underset{n \to +\infty}{\longrightarrow} 2\theta_1 = 2 \arcsin \frac{1}{u_1} = 2 \arcsin \frac{1}{\sqrt{1 + x^2}}. \quad \triangleleft$$

Dans l'exemple suivant, où on ne peut invoquer aucune méthode générale, on essaie de se ramener à l'étude d'une récurrence d'ordre 1.

2.40. Suite vérifiant $u_{n+1} = \ln(1 + u_n) + \ln(1 + u_{n-1})$

Étudier la suite définie par $u_0 > 0$ et $u_1 > 0$ et, pour $n \geqslant 1$,

$$u_{n+1} = \ln(1 + u_n) + \ln(1 + u_{n-1}).$$

(École polytechnique)

▷ **Solution.**

Une petite simulation numérique sur un ordinateur ou une calculatrice montre que, quel que soit le choix de u_0 et u_1, la suite $(u_n)_{n\in\mathbb{N}}$ semble converger vers un réel strictement positif. Si $(u_n)_{n\in\mathbb{N}}$ converge, le logarithme étant continu cela ne peut être que vers un point fixe de la fonction $x \longmapsto 2\ln(1 + x)$ sur $\mathbb{R}_+$. Une rapide étude de fonction laisse apparaître deux points fixes, 0 et ℓ (qui vaut 2,51 à 10^{-2} près).

Nous allons vérifier que $(u_n)_{n\in\mathbb{N}}$ converge vers ℓ, point fixe strictement positif de $x \longmapsto 2\ln(1 + x)$. Pour cela, nous allons encadrer u_n par deux suites monotones vérifiant la même relation de récurrence. Soit $(v_n)_{n\in\mathbb{N}}$ et $(w_n)_{n\in\mathbb{N}}$ deux suites vérifiant la même relation de récurrence que $(u_n)_{n\in\mathbb{N}}$ avec $v_0 = v_1 = m$, $w_0 = w_1 = \mathrm{M}$, m étant un réel strictement positif strictement inférieur à u_0, u_1 et ℓ, M un réel strictement supérieur à u_0, u_1 et ℓ. Une récurrence immédiate prouve alors que pour tout n entier naturel, $v_n \leqslant u_n \leqslant w_n$.

Démontrons que la suite $(v_n)_{n\in\mathbb{N}}$ est croissante. Vérifions par récurrence sur $n \geqslant 1$ que $v_n - v_{n-1} \geqslant 0$. On a $v_1 - v_0 = 0$, $v_2 - v_1 = 2\ln(1 + m) - m \geqslant 0$ car $m \in [0, \ell]$ et la fonction $x \longmapsto f(x) - x$ reste positive pour x entre 0 et ℓ. Pour n plus grand que 3,

$$v_n - v_{n-1} = \ln(1 + v_{n-1}) - \ln(1 + v_{n-3}) \geqslant 0,$$

car par hypothèse de récurrence, $v_{n-1} \geqslant v_{n-2} \geqslant v_{n-3}$.

On démontre, de même, que la suite $(w_n)_{n\in\mathbb{N}}$ est décroissante en établissant, par récurrence, que, pour $n \geqslant 1$, $w_n - w_{n-1} \leqslant 0$. On a $w_1 - w_0 = 0$ et $w_2 - w_1 = 2\ln(1 + \mathrm{M}) - \mathrm{M} \leqslant 0$ car $f(x) - x$ est négatif pour x plus grand que ℓ. Ensuite pour $n \geqslant 3$,

$$w_n - w_{n-1} = \ln(1 + w_{n-1}) - \ln(1 + w_{n-3}) \leqslant 0$$

par hypothèse de récurrence. Ainsi, pour $n \geqslant 0$, on a

$$0 < v_0 \leqslant v_n \leqslant v_{n+1} \leqslant w_{n+1} \leqslant w_n \leqslant w_0.$$

On en déduit que les suites $(v_n)_{n\in\mathbb{N}}$ et $(w_n)_{n\in\mathbb{N}}$ sont convergentes vers un réel, qui par passage à la limite dans la relation de récurrence, ne peut être que ℓ. Comme pour tout entier naturel n, $v_n \leqslant u_n \leqslant w_n$, le théorème d'encadrement permet de conclure que

$$\boxed{\lim_{n\to+\infty} u_n = \ell}.\triangleleft$$

Dans l'exercice suivant, on utilisera le fait qu'une suite bornée converge dès qu'elle ne possède qu'une seule valeur d'adhérence.

2.41. Suite vérifiant $u_{n+1} = \dfrac{f(u_1)+\cdots+f(u_n)}{n}$

Soit $f : [0,1] \to [0,1]$ continue. On définit la suite (u_n) par la relation de récurrence $u_{n+1} = \dfrac{f(u_1)+\cdots+f(u_n)}{n}$ et $u_1 = a \in [0,1]$. Montrer que u converge vers un point fixe de f.

(**École normale supérieure**)

Solution.

Notons déjà que si u converge vers a, alors a est un point fixe de la fonction f. En effet, $f(u_n)$ tend vers $f(a)$ par continuité de f donc $u_{n+1} = \dfrac{f(u_1)+\cdots+f(u_n)}{n}$ tend aussi vers $f(a)$ par le théorème de Cesàro. Or u_{n+1} tend vers a donc par unicité de la limite on a $a = f(a)$.

Observons aussi que

$$u_{n+1} - u_n = \frac{f(u_n)}{n} - \frac{f(u_1)+\cdots+f(u_{n-1})}{n(n-1)} = \frac{f(u_n)-u_n}{n}$$

et que cela tend vers 0 car u_n et $f(u_n)$ sont des suites bornées. On en déduit notamment que si u_N est un point fixe de f, alors la suite est stationnaire à partir de ce rang.

Pour prouver que u_n converge, il suffit de montrer qu'elle n'admet qu'une seule valeur d'adhérence[5]. Supposons donc par l'absurde que la suite possède deux valeurs d'adhérence $\alpha < \beta$. On va montrer que tout x de $]\alpha,\beta[$ est un point fixe de f. En effet, supposons qu'il existe $c \in]\alpha,\beta[$ avec par exemple $f(c) > c$. Par continuité de f, on peut trouver un intervalle $[a,b]$ tel que $\alpha < a < c < b < \beta$ et tel que $f(x) \geqslant x$ pour tout $x \in [a,b]$. Comme $u_{n+1} - u_n$ tend vers 0, on peut trouver un rang N tel

[5] D'après la constatation précédente l'ensemble des valeurs d'adhérence est un segment, mais cela ne va pas servir ici.

que $|u_{n+1}-u_n| < b-a$ pour $n \geqslant N$. Comme β est valeur d'adhérence, on peut trouver $p \geqslant N$ tel que $u_p \geqslant a$. Mais il est alors impossible à la suite u de repasser à gauche de a. En effet, si u_n est dans l'intervalle $[a,b]$ on a $u_{n+1} \geqslant u_n$ et si $u_n > b$, $u_{n+1} \geqslant a$ car l'écart entre u_n et u_{n+1} ne dépasse pas $b-a$. Cela est contradictoire avec le fait que α est valeur d'adhérence. On raisonne de manière analogue dans le cas où il existe $c \in]\alpha,\beta[$ avec $f(c) < c$. On a donc $f(x) = x$ pour tout $x \in]\alpha,\beta[$. Mais cela n'est pas possible car dès que la suite tombe dans cet intervalle (et cela arrive car pour n assez grand on a $|u_{n+1} - u_n| < \beta - \alpha$ et les termes de la suite ne peuvent pas passer d'un voisinage de α à un voisinage de β sans rencontrer l'intervalle $]\alpha,\beta[$), elle est stationnaire. D'où le résultat. ◁

Les exercices suivants étudient des suites vérifiant des relations de récurrence simultanées. L'étude de telles suites se ramène souvent à celle de deux suites adjacentes.

2.42. Suites de Schwob

Soit $a < b$ deux réels strictement positifs. Étudier les deux suites $(x_n)_{n\geqslant 0}$ et $(y_n)_{n\geqslant 0}$ définies par $x_0 = a$, $y_0 = b$ et pour tout $n \in \mathbb{N}$,

$$x_{n+1} = \frac{x_n + y_n}{2} \quad \text{et} \quad y_{n+1} = \sqrt{x_{n+1}y_n}.$$

(École polytechnique)

▷ **Solution.**

On a $x_0 < x_1 < y_0$, puis $x_0 < x_1 < y_1 < y_0$. Une récurrence immédiate montre que pour tout n, $x_n < x_{n+1} < y_{n+1} < y_n$. Les suites (x_n) et (y_n) sont convergentes car monotones et bornées. Si $\ell = \lim x_n$ et $\ell' = \lim y_n$, on a $\ell' = \dfrac{\ell+\ell'}{2}$ donc $\ell = \ell'$. En fait, cette limite commune peut être calculée.

Soit $\theta \in \left]0, \dfrac{\pi}{2}\right[$ tel que $a = b\cos\theta$. On a alors $x_1 = b\cos^2\dfrac{\theta}{2}$ et $y_1 = b\cos\dfrac{\theta}{2}$. Par récurrence sur n, on voit alors que $y_n = b\displaystyle\prod_{k=1}^{n}\cos\frac{\theta}{2^k}$ et $x_n = y_n\cos\dfrac{\theta}{2^n}$. Le produit définissant y_n se simplifie si on le multiplie par $\sin\dfrac{\theta}{2^n}$ et on obtient $y_n = \dfrac{b\sin\theta}{2^n\sin\dfrac{\theta}{2^n}}$. On en déduit que

$$\boxed{\lim_{n\to+\infty} x_n = \lim_{n\to+\infty} y_n = b\frac{\sin\theta}{\theta}}. \quad \triangleleft$$

Les suites (a_n) et (b_n) étudiées dans le prochain exercice sont définies de la manière suivante : a_{n+1} et b_{n+1} sont respectivement la moyenne arithmétique et la moyenne harmonique de a_n et b_n. On va montrer qu'elles convergent vers la moyenne géométrique de a_0 et b_0. Si on remplace la moyenne harmonique de a_n et b_n par leur moyenne géométrique, on obtient encore deux suites qui convergent vers la même limite appelée moyenne arithmético-géométrique de a_0 et b_0.

2.43. Moyenne arithmético-harmonique

On définit les suites $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ par la donnée de $a_0 > 0$ et $b_0 > 0$ et les relations de récurrence

$$a_{n+1} = \frac{a_n + b_n}{2} \quad \text{et} \quad \frac{1}{b_{n+1}} = \frac{1}{2}\left(\frac{1}{a_n} + \frac{1}{b_n}\right).$$

1. Étudier la convergence de ces suites.

2. Soit $\ell = \lim a_n$. Donner un équivalent de $a_n - \ell$.

(École polytechnique)

▷ **Solution.**

1. Les deux suites sont à termes strictement positifs. On a, pour tout $n \in \mathbb{N}$,

$$a_{n+1} - b_{n+1} = \frac{a_n + b_n}{2} - \frac{2a_nb_n}{a_n + b_n} = \frac{(a_n - b_n)^2}{2(a_n + b_n)} \geqslant 0.$$

Autrement dit $a_n \geqslant b_n$, pour tout $n \in \mathbb{N}^*$. On en déduit que, si $n \geqslant 1$,

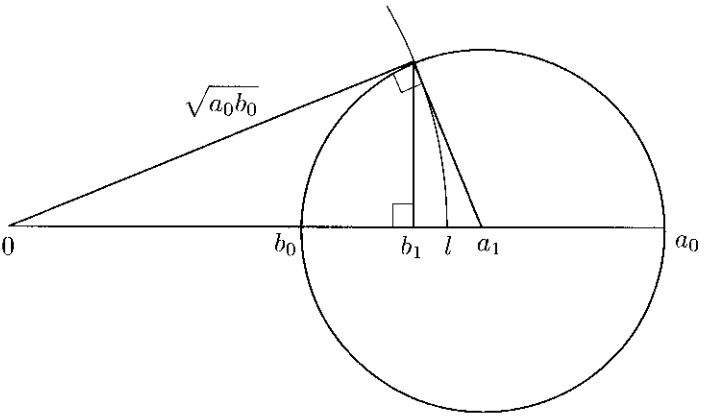
$$a_{n+1} - a_n = \frac{b_n - a_n}{2} \leqslant 0 \quad \text{et} \quad b_{n+1} - b_n = \frac{b_n(a_n - b_n)}{a_n + b_n} \geqslant 0.$$

La suite $(a_n)_{n\geqslant 1}$ décroît et est minorée par b_1 ; la suite $(b_n)_{n\geqslant 1}$ croît et est majorée par a_1. Elles convergent donc toutes les deux. En passant à la limite dans l'égalité $a_{n+1} = \dfrac{a_n + b_n}{2}$, on obtient qu'elles ont même limite. On appelle ℓ cette limite commune. Pour calculer ℓ, on remarque que, pour tout $n \in \mathbb{N}$, on a

$$b_{n+1} = \frac{a_n + b_n}{2a_nb_n} = \frac{a_{n+1}}{a_nb_n}$$

et donc $a_{n+1}b_{n+1} = a_nb_n$: la suite (a_nb_n) est constante et donc, pour tout n, $a_nb_n = a_0b_0$. En faisant tendre n vers $+\infty$, cela donne $\ell = \sqrt{a_0b_0}$.

La figure suivante montre la construction géométrique des termes des suites. La convergence semble assez rapide.



2. Pour répondre à cette question, on va d'abord calculer a_n. Pour tout $n \in \mathbb{N}$, on a

$$a_{n+1} = \frac{1}{2}\left(a_n + \frac{\ell^2}{a_n}\right) = \frac{a_n^2 + \ell^2}{2a_n},$$

et donc

$$a_{n+1} - \ell = \frac{(a_n - \ell)^2}{2a_n} \quad \text{et} \quad a_{n+1} + \ell = \frac{(a_n + \ell)^2}{2a_n}.$$

On obtient finalement $\dfrac{a_{n+1} - \ell}{a_{n+1} + \ell} = \left(\dfrac{a_n - \ell}{a_n + \ell}\right)^2$. Une récurrence aisée donne alors, pour tout $n \in \mathbb{N}$, $\dfrac{a_n - \ell}{a_n + \ell} = \left(\dfrac{a_0 - \ell}{a_0 + \ell}\right)^{2^n}$. Sachant que $a_n + \ell$ est équivalent à 2ℓ, on obtient

$$\boxed{a_n - \ell \underset{n\to\infty}{\sim} 2\ell \left(\frac{a_0 - \ell}{a_0 + \ell}\right)^{2^n}.} \quad \triangleleft$$

La suite (a_n) ci-dessus vérifie la récurrence $a_{n+1} = \dfrac{1}{2}\left(a_n + \dfrac{a_0 b_0}{a_n}\right)$. On reconnaît l'algorithme dit babylonien d'extraction d'une racine carrée, en l'occurrence $\sqrt{a_0 b_0}$. Il est basé sur le raisonnement suivant : si a_n est une valeur approchée par défaut (resp. par excès) de $\sqrt{a}$, alors $\dfrac{a}{a_n}$ en est une valeur approchée par excès (resp. par défaut) et la moyenne $a_{n+1} = \dfrac{1}{2}\left(a_n + \dfrac{a}{a_n}\right)$ est une meilleure valeur approchée que a_n. L'équivalent obtenu dans l'exercice montre que cet algorithme converge très rapidement si a_0 est proche de $\sqrt{a}$.

2.44. Récurrences simultanées

On se donne trois réels a_0, b_0, c_0 tels que $-1 \leqslant a_0 \leqslant b_0 \leqslant c_0 \leqslant 1$ et on pose, pour tout entier naturel n,

$$\begin{cases} a_{n+1} = \displaystyle\int_{-1}^{1} \min(x, b_n, c_n)\mathrm{d}x, \\ b_{n+1} = \displaystyle\int_{-1}^{1} \mathrm{mil}(x, a_n, c_n)\mathrm{d}x, \\ c_{n+1} = \displaystyle\int_{-1}^{1} \max(x, a_n, b_n)\mathrm{d}x, \end{cases}$$

où $\mathrm{mil}(a, b, c)$ est le terme médian de (a, b, c). Étudier la convergence des suites (a_n), (b_n) et (c_n).

(**École polytechnique**)

Solution.

Montrons, par récurrence sur n, que $a_n \leqslant b_n \leqslant c_n$. C'est vérifié pour $n = 0$ et si on suppose que la propriété est vraie au rang n, on obtient, pour tout $x \in [-1, 1]$:

⋆ si $x \leqslant a_n$,

$$\min(x, b_n, c_n) = x \leqslant a_n = \mathrm{mil}(x, a_n, c_n) \leqslant \max(x, a_n, b_n) = b_n;$$

⋆ si $a_n \leqslant x \leqslant c_n$,

$$\min(x, b_n, c_n) \leqslant x = \mathrm{mil}(x, a_n, c_n) \leqslant \max(x, a_n, b_n);$$

⋆ si $x \geqslant c_n$,

$$\min(x, b_n, c_n) = b_n \leqslant c_n = \mathrm{mil}(x, a_n, c_n) \leqslant x = \max(x, a_n, b_n).$$

On a donc, pour tout $x \in [-1, 1]$,

$$\min(x, b_n, c_n) \leqslant \mathrm{mil}(x, a_n, c_n) \leqslant \max(x, a_n, b_n).$$

On en déduit, en intégrant ces inégalités sur $[-1, 1]$ que

$$a_{n+1} \leqslant b_{n+1} \leqslant c_{n+1},$$

ce qui termine la récurrence.

On aimerait pouvoir dire que, pour tout n, $-1 \leqslant a_n \leqslant b_n \leqslant c_n \leqslant 1$. Il n'en est rien : si par exemple $b_0 = -1$, alors $a_0 = -2$. Nous allons cependant montrer que c'est réalisé pour n assez grand.

On a, pour tout $n \in \mathbb{N}$, $a_{n+1} \leqslant \displaystyle\int_{-1}^{1} x\mathrm{d}x = 0$ et de même $c_{n+1} \geqslant 0$. Autrement dit, $a_n \leqslant 0$ et $c_n \geqslant 0$ pour $n \geqslant 1$. On en déduit que, pour $n \geqslant 1$,

$$b_{n+1} + 1 = \int_{-1}^{1} \mathrm{mil}(x + 1, a_n + 1, c_n + 1)\mathrm{d}x \geqslant 0,$$

car $c_n + 1 \geqslant 0$ et pour tout $x \in [-1, 1]$, $x + 1 \geqslant 0$. On montre de même que $b_{n+1} - 1 \leqslant 0$. On a donc $b_{n+1} \in [-1, 1]$ pour $n \geqslant 1$, soit $b_n \in [-1, 1]$ pour $n \geqslant 2$. On peut donc écrire, pour $n \geqslant 2$,

$$\begin{aligned} a_{n+1} &= \int_{-1}^{1} \min(x, b_n)\mathrm{d}x = \int_{-1}^{b_n} x\mathrm{d}x + \int_{b_n}^{1} b_n\mathrm{d}x \\ &= \frac{1}{2}(b_n^2 - 1) + b_n(1 - b_n) = -\frac{(1 - b_n)^2}{2}. \end{aligned}$$

On trouve de même,

$$c_{n+1} = \int_{-1}^{b_n} b_n\mathrm{d}x + \int_{b_n}^{1} x\mathrm{d}x = \frac{(b_n + 1)^2}{2}.$$

Soit n un entier $\geqslant 2$. Les conditions $a_{n+1} \geqslant -1$ et $c_{n+1} \leqslant 1$ sont réalisées si $|b_n| \leqslant \sqrt{2} - 1$. On note que $|c_{n+1}| + |a_{n+1}| = c_{n+1} - a_{n+1} = 1 + b_n^2 \leqslant 2$. On ne peut pas avoir simultanément $a_{n+1} < -1$ et $c_{n+1} > 1$.

⋆ Si $|b_n| \leqslant \sqrt{2} - 1$, alors

$$\begin{aligned} b_{n+2} &= \int_{-1}^{a_{n+1}} a_{n+1}\mathrm{d}x + \int_{a_{n+1}}^{c_{n+1}} x\mathrm{d}x + \int_{c_{n+1}}^{1} c_{n+1}\mathrm{d}x \\ &= \frac{1}{2}(a_{n+1} + c_{n+1})(a_{n+1} - c_{n+1} + 2) \\ &= b_n(1 - b_n^2). \end{aligned}$$

On a alors $|b_{n+2}| \leqslant |b_n| \leqslant \sqrt{2} - 1$.

⋆ Si $b_n > \sqrt{2} - 1$, alors $c_{n+1} > 1$ et

$$\begin{aligned} b_{n+2} &= \int_{-1}^{1} \max(x, a_{n+1})\mathrm{d}x = \int_{-1}^{a_{n+1}} a_{n+1}\mathrm{d}x + \int_{a_{n+1}}^{1} x\mathrm{d}x \\ &= a_{n+1}(a_{n+1} + 1) + \frac{1}{2}(1 - a_{n+1}^2) = \frac{(a_{n+1} + 1)^2}{2}. \end{aligned}$$

On en déduit $b_{n+2} \in \left[0, \frac{1}{2}\right]$, puis $a_{n+3} = -\frac{(1 - b_{n+2})^2}{2} \in \left[-\frac{1}{2}, -\frac{1}{8}\right]$. Si on a encore $b_{n+2} > \sqrt{2} - 1$, alors on obtient $b_{n+4} = \frac{(a_{n+3} + 1)^2}{2}$ et, cette fois, on a $0 \leqslant b_{n+4} \leqslant \frac{49}{128} \leqslant \sqrt{2} - 1$.

⋆ On montre de même que si $b_n < -\sqrt{2} + 1$, alors $-\sqrt{2} + 1 \leqslant b_{n+4} \leqslant 0$.

On a donc, pour tout n assez grand ($n \geqslant 6$),

$$|b_n| \leqslant \sqrt{2} - 1, \quad b_{n+2} = b_n(1 - b_n^2) \quad \text{et} \quad |b_{n+2}| \leqslant |b_n|.$$

Les deux suites $(|b_{2n}|)$ et $(|b_{2n+1}|)$ sont donc décroissantes et bornées. Elles convergent. La limite de chacune vérifie $\ell = \ell(1 - \ell^2)$ et donc $\ell = 0$.

La suite (b_n) converge donc vers 0. Des relations $a_{n+1} = -\dfrac{(1-b_n)^2}{2}$ et $c_{n+1} = \dfrac{(b_n+1)^2}{2}$, on déduit que les suites (a_n) et (c_n) convergent respectivement vers $-\dfrac{1}{2}$ et $\dfrac{1}{2}$. ◁

Les exercices qui terminent ce chapitre étudient des suites définies implicitement. La démarche est toujours la même : montrer l'existence de x_n, démontrer la convergence de la suite x_n, déterminer un équivalent, puis un développement asymptotique de x_n. Pour cela il faut commencer par déterminer dans la relation qui définit x_n quels sont les termes prépondérants.

2.45. Suite définie implicitement (1)

On considère le polynôme $P_n = X^n + X^{n-1} + 2X - 1$.

1. Montrer que pour tout $n \geqslant 2$, il existe un unique réel $x_n > 0$ tel que $P_n(x_n) = x_n$.

2. Montrer que la suite $(x_n)_{n\in\mathbb{N}}$ est croissante et converge vers 1.

(**École polytechnique**)

Solution.

1. Posons, pour $n \geqslant 3$, $Q_n = P_n - X = X^n + X^{n-1} + X - 1$. On a $Q_n' = nX^{n-1} + (n-1)X^{n-2} + 1$. Ce polynôme est strictement positif sur $\mathbb{R}_+$. Il en résulte que Q_n réalise un homéomorphisme strictement croissant de $\mathbb{R}_+$ sur $[Q_n(0), +\infty[= [-1, +\infty[$. En particulier, il existe un unique réel $x_n > 0$ tel que $Q_n(x_n) = 0$.

2. Notons que puisque $Q_n(1) = 2 > 0$, on a $0 < x_n < 1$. En évaluant $Q_{n+1} - Q_n = (X+1)X^{n-1}(X-1)$ en x_n, on voit que $Q_{n+1}(x_n) < 0$. On a donc $x_{n+1} > x_n$: la suite $(x_n)_{n\in\mathbb{N}}$ est strictement croissante. Comme elle est majorée par 1, elle converge. Il ne reste plus qu'à prouver que sa limite est 1.

Soit $\varepsilon \in]0,1[$. On calcule $Q_n(1-\varepsilon) = (1-\varepsilon)^n + (1-\varepsilon)^{n-1} - \varepsilon$. On observe que $\lim\limits_{n\to+\infty} Q_n(1-\varepsilon) = -\varepsilon$. Pour n assez grand, on a donc $Q_n(1-\varepsilon) < 0$ et ainsi $1-\varepsilon < x_n < 1$. Cela prouve que (x_n) tend vers 1. ◁

2.46. Suite définie implicitement (2)

Soit a_n la plus grande racine réelle de $X^{2n} - 2nX + 1$. Donner un développement asymptotique à deux termes de a_n.

(**École polytechnique**)

▷ **Solution.**

Posons $f_n(x) = x^{2n} - 2nx + 1$. On a $f_n'(x) = 2n(x^{2n-1} - 1)$. La fonction f_n est donc strictement décroissante sur $]-\infty, 1]$ et strictement croissante sur $[1, +\infty[$. Comme $f_n(1) = 2(1-n) < 0$, la fonction f_n s'annule exactement deux fois pour $n > 1$. On a en particulier $a_n > 1$. Comme $f_n(2) = 4^n - 4n + 1 > 0$ on a $1 < a_n < 2$. Posons alors $a_n = 1 + \varepsilon_n$. En remplaçant dans l'égalité $f_n(a_n) = 0$, il vient

$$(1+\varepsilon_n)^{2n} = 2n(1+\varepsilon_n) + 1.$$

En passant au logarithme et en divisant par $2n$ on obtient,

$$\ln(1+\varepsilon_n) \sim \frac{\ln(2n)}{2n} \sim \frac{\ln n}{2n}.$$

Donc ε_n tend vers 0 et $\varepsilon_n \sim \ln(1+\varepsilon_n) \sim \dfrac{\ln n}{2n}$. On a donc le développement asymptotique suivant :

$$a_n = 1 + \frac{\ln n}{2n} + o\left(\frac{\ln n}{n}\right). \quad \triangleleft$$

2.47. Suite définie implicitement (3)

Soit $(u_n)_{n\in\mathbb{N}}$ une suite de réels telle que $u_n^5 + nu_n - 1 = 0$ pour tout $n \in \mathbb{N}$. Étudier cette suite et donner un développement asymptotique de u_n comportant deux termes.

(École polytechnique)

▷ **Solution.**

Considérons $f_n : x \longmapsto x^5 + nx - 1$ pour $n \geqslant 1$. La fonction f_n est dérivable et $f_n'(x) = x^4 + n > 0$. Elle est donc strictement croissante sur $\mathbb{R}$. Comme $f(0) = -1$ et $f(1) = n \geqslant 0$, le théorème des valeurs intermédiaires garantit l'existence de u_n, qui est dans l'intervalle $[0, 1]$. Son unicité est due à la stricte monotonie de f_n.

Pour $n \in \mathbb{N}$, $f_{n+1}(u_n) = u_n^5 + (n+1)u_n - 1 = f_n(u_n) + u_n = u_n \geqslant 0$. Comme f_{n+1} est strictement croissante, $u_{n+1} \leqslant u_n$. La suite $(u_n)_{n\in\mathbb{N}}$ est décroissante ; étant minorée, elle converge vers un réel $\ell \in [0, 1[$. On a $\lim\limits_{n\to+\infty} u_n^5 - 1 = \ell^5 - 1 = \lim\limits_{n\to+\infty} -nu_n$. Nécessairement, $(u_n)_{n\in\mathbb{N}}$ converge vers 0.

Comme u_n tend vers 0 lorsque n tend vers l'infini,

$$\lim_{n\to+\infty} nu_n = 1 \quad i.e. \quad u_n \underset{n\to\infty}{\sim} \frac{1}{n}.$$

De la relation $u_n^5 + nu_n - 1 = 0$, on tire pour n tendant vers l'infini,

$$nu_n - 1 = -u_n^5 \underset{n\to\infty}{\sim} -\frac{1}{n^5}, \quad i.e. \quad nu_n = 1 - \frac{1}{n^5} + o\left(\frac{1}{n^5}\right).$$

On conclut

$$\boxed{u_n = \frac{1}{n} - \frac{1}{n^6} + o\left(\frac{1}{n^6}\right).}$$

2.48. Suite définie implicitement (4)

1. Montrer que l'équation $x \sin x - c \cos x$ (où $c > 0$) admet une unique racine x_n dans tout intervalle $\left]n\pi, n\pi + \frac{\pi}{2}\right[$ et qu'elle n'en admet pas d'autre dans $\mathbb{R}_+$.

2. Trouver un équivalent θ_n de $x_n - n\pi$, puis un équivalent de $x_n - n\pi - \theta_n$.

(École polytechnique)

Solution.

1. Posons, pour x réel positif ou nul, $f(x) = x \sin x - c \cos x$. La fonction f est de classe $\mathcal{C}^\infty$ et sa dérivée vaut $f'(x) = x \cos x + (1+c) \sin x$. On remarque que sur un intervalle de la forme $\left]n\pi, n\pi + \frac{\pi}{2}\right[$ (n entier naturel), $\sin x$ et $\cos x$ ont le même signe, si bien que f' garde sur cet intervalle un signe constant : la fonction f est strictement monotone sur $\left]n\pi, n\pi + \frac{\pi}{2}\right[$. Or $f(n\pi) f\left(n\pi + \frac{\pi}{2}\right) = -c(-1)^n n\pi (-1)^n = -cn\pi < 0$, donc le théorème des valeurs intermédiaires nous assure l'existence d'une racine x_n de f dans $\left]n\pi, n\pi + \frac{\pi}{2}\right[$, unique par stricte monotonie.

D'autre part, si $f(x) = 0$, $\cos x \neq 0$ (car sinon, $f(x) = x \sin x \neq 0$). Dans ces conditions,

$$\tan x = \frac{\sin x}{\cos x} = \frac{c}{x} > 0.$$

Il s'ensuit que x est dans un intervalle du type $\left]n\pi, n\pi + \frac{\pi}{2}\right[$ ($n \geqslant 0$). En dehors des x_n, la fonction f n'a donc pas d'autre racine positive.

2. Pour n entier naturel non nul, $x_n \in \left]n\pi, n\pi + \frac{\pi}{2}\right[$ et donc

$$1 \leqslant \frac{x_n}{n\pi} \leqslant 1 + \frac{\pi}{2n\pi}.$$

Le théorème d'encadrement nous donne l'équivalent $x_n \sim n\pi$.

Posons $u_n = x_n - n\pi \in \left]0, \frac{\pi}{2}\right[$. La relation $f(x_n) = 0$ s'écrit

$$(-1)^n (u_n + n\pi) \sin u_n - c(-1)^n \cos u_n = 0,$$

ce qui se simplifie comme suit

$$\tan u_n = \frac{c}{n\pi + u_n} \text{ ou encore } u_n = \arctan \frac{c}{n\pi + u_n} \sim \frac{c}{n\pi + u_n} \sim \frac{c}{n\pi},$$

puisque $u_n = o(n\pi)$ et $\lim_{n\to+\infty} \dfrac{c}{n\pi + u_n} = 0$.

Posons ensuite $v_n = u_n - \dfrac{c}{n\pi} = o\left(\dfrac{1}{n}\right)$. La relation $u_n = \arctan \dfrac{c}{n\pi + u_n}$ devient

$$v_n + \frac{c}{n\pi} = \arctan\left(\frac{c}{n\pi + \dfrac{c}{n\pi} + o\left(\dfrac{1}{n}\right)}\right).$$

On a

$$\frac{c}{n\pi + \dfrac{c}{n\pi} + o\left(\dfrac{1}{n}\right)} = \frac{c}{n\pi\left(1 + \dfrac{c}{n^2\pi^2} + o\left(\dfrac{1}{n^2}\right)\right)} = \frac{c}{n\pi} - \frac{c^2}{n^3\pi^3} + o\left(\frac{1}{n^3}\right).$$

Sachant que pour x tendant vers 0, $\arctan x = x - \dfrac{1}{3}x^3 + o(x^3)$, on obtient

$$v_n + \frac{c}{n\pi} = \frac{c}{n\pi} - \frac{c^2}{n^3\pi^3} - \frac{1}{3}\frac{c^3}{n^3\pi^3} + o\left(\frac{1}{n^3}\right),$$

ce qui donne après simplification, $v_n = -\dfrac{c^2(c+3)}{3n^3\pi^3} + o\left(\dfrac{1}{n^3}\right)$ et finalement

$$\boxed{x_n = n\pi + \frac{c}{n\pi} - \frac{c^2(c+3)}{3n^3\pi^3} + o\left(\frac{1}{n^3}\right)}. \triangleleft$$

La commande Maple `solve(series(arctan(c/(Pi/y + x)), y) = x, x)` *permet de trouver le résultat.*

2.49. Zéros des polynômes de Taylor d'ordre impair de l'exponentielle

Soit $P_n(X) = 1 + \dfrac{X}{1!} + \cdots + \dfrac{X^n}{n!}$.

1. Montrer que P_n admet au plus une racine réelle.

2. Soit a_n l'unique zéro de P_{2n+1}. Étudier la limite de la suite (a_n).

3. Déterminer un équivalent de a_n.

(École polytechnique)

Solution.

1. Il est clair pour commencer que $P_n(x) \geqslant 1$ pour $x \geqslant 0$. Les éventuels zéros seront donc négatifs. Pour tout $n \geqslant 1$ on a $P'_n = P_{n-1}$. Cela invite clairement à faire une démonstration par récurrence. Prenons l'hypothèse suivante

(H_n) : la fonction P_{2n} est strictement positive sur $\mathbb{R}$ et P_{2n+1} réalise un homéomorphisme strictement croissant de $\mathbb{R}$ sur lui-même, et en particulier s'annule pour une unique valeur $a_n < 0$.

Pour $n = 0$ c'est clair : $P_0 = 1$ et $P_1(X) = 1 + X$. On a $a_0 = -1$.

Supposons (H_n) vraie. Alors comme $P'_{2n+2} = P_{2n+1}$, la fonction P_{2n+2} atteint son minimum en a_n.Puisque

$$P_{2n+2}(a_n) = P_{2n+1}(a_n) + \frac{a_n^{2n+2}}{(2n+2)!} = \frac{a_n^{2n+2}}{(2n+2)!} > 0,$$

on a donc $P_{2n+2} > 0$. Enfin, comme $P'_{2n+3} = P_{2n+2}$, la fonction P_{2n+3} est strictement croissante. Comme ses limites en $\pm\infty$ sont respectivement $\mp\infty$ la seconde partie de (H_{n+1}) est vérifiée.

2. Essayons maintenant d'étudier la monotonie de la suite (a_n). Pour localiser a_{n+1} par rapport à a_n il suffit de connaître le signe de $P_{2n+3}(a_n)$. Or nous avons

$$P_{2n+3}(x) - P_{2n+1}(x) = \frac{x^{2n+2}}{(2n+2)!} + \frac{x^{2n+3}}{(2n+3)!} = \frac{x^{2n+2}}{(2n+3)!}(2n+3+x)$$

donc $P_{2n+3}(a_n)$ a le signe de $2n + 3 + a_n$. On remarque que

$$P_{2n+3}(-2n-3) = P_{2n+1}(-2n-3) \leqslant P_{2n+1}(-2n-1).$$

Il en résulte que la suite $(P_{2n+1}(-2n-1))_{n\geqslant 0}$ est décroissante. En particulier, $P_{2n+1}(-2n-1) \leqslant P_1(-1) = 0$ ce qui implique $-2n-1 \leqslant a_n$. En conséquence, $a_n + 2n + 3 \geqslant 2 > 0$, ce qui implique $P_{2n+3}(a_n) > 0$, $a_{n+1} \leqslant a_n$ et la décroissance stricte de la suite (a_n).

On va maintenant prouver que la suite (a_n) tend vers $-\infty$. Supposons qu'elle converge vers une limite finie l. Alors, par croissance stricte des P_{2n+1}, on aurait, pour tout n, $P_{2n+1}(l) < 0$ et donc, en faisant tendre n vers l'infini, $e^l \leqslant 0$ ce qui est absurde. On en déduit que $\lim\limits_{n\to+\infty} a_n = -\infty$.

3. Cherchons un équivalent de a_n. Pour tout réel x, la suite $(P_n(x))$ converge vers e^x. De plus, pour $x < 0$, la série définissant e^x est alternée, donc e^x est encadré par $P_n(x)$ et $P_{n+1}(x)$. D'où l'idée d'encadrer e^{a_n} par $P_{2n+2}(a_n)$ et $P_{2n+3}(a_n)$ (on sait déjà que $P_{2n+1}(a_n) = 0$).

Par la formule de Taylor-Lagrange, pour tout $x < 0$, il existe $c \in]x, 0[$ tel que $e^x - P_{2n+2}(x) = \dfrac{e^c}{(2n+3)!}x^{2n+3} < 0$. En particulier, pour $x = a_n$,

on obtient $e^{a_n} - \dfrac{a_n^{2n+2}}{(2n+2)!} \leqslant 0$. De même, la formule de Taylor-Lagrange montre que $\mathrm{P}_{2n+3}(x) \leqslant e^x$ pour tout réel x et on obtient, pour $x = a_n$,

$$e^{a_n} \geqslant \frac{a_n^{2n+2}}{(2n+2)!} + \frac{a_n^{2n+3}}{(2n+3)!} = \frac{a_n^{2n+2}}{(2n+3)!}(2n+3+a_n) \geqslant 2\frac{a_n^{2n+2}}{(2n+3)!}$$

(on a utilisé l'inégalité $a_n + 2n + 3 \geqslant 2$ démontrée plus haut). En passant au logarithme dans les deux inégalités précédentes, on obtient l'encadrement

$$\ln 2 + (2n+2)\ln(-a_n) - \ln(2n+3)! \leqslant a_n \leqslant (2n+2)\ln(-a_n) - \ln(2n+2)!.$$

On rappelle les premiers termes du développement asymptotique de Stirling

$$\ln k! = k\ln k - k + o(k).$$

Comme $\ln(2n+3)! = \ln(2n+3) + \ln(2n+2)! = (2n+2)\ln(2n+2) - (2n+2) + o(n)$, il vient

$$(2n+2)\ln\frac{-a_n}{2n+2} + (2n+2) + o(n) \leqslant a_n \leqslant (2n+2)\ln\frac{-a_n}{2n+2} + (2n+2) + o(n).$$

En divisant l'inégalité ci-dessus par $2n+2$ et en notant $x_n = -\dfrac{a_n}{2n+2}$, on a donc $1 + \ln(x_n) + o(1) \leqslant -x_n \leqslant 1 + \ln x_n + o(1)$. Il en résulte que $\lim\limits_{n\to+\infty}(x_n + \ln x_n) = -1$. Or, la fonction $\psi : t \longmapsto t + \ln t$ réalise un homéomorphisme strictement monotone de $\mathbb{R}_+^*$ sur $\mathbb{R}$. Par continuité de ψ^{-1}, x_n tend vers l'unique réel α tel que $\alpha + \ln\alpha = -1$ (α vaut environ 0,278) et on a l'équivalent cherché $\boxed{a_n \sim -2\alpha n}$. ◁

Chapitre 3

Séries numériques

Dans une tentative d'historique des séries numériques, nous pourrions faire remonter leurs origines aux travaux développés dès la fin du XVII^{e} *siècle autour du comportement asymptotique de sommes du type* $\sum_{k=1}^{n} f(k)$. *Quelques années après les travaux de Bernoulli sur ce sujet, Euler et Mac-Laurin produisent indépendamment une «formule sommatoire» obtenue par inversion d'identités tayloriennes :*

$$\sum_{k=1}^{n} f(k) = \int_0^n f + \frac{f(0)+f(n)}{2} + \frac{f'(n)-f'(0)}{12} - \frac{f'''(n)-f'''(0)}{720} + \cdots$$

Si l'expression générale donnant les coefficients de cette formule leur échappe dans un premier temps, Euler établit leur lien avec les coefficients du développement en série de $\dfrac{x}{e^x-1}$ *et les nombres de Bernoulli, introduits par celui-ci dans le calcul des sommes* $\sum_{k=1}^{n} k^p$. *Euler en déduit, par de jolis calculs, les sommes des séries* $\sum_{n=1}^{+\infty} \dfrac{1}{n^{2s}}$ *pour* s *entier naturel non nul. Cependant, le problème de la convergence des sommes en question n'est jamais au centre de leurs réflexions et l'aspect formel l'emporte, ce qui conduit parfois les plus grands mathématiciens du siècle à commettre de lourdes erreurs. Vers 1768, d'Alembert commence à douter de la validité de l'emploi de séries non convergentes. En 1826, le mot d'Abel*[1] *illustre parfaitement cette nouvelle préoccupation : «Les séries divergentes sont des inventions du diable, et c'est une honte que l'on ose fonder sur elles la moindre démonstration. On peut en tirer tout ce qu'on veut quand on les emploie et ce sont elles qui ont produit tant d'échecs et tant de paradoxes.»*

Mais ce sont les nécessités du calcul numérique qui imposent vraiment un effort de rigueur dont Gauss, s'étant fait une idée claire de la notion de limite, sera le principal artisan. À partir de là, il paraît naturel d'établir des critères simples de convergence : on en doit plusieurs à Cauchy[2]*, et notamment celui qui porte son nom : si la suite de réels positifs* $(a_n)_{n\geqslant 0}$ *est telle que la limite supérieure de* $\sqrt[n]{a_n}$ *est strictement inférieure à* 1*, alors la série* $\sum a_n$ *est convergente.*

---

1 ABEL (N.), *Œuvres*, 1881.

2 CAUCHY (A.), *Analyse algébrique*, 1821.

Les premiers exercices sont consacrés aux séries à termes positifs. Dans ce cas, la suite des sommes partielles est croissante, de sorte qu'une série à termes positifs converge si et seulement si la suite des sommes partielles est majorée.

3.1. Étude de convergence

Soit $(a_n)_{n\geqslant 1}$ une suite de réels positifs ou nuls telle que la série $\sum a_n$ converge.

1. Montrer que si $\alpha > \dfrac{1}{2}$, la série $\sum \dfrac{\sqrt{a_n}}{n^\alpha}$ converge.
2. Que dire dans le cas $\alpha = \dfrac{1}{2}$?

(École polytechnique)

▷ **Solution.**

1. On peut envisager de majorer la somme partielle $\displaystyle\sum_{n=1}^{N} \frac{\sqrt{a_n}}{n^\alpha}$ à l'aide de l'inégalité de Cauchy-Schwarz pour faire apparaître la somme des a_n sur laquelle nous avons une hypothèse ; en effet, pour tout $N \geqslant 1$, on peut écrire

$$\begin{aligned}\sum_{n=1}^{N} \frac{\sqrt{a_n}}{n^\alpha} &\leqslant \left(\sum_{n=1}^{N} (\sqrt{a_n})^2\right)^{1/2} \left(\sum_{n=1}^{N} \frac{1}{n^{2\alpha}}\right)^{1/2} \\ &\leqslant \left(\sum_{n=1}^{+\infty} a_n\right)^{1/2} \left(\sum_{n=1}^{+\infty} \frac{1}{n^{2\alpha}}\right)^{1/2},\end{aligned}$$

puisque la série de terme général $\dfrac{1}{n^{2\alpha}}$ converge (car $2\alpha > 1$). Les sommes partielles de la série $\sum \dfrac{\sqrt{a_n}}{n^\alpha}$ étant majorées, cette série à termes positifs converge.

On pouvait également majorer directement le terme général en écrivant : $\dfrac{\sqrt{a_n}}{n^\alpha} \leqslant \dfrac{1}{2}\left(\dfrac{1}{n^{2\alpha}} + a_n\right)$ *(c'est tout simplement l'inégalité arithmético-géométrique). Le théorème de comparaison des séries à termes positifs permet alors de conclure.*

2. On ne peut répondre de manière générale : si les a_n sont tous nuls la série $\sum \dfrac{\sqrt{a_n}}{n^{1/2}}$ converge. En revanche si $a_n = \dfrac{1}{n \ln^2 n}$ (pour $n \geqslant 2$), $\dfrac{\sqrt{a_n}}{n^{1/2}} = \dfrac{1}{n \ln n}$ est le terme général d'une série divergente. ◁

L'exercice suivant fournit un premier exemple très simple de transformation d'Abel. On trouvera des développements sur cette technique dans les exercices 3.35 à 3.37.

3.2. Une transformation d'Abel

Soit $(u_n)_{n\geqslant 1}$ une suite positive décroissante qui converge vers 0. Montrer que les séries $\sum u_n$ et $\sum n(u_n - u_{n+1})$ sont de même nature. Dans le cas de la convergence, montrer que

$$\sum_{n=1}^{+\infty} u_n = \sum_{n=1}^{+\infty} n(u_n - u_{n+1}).$$

(École polytechnique)

Solution.

Notons que les séries en question sont à termes positifs. Posons pour $n \geqslant 1$, $S_n = \sum_{k=1}^{n} u_k$ et $T_n = \sum_{k=1}^{n} k(u_k - u_{k+1})$. On a, pour $n \geqslant 1$,

$$\begin{aligned} T_n &= \sum_{k=1}^{n} ku_k - \sum_{k=1}^{n} ku_{k+1} = \sum_{k=1}^{n} ku_k - \sum_{k=2}^{n+1} (k-1)u_k \\ &= \sum_{k=1}^{n} (k-(k-1))u_k + (1-1)u_1 - nu_{n+1} = S_n - nu_{n+1}. \end{aligned}$$

• Supposons que $\sum u_k$ converge et notons S sa somme. Dans ces conditions, on a $T_n \leqslant S_n \leqslant S$ pour tout entier n. Donc, les sommes partielles de la série à termes positifs $\sum n(u_n - u_{n+1})$ sont majorées par S. Cette série converge et sa somme T est, par passage à la limite dans l'inégalité $T_n \leqslant S$, inférieure ou égale à S (on pourrait ici invoquer le résultat de l'exercice 3.10 pour avoir directement S = T, mais ce n'est pas indispensable car on va obtenir l'égalité dans la preuve du sens réciproque).

• Réciproquement, supposons que $\sum n(u_n - u_{n+1})$ converge et notons T sa somme. Posons $v_k = k(u_k - u_{k+1})$. On va montrer que la suite nu_{n+1} tend vers 0, ce qui impliquera à la fois la convergence de la série $\sum u_n$ et le fait que les deux séries ont la même somme d'après l'identité $T_n = S_n - nu_{n+1}$. On a pour tout entier $k \geqslant 1$, $u_k - u_{k+1} = \dfrac{v_k}{k}$, ce qui donne, puisque u_n tend vers 0,

$$u_n = \sum_{k=n}^{+\infty} \frac{v_k}{k}.$$

On en déduit que $0 \leqslant nu_{n+1} \leqslant nu_n \leqslant \sum_{k=n}^{+\infty} v_k$, quantité qui tend vers 0 en tant que reste d'une série convergente. La limite de nu_{n+1} est donc nulle.

Conclusion. Les séries $\sum u_n$ et $\sum v_n$ sont de même nature et en cas de convergence ont la même somme :

$$\boxed{\sum_{n=1}^{+\infty} u_n = \sum_{n=1}^{+\infty} n(u_n - u_{n+1}).} \triangleleft$$

Cet exercice est très souvent posé à l'oral et parfois sous la forme «dérivée» suivante : si $\sum x_n$ est une série à termes positifs qui converge et si $\mathrm{R}_n = \sum_{k=n+1}^{+\infty} x_k$ désigne son reste d'ordre n, alors les séries $\sum \mathrm{R}_n$ et $\sum nx_n$ sont de même nature et en cas de convergence ont la même somme : $\sum_{n=0}^{+\infty} \mathrm{R}_n = \sum_{n=0}^{+\infty} nx_n$. Il s'agit bien entendu du même exercice puisque la suite R_n tend vers 0 en décroissant. Sous cette forme il apparaît très clairement que le résultat découle immédiatement du théorème de Fubini pour les séries doubles positives :

$$\sum_{n=0}^{+\infty} \mathrm{R}_n = \sum_{n=0}^{+\infty} \sum_{k=n+1}^{+\infty} x_k = \sum_{k=1}^{+\infty} \sum_{n=0}^{k-1} x_k = \sum_{k=1}^{+\infty} kx_k.$$

Les théorèmes de comparaison des séries à termes positifs sont des outils puissants qui permettent de conclure sur des problèmes de convergence en étudiant le terme général de la série plutôt que la suite des sommes partielles. Rappelons les résultats les plus utilisés. Si $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ sont deux suites réelles positives, on a

$$(\forall n \in \mathbb{N},\ u_n \leqslant v_n) \text{ et } \sum v_n \text{ converge} \implies \sum u_n \text{ converge};$$

$$u_n = O(v_n) \text{ et } \sum v_n \text{ converge} \implies \sum u_n \text{ converge};$$

$$u_n \underset{n\to\infty}{\sim} v_n \implies \sum u_n \text{ et } \sum v_n \text{ sont de même nature.}$$

Le plus souvent on utilise ces résultats avec les séries de référence que sont les séries de Riemann $\sum \dfrac{1}{n^\alpha}$ (qui convergent si et seulement si $\alpha > 1$) et les séries de Bertrand $\sum \dfrac{1}{n^\alpha \ln^\beta n}$ (qui convergent si et seulement si $\alpha > 1$ ou $\alpha = 1$ et $\beta > 1$). On en trouvera plusieurs exemples dans les exercices qui suivent.

3.3. Nature d'une série (1)

Quelle est la nature de la série $\sum \ln\left(\dfrac{(\ln(n+1))^2}{\ln n \ln(n+2)}\right)$?

(École polytechnique)

Solution.

Observons pour commencer que la série est positive. En effet, on a

$$u_n = \ln\left(\frac{(\ln(n+1))^2}{\ln n.\ln(n+2)}\right) = 2\ln\ln(n+1) - \ln\ln n - \ln\ln(n+2) \geqslant 0$$

par concavité de la fonction $t \longmapsto \ln\ln t$. On va maintenant estimer u_n par un développement asymptotique. Posons $u_n = \ln v_n$. On peut écrire pour n tendant vers $+\infty$

$$\begin{aligned} v_n &= \frac{\left[\ln n + \ln\left(1+\frac{1}{n}\right)\right]^2}{\ln n\left[\ln n + \ln\left(1+\frac{2}{n}\right)\right]} = \frac{\left[\ln n + \frac{1}{n} + O\left(\frac{1}{n^2}\right)\right]^2}{\ln n\left[\ln n + \frac{2}{n} + O\left(\frac{1}{n^2}\right)\right]} \\ &= \frac{\left[1 + \frac{1}{n\ln n} + O\left(\frac{1}{n^2\ln n}\right)\right]^2}{1 + \frac{2}{n\ln n} + O\left(\frac{1}{n^2\ln n}\right)} \\ &= \left(1 + \frac{2}{n\ln n} + O\left(\frac{1}{n^2\ln n}\right)\right) \times \left(1 - \frac{2}{n\ln n} + O\left(\frac{1}{n^2\ln n}\right)\right) \\ &= 1 + O\left(\frac{1}{n^2\ln n}\right), \end{aligned}$$

et par conséquent,

$$u_n = \ln v_n = \ln\left(1 + O\left(\frac{1}{n^2\ln n}\right)\right) = O\left(\frac{1}{n^2\ln n}\right).$$

Comme la série $\sum \dfrac{1}{n^2\ln n}$ converge, le théorème de comparaison des séries à termes positifs assure la convergence de la série étudiée. ◁

En fait on peut même calculer sa somme, comme le lecteur l'aura peut-être fait. Il suffit de constater que dans la somme partielle

$$S_n = \sum_{k=2}^{n} u_k = \sum_{k=2}^{n} (2\ln\ln(k+1) - \ln\ln k - \ln\ln(k+2))$$

la plupart des termes s'éliminent. On obtient

$$S_n = \ln\ln(n+1) - \ln\ln(n+2) + \ln\ln 3 - \ln\ln 2$$

qui tend vers $\ln\ln 3 - \ln\ln 2$ *lorsque* n *tend vers* $+\infty$.

3.4. Nature d'une série (2)

Soit (a_n) une suite réelle strictement positive et (b_n) une suite réelle. On suppose que $b_n = 1 + \mathcal{O}\left(\dfrac{1}{\ln n}\right)$ et que la série $\sum a_n$ converge. Étudier la nature de la série $\sum a_n^{b_n}$.

(École normale supérieure)

▷ **Solution.**

Pour la suite $(b_n)_{n\in\mathbb{N}}$ constante égale à 1, la série converge. On va montrer que c'est toujours le cas. Pour n assez grand, $a_n < 1$ et on majore alors $a_n^{b_n}$ en minorant b_n. Il existe $N \geqslant 0$ et $K > 0$ tel que pour tout $n \geqslant N$, $b_n \geqslant 1 - \dfrac{K}{\ln n}$ et dans ces conditions,

$$0 \leqslant a_n^{b_n} \leqslant a_n \exp\left(-K\frac{\ln a_n}{\ln n}\right).$$

On remarque que plus a_n est proche de zéro, plus la valeur de $\exp\left(-K\dfrac{\ln a_n}{\ln n}\right)$ est grande et inversement. Soit $M > 0$ à fixer plus tard. On a l'équivalence

$$\exp\left(-K\frac{\ln a_n}{\ln n}\right) \geqslant \exp M \iff -K\ln a_n \geqslant M\ln n \iff \ln a_n \leqslant -\frac{M}{K}\ln n,$$

ce qui donne

$$\exp\left(-K\frac{\ln a_n}{\ln n}\right) \geqslant \exp M \iff a_n \leqslant \frac{1}{n^{M/K}}.$$

Choisissons $M = 2K$. On peut alors majorer de la sorte

$$0 \leqslant a_n^{b_n} \leqslant \max\left(e^M a_n, \left(\frac{1}{n^2}\right)^{1-\frac{K}{\ln n}}\right) \leqslant e^M a_n + \left(\frac{1}{n^2}\right)^{1-\frac{K}{\ln n}}.$$

Or, $\left(\dfrac{1}{n^2}\right)^{1-\frac{K}{\ln n}} = \exp(-2\ln n + 2K) = \dfrac{e^{2K}}{n^2}$ est le terme général d'une série convergente. Le théorème de comparaison des séries à termes positifs permet de conclure : $\sum a_n^{b_n}$ converge. ◁

On rencontre parfois à l'oral le cas particulier $\sum a_n^{1-\frac{1}{n}}$.

3.5. Majoration à l'aide d'une intégrale

Soit $(u_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs. Pour $n \in \mathbb{N}$ on pose $S_n = \sum_{k=0}^{n} u_k$ et, en cas d'existence, $R_n = \sum_{k=n}^{+\infty} u_k$.

1. On suppose que $\sum u_n$ diverge. Prouver que

$$\sum \frac{u_n}{S_n^\alpha} \text{ converge} \iff \alpha > 1.$$

2. On suppose que $\sum u_n$ converge. Prouver que

$$\sum \frac{u_n}{R_n^\alpha} \text{ converge} \iff \alpha < 1.$$

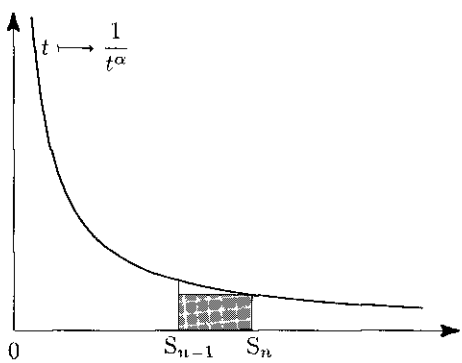
(École polytechnique)

Solution.

1. Comme la série $\sum u_n$ diverge, la suite (S_n) est croissante et tend vers $+\infty$ lorsque n tend vers l'infini.

• Prouvons pour commencer que la série $\dfrac{u_n}{S_n^\alpha}$ converge lorsque $\alpha > 1$. Pour cela, nous allons majorer $\dfrac{u_n}{S_n^\alpha}$ à l'aide d'une intégrale. Pour tout $n \geqslant 1$ on a,

$$\frac{u_n}{S_n^\alpha} = \frac{S_n - S_{n-1}}{S_n^\alpha} \leqslant \int_{S_{n-1}}^{S_n} \frac{dt}{t^\alpha}.$$



Comme $\alpha > 1$, la fonction $t \longmapsto \dfrac{1}{t^\alpha}$ est intégrable sur $[S_0, +\infty[$ et on a

$$\sum_{k=1}^{n} \frac{u_k}{S_k^\alpha} \leqslant \int_{S_0}^{S_n} \frac{dt}{t^\alpha} \leqslant \int_{S_0}^{+\infty} \frac{dt}{t^\alpha}.$$

La série étant à termes positifs, elle converge.

• On va maintenant montrer que la série diverge lorsque $\alpha \leqslant 1$. En fait, il suffit de traiter le cas $\alpha = 1$. En effet, pour n assez grand on a $S_n \geqslant 1$ et donc $\dfrac{u_n}{S_n^\alpha} \geqslant \dfrac{u_n}{S_n}$ pour $\alpha < 1$. Si on sait que la série $\sum \dfrac{u_n}{S_n}$ diverge, le théorème de comparaison des séries à termes positifs permet alors d'affirmer que la série $\sum \dfrac{u_n}{S_n^\alpha}$ diverge aussi. Pour prouver que $\sum \dfrac{u_n}{S_n}$ diverge, on va montrer qu'elle ne vérifie pas le critère de Cauchy. Soit N et p deux entiers naturels. Comme la suite $(S_n)_{n\in\mathbb{N}}$ est croissante, on peut écrire

$$\sum_{n=N}^{N+p} \frac{u_n}{S_n} \geqslant \sum_{n=N}^{N+p} \frac{u_n}{S_{N+p}} = \frac{S_{N+p} - S_{N-1}}{S_{N+p}} = 1 - \frac{S_{N-1}}{S_{N+p}} \xrightarrow[p\to\infty]{} 1.$$

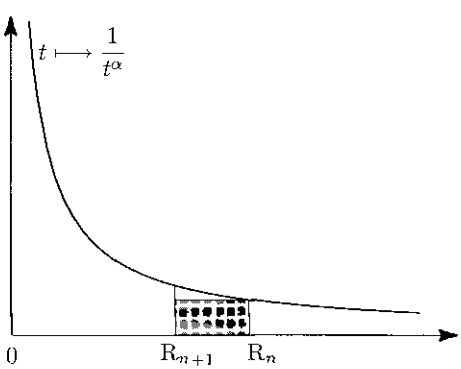
Ainsi, pour p assez grand, la tranche de Cauchy $\sum_{n=N}^{N+p} \dfrac{u_n}{S_n}$ est minorée par $\dfrac{1}{2}$. Le critère de Cauchy n'est pas vérifié et la série diverge donc.

Cette question montre donc que si $\sum u_n$ est une série divergente à termes positifs, on peut toujours trouver une suite positive v_n négligeable devant u_n, telle que la série $\sum v_n$ diverge encore : il suffit de prendre $v_n = \dfrac{u_n}{S_n^\alpha} = o(u_n)$ avec $0 < \alpha \leqslant 1$.

2. Remarquons que R_n décroît et tend vers 0 lorsque n tend vers l'infini.

• Supposons $\alpha < 1$. On utilise, comme dans la question précédente, une majoration à l'aide d'une intégrale. On peut supposer $\alpha \geqslant 0$ car si $\alpha < 0$ on a $\dfrac{u_n}{R_n^\alpha} = o(u_n)$ et le théorème de comparaison permet de conclure que la série $\sum \dfrac{u_n}{R_n^\alpha}$ converge. Soit n un entier naturel. La fonction $t \longmapsto \dfrac{1}{t^\alpha}$ étant décroissante, on a

$$\frac{u_n}{R_n^\alpha} = \frac{R_n - R_{n+1}}{R_n^\alpha} \leqslant \int_{R_{n+1}}^{R_n} \frac{dt}{t^\alpha}.$$



Il en résulte que

$$\sum_{k=0}^{n} \frac{u_k}{\mathrm{R}_k^\alpha} \leqslant \int_{\mathrm{R}_{n+1}}^{\mathrm{R}_0} \frac{\mathrm{d}t}{t^\alpha} \leqslant \int_0^{\mathrm{R}_0} \frac{\mathrm{d}t}{t^\alpha},$$

car la fonction $t \longmapsto \dfrac{1}{t^\alpha}$ est intégrable sur $]0, \mathrm{R}_0]$ puisque $\alpha < 1$. La série étant à termes positifs, elle converge.

Ainsi, si u_n est le terme général d'une série convergente, il existe une suite (v_n) de $\mathbb{R}_+$ devant laquelle (u_n) est négligeable et qui est le terme général d'une série convergente. Il suffit de prendre $v_n = \dfrac{u_n}{\mathrm{R}_n^\alpha}$ avec $0 < \alpha < 1$.

• Supposons maintenant $\alpha = 1$ et montrons que la série $\sum \dfrac{u_n}{\mathrm{R}_n}$ ne vérifie pas le critère de Cauchy. Soit N et p des entiers naturels. On minore la tranche de Cauchy entre N et $\mathrm{N}+p$ par :

$$\sum_{n=\mathrm{N}}^{\mathrm{N}+p} \frac{u_n}{\mathrm{R}_n} \geqslant \sum_{n=\mathrm{N}}^{\mathrm{N}+p} \frac{u_n}{\mathrm{R}_\mathrm{N}} = \frac{\mathrm{R}_\mathrm{N} - \mathrm{R}_{\mathrm{N}+p+1}}{\mathrm{R}_\mathrm{N}} = 1 - \frac{\mathrm{R}_{\mathrm{N}+p+1}}{\mathrm{R}_\mathrm{N}} \xrightarrow[p\to\infty]{} 1.$$

Ainsi, pour p assez grand, $\sum\limits_{n=\mathrm{N}}^{\mathrm{N}+p} \dfrac{u_n}{\mathrm{R}_n}$ est minorée par $\dfrac{1}{2}$. Cela prouve que la série diverge.

• Supposons pour terminer $\alpha > 1$. Pour n assez grand, $\mathrm{R}_n \leqslant 1$ et donc $\mathrm{R}_n^\alpha \leqslant \mathrm{R}_n$ et $\dfrac{u_n}{\mathrm{R}_n^\alpha} \geqslant \dfrac{u_n}{\mathrm{R}_n}$. Le théorème de comparaison des séries à termes positifs assure la divergence de $\sum \dfrac{u_n}{\mathrm{R}_n^\alpha}$. ◁

Dans l'exercice suivant, on se donne une suite $(a_n)_{n\in\mathbb{N}}$ de $\mathbb{R}_+$, croissante et tendant vers $+\infty$. Aussi lente que soit la croissance de a_n, on peut construire une suite positive $(u_n)_{n\in\mathbb{N}}$ avec u_n assez petit pour que la série $\sum u_n$ converge, mais pas trop pour que la série $\sum a_n u_n$ diverge.

3.6. Sur la négligeabilité

Soit $(a_n)_{n\in\mathbb{N}}$ une suite croissante de réels positifs qui tend vers l'infini. Construire une suite $(u_n)_{n\in\mathbb{N}}$ de réels positifs telle que $\sum u_n$ converge et $\sum a_n u_n$ diverge.

(École polytechnique)

Solution.

La suite $(a_n)_{n\in\mathbb{N}}$ est strictement positive à partir d'un certain rang

n_0. Comme le comportement d'une série ne dépend pas des premiers termes de la suite on peut, quitte à se placer au delà du rang n_0, supposer que $a_n > 0$ pour tout n.

Posons alors $u_n = \dfrac{1}{a_n} - \dfrac{1}{a_{n+1}}$ pour $n \geqslant 0$. La suite $(u_n)_{n\in\mathbb{N}^*}$ est à termes positifs puisque $(a_n)_{n\in\mathbb{N}}$ est croissante. On a $\sum_{k=0}^{n} u_k = \dfrac{1}{a_0} - \dfrac{1}{a_{n+1}} \xrightarrow[n\to\infty]{} \dfrac{1}{a_0}$ de sorte que $\sum u_n$ converge.

Par ailleurs, on a $a_n u_n = 1 - \dfrac{a_n}{a_{n+1}}$. Si $\dfrac{a_n}{a_{n+1}}$ ne tend pas vers 1, alors $\sum a_n u_n$ diverge. Si $\dfrac{a_n}{a_{n+1}}$ tend vers 1, alors

$$1 - \frac{a_n}{a_{n+1}} \underset{n\to\infty}{\sim} -\ln\left(\frac{a_n}{a_{n+1}}\right) = \ln a_{n+1} - \ln a_n.$$

La suite $(\ln a_{n+1} - \ln a_n)_{n\in\mathbb{N}}$ est à termes positifs et

$$\sum_{k=0}^{n} (\ln a_{k+1} - \ln a_k) = \ln a_{n+1} - \ln a_0 \xrightarrow[n\to\infty]{} +\infty.$$

Donc $\sum(\ln a_{n+1} - \ln a_n)$ diverge et $\sum a_n u_n$ également d'après le théorème de comparaison des séries à termes positifs. ◁

Si $\sum x_n$ est une série convergente à termes strictement positifs, et si on note $\mathrm{R}_n = \sum_{k=n}^{+\infty} x_k$, on voit grâce à l'exercice 3.5 que la série $\sum \dfrac{x_n}{\mathrm{R}_n}$ diverge. Or, la suite $a_n = \dfrac{1}{\mathrm{R}_n}$ tend vers $+\infty$ en croissant. On a donc divergence de la série $\sum a_n x_n$ et convergence de la série $\sum x_n$. Mais $x_n = \mathrm{R}_n - \mathrm{R}_{n+1} = \dfrac{1}{a_n} - \dfrac{1}{a_{n+1}}$. Cela éclaire la solution donnée ci-dessus.

Dans les deux exercices suivants, comme dans beaucoup d'autres, on utilise le fait que la suite $(a_n)_{n\in\mathbb{N}}$ et la série $\sum(a_{n+1} - a_n)$ sont de même nature, puisque pour $n \geqslant 1$, $\sum_{k=0}^{n-1} (a_{k+1} - a_k) = a_n - a_0$.

3.7. Série et suite

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs. On considère la suite $(\alpha_n)_{n\in\mathbb{N}}$ définie par le choix de α_0 et α_1 dans $]0, +\infty[$ et la relation de récurrence $\alpha_{n+1} = \alpha_n + a_n\alpha_{n-1}$ pour $n \geqslant 1$. Trouver une condition nécessaire et suffisante sur la suite $(a_n)_{n\in\mathbb{N}}$ pour que la suite $(\alpha_n)_{n\in\mathbb{N}}$ converge.

(École polytechnique)

▷ **Solution.**

Il est clair que la suite (α_n) est à termes strictement positifs et est croissante à partir du rang 1. Donc soit elle converge, soit elle diverge vers $+\infty$.

Supposons que la suite (α_n) converge et notons $\ell > 0$ sa limite. On obtient

$$a_n = \frac{\alpha_{n+1} - \alpha_n}{\alpha_{n-1}} \underset{n\to\infty}{\sim} \frac{1}{\ell}(\alpha_{n+1} - \alpha_n).$$

Comme la série $\sum(\alpha_{n+1} - \alpha_n)$ converge, on en déduit que la série $\sum a_n$ converge.

Montrons que cette condition est suffisante. On suppose donc que $\sum a_n$ converge. En particulier a_n tend vers 0 et

$$\alpha_{n+1} - \alpha_n = a_n\alpha_{n-1} = o(\alpha_{n-1}) = o(\alpha_n),$$

car $0 < \alpha_{n-1} < \alpha_n$. On en déduit que $\lim\limits_{n\to+\infty} \dfrac{\alpha_{n+1}}{\alpha_n} = 1$, puis que

$$\frac{\alpha_{n+1}}{\alpha_n} - 1 = a_n \frac{\alpha_{n-1}}{\alpha_n} \underset{n\to\infty}{\sim} a_n.$$

Mais on sait aussi, puisque $\dfrac{\alpha_{n+1}}{\alpha_n}$ tend vers 1, que

$$\frac{\alpha_{n+1}}{\alpha_n} - 1 \underset{n\to\infty}{\sim} \ln\left(\frac{\alpha_{n+1}}{\alpha_n}\right) = \ln\alpha_{n+1} - \ln\alpha_n.$$

On obtient donc finalement $a_n \underset{n\to\infty}{\sim} \ln\alpha_{n+1} - \ln\alpha_n$. Par le théorème de comparaison, la série $\sum(\ln\alpha_{n+1} - \ln\alpha_n)$ converge, ce qui équivaut à la convergence de la suite $(\ln\alpha_n)$. Par continuité de la fonction exponentielle, la suite (α_n) converge.

Conclusion. La suite (α_n) converge si et seulement si la série $\sum a_n$ converge. ◁

3.8. Un théorème de comparaison

Soit (a_n) et (b_n) deux suites réelles strictement positives. On suppose que $\sum b_n$ diverge, et que

$$\lim_{n\to+\infty} \frac{1}{b_n}\frac{a_n}{a_{n+1}} - \frac{1}{b_{n+1}} = \ell.$$

Montrer que si $\ell > 0$, $\sum a_n$ converge et que si $\ell < 0$, $\sum a_n$ diverge.

(École polytechnique)

▷ **Solution.**

Notons que, comme $\ell \neq 0$, l'hypothèse s'écrit $\dfrac{a_n}{b_n} - \dfrac{a_{n+1}}{b_{n+1}} \underset{n\to\infty}{\sim} \ell a_{n+1}$. On va raisonner par l'absurde dans les deux cas.

• Supposons pour commencer $\ell > 0$. Si la série $\sum a_n$ diverge, alors $\sum\left(\dfrac{a_n}{b_n} - \dfrac{a_{n+1}}{b_{n+1}}\right)$ diverge également. Cette série est à termes positifs, pour n assez grand, car $\ell a_{n+1} > 0$. On en déduit que

$$\lim_{n\to+\infty} \sum_{k=0}^{n} \left(\frac{a_k}{b_k} - \frac{a_{k+1}}{b_{k+1}}\right) = +\infty.$$

Mais ceci est impossible car $\displaystyle\sum_{k=0}^{n} \left(\frac{a_k}{b_k} - \frac{a_{k+1}}{b_{k+1}}\right) = \frac{a_0}{b_0} - \frac{a_{n+1}}{b_{n+1}} \leqslant \frac{a_0}{b_0}$. Cette contradiction prouve que la série $\sum a_n$ converge.

• Supposons $\ell < 0$. Si $\sum a_n$ converge, alors $\sum\left(\dfrac{a_n}{b_n} - \dfrac{a_{n+1}}{b_{n+1}}\right)$ converge également. Il existe un entier n_0 tel que $\dfrac{a_n}{b_n} - \dfrac{a_{n+1}}{b_{n+1}} < 0$ pour $n \geqslant n_0$. On a alors $\dfrac{a_{n_0}}{b_{n_0}} - \dfrac{a_n}{b_n} = \displaystyle\sum_{k=n_0}^{n-1} \left(\frac{a_k}{b_k} - \frac{a_{k+1}}{b_{k+1}}\right)$ qui a pour limite $\mathrm{S} = \displaystyle\sum_{k=n_0}^{+\infty} \left(\frac{a_k}{b_k} - \frac{a_{k+1}}{b_{k+1}}\right) < 0$. On en déduit que $\lim\limits_{n\to+\infty} \dfrac{a_n}{b_n} = \dfrac{a_{n_0}}{b_{n_0}} - \mathrm{S} = \mathrm{K} > 0$, c'est-à-dire que $a_n \underset{n\to\infty}{\sim} \mathrm{K} b_n$, ce qui est impossible car $\sum a_n$ converge et $\sum b_n$ diverge. Donc la série $\sum a_n$ diverge. ◁

Dans le cas où $\ell = 0$ on ne peut rien dire : si on prend pour (b_n) la suite constante égale à 1, l'hypothèse est que $\dfrac{a_n}{a_{n+1}}$ tend vers 1 et on sait que cela ne suffit pas pour connaître la nature de $\sum a_n$ (considérer les séries de Riemann).

Le regroupement de termes consécutifs d'une série $\sum u_n$ conduit à la notion de tranche de Cauchy, déjà utilisée dans l'exercice 3.5. Il s'agit des sommes finies de la forme $u_{n+1} + u_{n+2} + \cdots + u_{n+p} = \mathrm{U}_{n+p} - \mathrm{U}_n$ où l'on note $\mathrm{U}_n = \displaystyle\sum_{k=0}^{n} u_k$ la suite des sommes partielles. Lorsque la série réelle ou complexe $\sum u_n$ converge, pour tout $\varepsilon > 0$, il existe un rang N tel que pour tout $n \geqslant \mathrm{N}$ et tout $p \geqslant 0$, $\left|\displaystyle\sum_{k=n+1}^{n+p} u_k\right| \leqslant \varepsilon$. Ce résultat demeure pour une série $\sum u_n$ d'un espace vectoriel normé (le module, ou la valeur absolue, étant remplacé par la norme). La réciproque est vraie lorsque l'espace est complet (ce qui est bien entendu le cas de $\mathbb{R}$ et $\mathbb{C}$) : c'est le critère de convergence de Cauchy. L'exercice suivant utilise cette notion, toujours dans le cadre des séries réelles positives.

3.9. Étude de convergence

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs qui décroît vers 0. Étudier la nature de la série $\sum \dfrac{a_n - a_{n+1}}{a_n}$.

(**École polytechnique**)

▷ **Solution.**

Pour avoir une idée du résultat, on commence par regarder un ou deux exemples. On a $\dfrac{a_n - a_{n+1}}{a_n} = 1 - \dfrac{a_{n+1}}{a_n}$. Pour avoir convergence il est donc nécessaire que $a_n \sim a_{n+1}$, c'est-à-dire que la suite $(a_n)_{n\in\mathbb{N}}$ ne tende pas trop vite vers 0. Si on prend par exemple $a_n = \dfrac{1}{n}$, on obtient $\dfrac{a_n - a_{n+1}}{a_n} = \dfrac{1}{n+1}$ qui conduit à une série divergente. On obtient un résultat identique avec d'autres exemples. On va donc essayer de montrer qu'il y a toujours divergence, et pour cela nous allons minorer une tranche de Cauchy de la série. Soit $\mathrm{N} \geqslant 1$ et $p \geqslant 0$. Puisque la suite $(a_n)_{n\in\mathbb{N}}$ est décroissante, on a

$$\sum_{n=\mathbb{N}}^{\mathrm{N}+p} \frac{a_n - a_{n+1}}{a_n} \geqslant \sum_{n=\mathbb{N}}^{\mathrm{N}+p} \frac{a_n - a_{n+1}}{a_{\mathrm{N}}} = \frac{a_{\mathrm{N}} - a_{\mathrm{N}+p+1}}{a_{\mathrm{N}}}$$
$$\geqslant 1 - \frac{a_{\mathrm{N}+p+1}}{a_{\mathrm{N}}} \xrightarrow[p\to\infty]{} 1.$$

Donc pour p assez grand, la somme $\displaystyle\sum_{n=\mathrm{N}}^{\mathrm{N}+p} \frac{a_n - a_{n+1}}{a_n}$ est supérieure à $\dfrac{1}{2}$; ainsi, le critère de Cauchy n'est pas vérifié et la série $\sum \dfrac{a_n - a_{n+1}}{a_n}$ diverge. ◁

Le lecteur aura remarqué que cet exercice est une autre formulation d'une partie de l'exercice 3.5. En effet, si on pose $u_n = a_n - a_{n+1}$ la série $\sum u_n$ converge et son reste R_n d'ordre n est égal à a_n de sorte que $\dfrac{a_n - a_{n+1}}{a_n} = \dfrac{u_n}{\mathrm{R}_n}$.

Le résultat de l'exercice suivant est classique et important.

3.10. Condition nécessaire de convergence

Soit $(u_n)_{n\in\mathbb{N}}$ une suite réelle décroissante de limite nulle. Montrer que si la série $\sum u_n$ converge, alors la suite $(nu_n)_{n\in\mathbb{N}}$ tend vers 0.

(**École polytechnique**)

▷ **Solution.**

L'idée consiste à majorer la quantité nu_n par une tranche de Cauchy contenant n termes consécutifs en utilisant la décroissance de la suite $(u_n)_{n\in\mathbb{N}}$. La majoration directe $nu_n \leqslant u_1 + \cdots + u_n$ est toutefois trop grossière pour conclure. On va commencer par la suite extraite $(2nu_{2n})$ afin d'avoir une tranche de Cauchy qui tend vers 0. On a, pour tout $n \geqslant 0$,

$$2nu_{2n} \leqslant 2(u_{2n} + u_{2n-1} + \cdots + u_{n+1}) = 2(S_{2n} - S_n),$$

avec $S_n = \sum_{k=0}^{n} u_k$. Puisque la série $\sum u_n$ converge, la suite $S_{2n} - S_n$ tend vers 0. Il s'ensuit que $2nu_{2n}$ converge vers 0. D'autre part, on peut écrire

$$0 \leqslant (2n+1)u_{2n+1} = 2nu_{2n+1} + u_{2n+1} \leqslant 2nu_{2n} + u_{2n+1} \xrightarrow[n\to\infty]{} 0.$$

On a donc $\lim_{n\to+\infty} 2nu_{2n} = \lim_{n\to+\infty} (2n+1)u_{2n+1} = 0$ et par conséquent la suite $(nu_n)_{n\in\mathbb{N}}$ tend vers 0. ◁

L'hypothèse de décroissance est bien entendu essentielle : $\sum u_n$ peut converger sans que u_n soit négligeable devant $\frac{1}{n}$. Par exemple, il suffit de poser $u_n = \frac{1}{n}$ si n est une puissance de 10 et 0 sinon. Par ailleurs, il est bien connu que si $(u_n)_{n\in\mathbb{N}}$ vérifie les conditions de l'énoncé, l'hypothèse $u_n = o\left(\frac{1}{n}\right)$ n'est pas une condition suffisante pour avoir convergence de la série $\sum u_n$: on peut considérer par exemple $u_n = \frac{1}{n \ln n}$ pour $n \geqslant 2$.

Dans l'étude de la nature d'une série $\sum u_n$, on peut être amené à regrouper des termes afin de constituer une nouvelle série. On se donne une suite $(k_n)_{n\geqslant 0}$ strictement croissante de $\mathbb{N}$ avec $k_0 = 0$, et on considère $p_n = \sum_{k=k_n}^{k_{n+1}-1} u_k$ pour $n \geqslant 0$ (ce sont les «paquets»). Les sommes partielles de la série $\sum p_n$ forment une suite extraite de la suite des sommes partielles de $\sum u_n$. Donc si la série $\sum u_n$ converge il en est de même de la série $\sum p_n$ et les deux séries ont la même somme. Bien entendu, la réciproque n'est pas toujours vraie. C'est toutefois le cas lorsque $\sum u_n$ est une série à termes positifs, car dans ce cas la suite des sommes partielles est croissante (on peut aussi voir cela comme un cas particulier du théorème d'associativité des familles sommables.) Dans le cas où la série n'est pas à termes positifs, on peut donner diverses conditions supplémentaires qui permettent d'avoir la réciproque. Nous invitons le lecteur à se reporter à l'exercice 3.34 où il trouvera un exemple très intéressant de sommation par paquets.

Les exercices qui suivent sont des applications dans le cadre de séries positives.

3.11. Entiers qui s'écrivent sans le chiffre 9

Déterminer la nature de la série $\sum \frac{1}{p_n}$ où p_n est le n-ième entier naturel non nul dont l'écriture décimale ne comporte pas de 9.

(**École polytechnique**)

▷ **Solution.**

Les entiers naturels (0 compris) inférieurs strictement à 10^n s'écrivent avec n chiffres en notation décimale. Ceux qui ne comportent pas de 9 s'écrivent donc avec n chiffres parmi les entiers de 0 à 8. On a 9^n possibilités pour les choix des n chiffres. Il y a donc 9^n entiers naturels (0 compris) strictement inférieur à 10^n sans 9 dans leur écriture décimale. Le nombre 10^n qui s'écrit avec 1 suivi de n zéros est donc le 9^n-ième entier naturel non nul s'écrivant sans 9 : $p_{9^n} = 10^n$.

La suite $(p_n)_{n\geqslant 1}$ est croissante. Pour $n \geqslant 0$, on a la majoration suivante :

$$S_n = \sum_{k=9^n}^{9^{n+1}-1} \frac{1}{p_k} \leqslant \sum_{k=9^n}^{9^{n+1}-1} \frac{1}{p_{9^n}} = \sum_{k=9^n}^{9^{n+1}-1} \frac{1}{10^n} \leqslant \frac{9^{n+1}}{10^n} = 9\left(\frac{9}{10}\right)^n.$$

Comme la série géométrique $\sum 9\left(\frac{9}{10}\right)^n$ converge, le théorème de comparaison des séries à termes positifs assure la convergence de la série $\sum S_n$. Le théorème de sommation par paquets permet de conclure à la convergence de la série à termes positifs $\sum \frac{1}{p_n}$. ◁

Le résultat de l'exercice suivant est une généralisation d'une propriété utilisée par Cauchy.

3.12. Critère de condensation de Cauchy

Soit $(u_n)_{n\geqslant 1}$ une suite réelle décroissante, qui converge vers 0.

1. Soit $p \geqslant 2$ un entier. Établir l'équivalence :

$$\sum u_n \text{ converge} \iff \sum p^n u_{p^n} \text{ converge.}$$

C'est le critère de condensation de Cauchy.

2. Montrer que si $\sum u_n$ diverge, alors $\sum \min(u_n, 1/n)$ diverge aussi.

(École polytechnique)

▷ **Solution.**

1. Dans la série $\sum p^n u_{p^n}$, on ne prend que les termes u_k dont l'indice k appartient à la suite géométrique (p^n) mais chaque terme u_{p^n} est affecté d'un poids de p^n (ce qui est l'ordre de grandeur de $p^{n+1} - p^n$). La monotonie de la suite va permettre de montrer que les deux séries sont de même nature. On va procéder en regroupant les termes u_k pour k compris entre p^n et p^{n+1}.

Pour $p^n \leqslant k < p^{n+1}$, on pose $v_k = u_{p^{n+1}}$ et $w_k = u_{p^n}$. On a alors l'inégalité $v_k \leqslant u_k \leqslant w_k$ et donc,

$$(p^{n+1} - p^n)u_{p^{n+1}} = \sum_{k=p^n}^{p^{n+1}-1} v_k \leqslant \sum_{k=p^n}^{p^{n+1}-1} u_k \leqslant \sum_{k=p^n}^{p^{n+1}-1} w_k = (p^{n+1} - p^n)u_{p^n},$$

ou encore

$$\left(1 - \frac{1}{p}\right) p^{n+1} u_{p^{n+1}} \leqslant \sum_{k=p^n}^{p^{n+1}-1} u_k \leqslant (p-1)p^n u_{p^n}.$$

D'après le théorème de comparaison des séries à termes positifs, la série de terme général $p^n u_{p^n}$ converge si, et seulement si, celle de terme général $\sum_{k=p^n}^{p^{n+1}-1} u_k$ converge. Le théorème de sommation par paquets assure alors que les séries à termes positifs $\sum p^n u_{p^n}$ et $\sum u_n$ sont de même nature.

Dans le cas où $u_n = \dfrac{1}{n}$ pour $n \geqslant 1$ et $p = 2$, on a $2^n u_{2^n} = 1$ ce qui n'est pas le terme général d'une série convergente : c'est ainsi que Cauchy prouva la divergence de la série harmonique $\sum \dfrac{1}{n}$. On peut également prouver la convergence des séries de Riemann $\sum \dfrac{1}{n^\alpha}$ lorsque $\alpha > 1$. En effet, toujours en prenant $p = 2$, on a $2^n \dfrac{1}{(2^n)^\alpha} = \left(\dfrac{1}{2^{\alpha-1}}\right)^n$ qui est bien le terme général d'une série convergente (puisque $\dfrac{1}{2^{\alpha-1}} < 1$). Remarquons enfin que le critère de condensation donne une solution immédiate de l'exercice 3.11 en prenant $p = 9$ puisque $p_{9^n} = 10^n$.

2. Posons $v_n = \min(u_n, 1/n)$. On observe qu'il s'agit d'une suite décroissante de limite nulle. On peut donc utiliser le résultat de la question précédente. Raisonnons par l'absurde et supposons que $\sum v_n$

converge. En prenant $p = 2$ dans le critère de condensation, la série $\sum 2^n v_{2^n}$ converge. Or, pour $n \geqslant 0$, on peut écrire

$$2^n v_{2^n} = 2^n \min\left(u_{2^n}, \frac{1}{2^n}\right) = \min(2^n u_{2^n}, 1)$$

Comme le terme général d'une série convergente tend vers 0, on doit avoir $2^n u_{2^n} \leqslant 1$ à partir d'un certain rang, soit $\min(2^n u_{2^n}, 1) = 2^n u_{2^n}$. Par conséquent, $\sum 2^n u_{2^n}$ converge. Le critère de condensation appliqué à la série $\sum u_n$ assure alors la convergence de $\sum u_n$, ce qui est contraire à l'hypothèse. ◁

3.13. Séries définies à l'aide d'une permutation de $\mathbb{N}^*$

Soit σ une permutation de $\mathbb{N}^*$.

1. Quelle est la nature de la série $\sum \dfrac{\sigma(n)}{n^2}$?

2. Même question avec $\sum \dfrac{\sigma(n)}{n^2 \ln n}$.

3. Même question avec $\sum \dfrac{\sigma(n)}{n^3}$.

(École polytechnique)

Solution.

1. En prenant le cas particulier où σ est l'identité, on obtient la série harmonique qui est divergente. On va montrer qu'il y a toujours divergence. Pour montrer cela, nous allons minorer une tranche de Cauchy. Soit $N \geqslant 1$. Si on considère $2N$ entiers naturels non nuls deux à deux distincts, N d'entre eux nécessairement sont supérieurs ou égaux à N. C'est le cas notamment pour la famille des $\sigma(n)$ lorsque n décrit les entiers compris entre $N + 1$ et $3N$. Par conséquent, on peut minorer la tranche de Cauchy entre $N + 1$ et $3N$ par :

$$\sum_{n=N+1}^{3N} \frac{\sigma(n)}{n^2} \geqslant \sum_{n=N+1}^{3N} \frac{\sigma(n)}{(3N)^2} \geqslant N \cdot \frac{N}{9N^2} = \frac{1}{9}.$$

Le critère de Cauchy n'étant pas vérifié, la série $\sum \dfrac{\sigma(n)}{n^2}$ diverge.

2. Il est également légitime de penser que cette seconde série est toujours divergente puisque si σ est l'identité, on obtient la série de Bertrand $\sum \dfrac{1}{n \ln n}$ qui diverge. Comme la série $\sum \dfrac{\sigma(n)}{n^2 \ln n}$ est à termes positifs, elle est de même nature que la série de terme général $a_N = \sum_{n=2^N+1}^{2^{N+1}} \dfrac{\sigma(n)}{n^2 \ln n}$, d'après le théorème de sommations par paquets. Les

$\sigma(n)$, pour n compris entre $2^N + 1$ et 2^{N+1}, constituent un ensemble de 2^N entiers naturels non nuls distincts deux à deux. Aussi 2^{N-1} d'entre eux, au moins, sont supérieurs où égaux à 2^{N-1}. On minore a_N en ne considérant que ces derniers (les autres étant positifs),

$$a_N = \sum_{n=2^N+1}^{2^{N+1}} \frac{\sigma(n)}{n^2 \ln n} \geqslant \sum_{n=2^N+1}^{2^{N+1}} \frac{\sigma(n)}{(2^{N+1})^2 \ln 2^{N+1}} \geqslant 2^{N-1} \frac{2^{N-1}}{2^{2N+2}(N+1)\ln 2}$$

et finalement, $a_N \geqslant \dfrac{1}{16(N+1)\ln 2}$ qui est le terme général d'une série divergente. Aussi, la série $\sum a_N$ diverge et $\sum \dfrac{\sigma(n)}{n^2 \ln n}$ de même.

3. La troisième série peut en revanche converger, ce qui est le cas si par exemple σ est l'identité. Mais ce n'est pas général, car pour d'autres permutations on peut avoir une série divergente. Il suffit par exemple de trouver une permutation σ telle que $\sigma(n) = n^3$ pour une infinité de n. Cherchons à expliciter un tel exemple. Posons $A = \{(2k)^3,\ k \in \mathbb{N}^*\}$ et $B = \mathbb{N}^* \setminus A$. Les deux ensembles A et B sont infinis. On notera $\alpha_1 < \alpha_2 < \cdots$ la suite croissante des éléments de B. On considère σ la permutation de $\mathbb{N}^*$ définie par $\sigma(2k) = (2k)^3$ pour tout k et $\sigma(2k+1) = \alpha_{k+1}$ pour tout $k \geqslant 0$. La suite $\dfrac{\sigma(n)}{n^3}$ ne tend pas vers 0 donc la série $\sum \dfrac{\sigma(n)}{n^3}$ diverge.

On aurait pu aussi prendre pour σ la permutation qui échange p et p^3 pour tout nombre premier p et qui laisse fixes les entiers autres que les nombres premiers et leurs cubes. ◁

Le lecteur constatera aisément que l'argument utilisé en 3 se généralise et donc, pour tout exposant m, il existe des permutations σ qui rendent la série $\sum \dfrac{\sigma(n)}{n^m}$ divergente. On peut donner une autre solution de la question 1 en utilisant l'inégalité du réordonnement établie dans l'exercice 1.15 du chapitre 1. En effet, soit N un entier fixé. Notons $\alpha_1 < \alpha_2 < \cdots < \alpha_N$ les entiers de l'ensemble $\{\sigma(1), \ldots, \sigma(N)\}$ rangés dans l'ordre croissant. L'inégalité du réordonnement assure que

$$\sum_{k=1}^{N} \frac{\sigma(k)}{k^2} \geqslant \sum_{k=1}^{N} \frac{\alpha_k}{k^2} \geqslant \sum_{k=1}^{N} \frac{1}{k},$$

car on a clairement $\alpha_k \geqslant k$ pour tout k. On conclut en invoquant la divergence de la série harmonique.

Lorsque les deux suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ sont positives et équivalentes au voisinage de $+\infty$, on peut «sommer» les équivalents dans les deux cas suivants :

- *si* $\sum u_n$ *converge, les restes sont équivalents :* $\sum_{k=n+1}^{+\infty} u_k \sim \sum_{k=n+1}^{+\infty} v_k$;
- *si* $\sum u_n$ *diverge, ce sont les sommes partielles qui sont équivalentes :* $\sum_{k=0}^{n} u_k \sim \sum_{k=0}^{n} v_k$.

En considérant la différence $u_n - v_n = o(u_n)$, *on a l'énoncé suivant : si* $(u_n)_{n\in\mathbb{N}}$ *est une suite réelle positive et* $(w_n)_{n\in\mathbb{N}}$ *une suite telle que* $w_n = o(u_n)$, *alors le reste* $\sum_{k=n+1}^{+\infty} w_k$ *est négligeable devant* $\sum_{k=n+1}^{+\infty} u_k$ *dans le cas où* $\sum u_n$ *converge. Si cette série diverge, c'est la somme partielle* $\sum_{k=0}^{n} w_k$ *qui est négligeable devant* $\sum_{k=0}^{n} u_k$.

Le théorème de Cesàro pour les suites réelles n'est qu'un cas particulier de ce résultat.

3.14. Recherche d'un équivalent (1)

On considère une suite $(a_n)_{n\in\mathbb{N}}$ de nombres réels telle que $\lim_{n\to+\infty} a_n \sum_{k=0}^{n} a_k^2 = 1$. Déterminer un équivalent de a_n.

(École polytechnique)

Solution.

Posons $S_n = \sum_{k=0}^{n} a_k^2$. Si la série de terme général a_n^2 converge, S_n tend vers une limite finie lorsque n tend vers $+\infty$ et $\lim_{n\to+\infty} a_n = 0$. On ne peut avoir $\lim_{n\to+\infty} a_n S_n = 1$. Comme la série de terme général $a_n^2 \geqslant 0$ ne converge pas, $\lim_{n\to+\infty} S_n = +\infty$.

Comme pour tout $n \geqslant 0$, $a_n^2 = S_n - S_{n-1}$, l'hypothèse s'écrit

$$S_n - S_{n-1} \underset{n\to\infty}{\sim} \frac{1}{S_n^2} \quad (S_n > 0 \text{ pour } n \text{ assez grand}).$$

On constate alors que

$$\begin{aligned} S_n^3 - S_{n-1}^3 &= (S_n - S_{n-1})(S_n^2 + S_n S_{n-1} + S_{n-1}^2) \\ &\underset{n\to\infty}{\sim} \frac{1}{S_n^2}(S_n^2 + S_n S_{n-1} + S_{n-1}^2) \\ &= \left(1 + \frac{S_{n-1}}{S_n} + \left(\frac{S_{n-1}}{S_n}\right)^2\right) \xrightarrow[n\to+\infty]{} 3, \end{aligned}$$

car $S_{n-1} \sim S_n$ lorsque n tend vers l'infini. En effet, on a

$$1-\frac{S_{n-1}}{S_n} \underset{n\to\infty}{\sim} \frac{1}{S_n^3} \xrightarrow[n\to+\infty]{} 0 \quad \text{puisque} \lim_{n\to+\infty} S_n = +\infty.$$

Le théorème de Cesàro conduit à

$$S_n^3 = \sum_{k=1}^{n} (S_k^3 - S_{k-1}^3) \underset{n\to\infty}{\sim} 3n.$$

D'où l'on déduit $S_n \underset{n\to\infty}{\sim} (3n)^{1/3}$ et finalement

$$\boxed{a_n \underset{n\to\infty}{\sim} \frac{1}{(3n)^{1/3}}.} \quad \triangleleft$$

3.15. Recherche d'un équivalent (2)

Soit (x_n) une suite réelle définie par $x_1 = a > 0$ et la relation de récurrence $x_{n+1} = x_n + \dfrac{1}{S_n}$ où $S_n = x_1 + x_2 + \cdots + x_n$. Déterminer un équivalent de x_n.

(École polytechnique)

▷ **Solution.**

La suite (x_n) est strictement croissante. Supposons qu'elle converge et notons α sa limite. Le théorème de Cesàro permet d'affirmer que $S_n \sim n\alpha$. Mais alors $x_{n+1} - x_n \sim \dfrac{1}{n\alpha}$ et la série $\sum(x_{n+1} - x_n)$ diverge, ce qui contredit la convergence de (x_n). Donc la suite (x_n) tend vers $+\infty$ et il en est de même de la suite (S_n).

On a $x_{n+1} \underset{n\to\infty}{\sim} x_n$ et donc $x_{n+1}^2 - x_n^2 = \dfrac{x_n + x_{n+1}}{S_n} \underset{n\to\infty}{\sim} \dfrac{2x_n}{S_n}$. Prenons l'inverse qui se manipule mieux. En effet, on observe qu'on a

$$\frac{S_{n+1}}{x_{n+1}} = \frac{S_n + x_{n+1}}{x_{n+1}} = \frac{S_n}{x_{n+1}} + 1$$

et donc

$$\frac{S_{n+1}}{x_{n+1}} - \frac{S_n}{x_n} = 1 + \frac{S_n(x_n - x_{n+1})}{x_n x_{n+1}} = 1 - \frac{1}{x_n x_{n+1}} \xrightarrow[n\to+\infty]{} 1.$$

Le théorème de Cesàro implique $\dfrac{S_n}{x_n} \underset{n\to\infty}{\sim} n$ et donc $x_{n+1}^2 - x_n^2 \sim \dfrac{2}{n}$. La somme partielle de la série harmonique étant équivalente à $\ln n$, une application du théorème de sommation des équivalents conduit au résultat.

$$\boxed{x_n \sim \sqrt{2\ln n}}. \quad \lhd$$

L'énoncé suivant porte sur le même thème mais il est plus théorique.

3.16. Sommation d'équivalents

1. Soit $(u_n)_{n\in\mathbb{N}}$ une suite de réels strictement positifs. On pose $S_n = \sum_{k=0}^{n} u_k$ et on suppose que $\dfrac{S_n}{nu_n}$ tend vers $\alpha > 0$. Étudier la convergence et la limite de $\dfrac{1}{n^2 u_n} \sum_{k=0}^{n} k u_k$.

2. Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites de réels strictement positifs. $S_n = \sum_{k=0}^{n} u_k$ et $T_n = \sum_{k=0}^{n} v_k$ pour $n \geqslant 0$. On suppose que $\dfrac{S_n}{nu_n}$ (resp. $\dfrac{T_n}{nv_n}$) tend vers une limite $\alpha > 0$ (resp. $\beta > 0$). Étudier la convergence et la limite de $\dfrac{1}{n^2 u_n v_n} \sum_{k=0}^{n} k u_k v_k$.

(École polytechnique)

Solution.

1. Notons que $\sum u_n$ diverge. Dans le cas contraire, on aurait, en notant S la somme de la série, $u_n \underset{n\to\infty}{\sim} \dfrac{S}{\alpha n}$, ce qui est contradictoire avec la convergence de $\sum u_n$, car la série harmonique diverge. La suite (S_n) tend donc vers l'infini. Puisque $\sum S_n$ diverge, que $\alpha n u_n \underset{n\to\infty}{\sim} S_n$ et que les séries en présence sont à termes positifs, le théorème de sommation des relations de comparaison permet d'affirmer que $\alpha \sum_{k=0}^{n} k u_k \underset{n\to\infty}{\sim} \sum_{k=0}^{n} S_k$. On remarque alors que

$$\begin{aligned}\sum_{k=0}^{n} S_k &= \sum_{k=0}^{n} \sum_{i=0}^{k} u_i \\ &= \sum_{i=0}^{n} \sum_{k=i}^{n} u_i \quad \text{(par interversion de sommations.)} \\ &= \sum_{i=0}^{n} (n+1-i) u_i = (n+1) S_n - \sum_{i=0}^{n} i u_i.\end{aligned}$$

Compte tenu de la relation précédente, on obtient

$$\frac{(n+1)S_n}{\sum_{k=0}^{n} ku_k} = 1 + \frac{\sum_{k=0}^{n} S_k}{\sum_{k=0}^{n} ku_k} \xrightarrow[n\to+\infty]{} \alpha + 1,$$

et donc $\sum_{k=0}^{n} ku_k \sim \frac{(n+1)S_n}{\alpha+1} \sim \frac{nS_n}{\alpha+1} \sim \frac{\alpha n^2 u_n}{\alpha+1}$ pour n tendant vers $+\infty$. On en déduit que

$$\boxed{\lim_{n\to+\infty} \frac{1}{n^2 u_n} \sum_{k=0}^{n} ku_k = \frac{\alpha}{\alpha+1}.}$$

2. Cette seconde question est une généralisation de la précédente que l'on retrouve en prenant $v_n = 1$. On va utiliser la même méthode que précédemment. On a $\alpha n u_n v_n \underset{n\to\infty}{\sim} S_n v_n$, par hypothèse. La série $\sum v_n$ diverge ; la suite (S_n) tendant vers $+\infty$, il en est de même, *a fortiori*, de la série $\sum S_n v_n$. Le théorème de sommation entraîne $\alpha \sum_{k=0}^{n} ku_k v_k \underset{n\to\infty}{\sim} \sum_{k=0}^{n} S_k v_k$. De la même manière, on obtient $\beta \sum_{k=0}^{n} ku_k v_k \underset{n\to\infty}{\sim} \sum_{k=0}^{n} T_k u_k$. On a donc

$$\frac{\sum_{k=0}^{n} S_k v_k + \sum_{k=0}^{n} T_k u_k}{\sum_{k=0}^{n} ku_k v_k} \xrightarrow[n\to+\infty]{} \alpha + \beta$$

et

$$(\alpha+\beta) \sum_{k=0}^{n} ku_k v_k \underset{n\to\infty}{\sim} \sum_{k=0}^{n} S_k v_k + \sum_{k=0}^{n} T_k u_k.$$

On observe alors que

$$\begin{aligned} \sum_{k=0}^{n} S_k v_k + \sum_{k=0}^{n} T_k u_k &= \sum_{0\leqslant l\leqslant k\leqslant n} u_l v_k + \sum_{0\leqslant l\leqslant k\leqslant n} v_l u_k \\ &= \left(\sum_{0\leqslant k,l\leqslant n} u_k v_l \right) + u_1 v_1 + \cdots + u_n v_n, \\ &= S_n T_n + \sum_{k=0}^{n} u_k v_k. \end{aligned}$$

La série $\sum ku_k v_k$ étant divergente, et pour n tendant vers l'infini, $u_n v_n$ étant négligeable devant $nu_n v_n$, le théorème de sommation des

équivalents entraîne que $\sum_{k=0}^{n} u_k v_k$ est négligeable devant $\sum_{k=0}^{n} k u_k v_k$. Sachant que $(\alpha+\beta)\sum_{k=0}^{n} k u_k v_k \underset{n\to\infty}{\sim} S_n T_n + \sum_{k=0}^{n} u_k v_k$, on en déduit que

$$(\alpha+\beta)\sum_{k=0}^{n} k u_k v_k \underset{n\to\infty}{\sim} S_n T_n \underset{n\to\infty}{\sim} \alpha\beta n^2 u_n v_n,$$

c'est-à-dire que

$$\boxed{\lim_{n\to+\infty} \frac{1}{n^2 u_n v_n} \sum_{k=0}^{n} k u_k v_k = \frac{\alpha\beta}{\alpha+\beta}.} \quad \triangleleft$$

La comparaison série-intégrale est un outil essentiel d'étude des séries à termes positifs pour leur convergence, l'estimation des restes, le comportement asymptotique des sommes partielles... Les exercices suivants en sont une illustration.

3.17. Estimation d'une somme partielle

Déterminer la partie entière de $\displaystyle\sum_{n=1}^{10^9} \frac{1}{n^{2/3}}$.

(École polytechnique)

Solution.

La fonction $t \longmapsto \dfrac{1}{t^{2/3}}$ est positive, continue et strictement décroissante sur $\mathbb{R}_+^*$. Pour $n \geqslant 2$, on a donc l'encadrement

$$\int_n^{n+1} \frac{\mathrm{d}t}{t^{2/3}} < \frac{1}{n^{2/3}} < \int_{n-1}^{n} \frac{\mathrm{d}t}{t^{2/3}}.$$

En sommant de 2 à 10^9, on obtient

$$\int_2^{10^9+1} \frac{\mathrm{d}t}{t^{2/3}} < \sum_{n=2}^{10^9} \frac{1}{n^{2/3}} < \int_1^{10^9} \frac{\mathrm{d}t}{t^{2/3}}.$$

On se contente d'encadrer la somme à partir de $n=2$ car le premier terme est entier égal à 1. Après calcul de l'intégrale, on obtient

$$3(\sqrt[3]{10^9+1} - \sqrt[3]{2}) < \sum_{n=2}^{10^9} \frac{1}{n^{2/3}} < 3(\sqrt[3]{10^9} - 1) = 2997.$$

On constate que le minorant $3(\sqrt[3]{10^9+1} - \sqrt[3]{2})$ est plus grand que $3(\sqrt[3]{10^9} - \sqrt[3]{2}) = 3(1000 - \sqrt[3]{2}) \geqslant 2996$. En rajoutant le premier terme, on a

$$2997 < \sum_{n=1}^{10^9} \frac{1}{n^{2/3}} < 2998,$$

et la partie entière cherchée est 2997. ◁

On peut noter que le calcul d'une valeur approchée de cette somme dépasse les capacités de Maple.

La divergence de la série harmonique $\sum \frac{1}{n}$ est connue au moins depuis Johann Bernoulli. On peut résumer ainsi son argumentation : si on suppose que cette série converge, on peut écrire[3]

$$\begin{aligned} S &= \sum_{n=1}^{+\infty} \frac{1}{n} = 1 + \sum_{n=2}^{+\infty} \frac{n-1}{(n-1)n} = 1 + \sum_{n=1}^{+\infty} \frac{n}{n(n+1)} \\ &= 1 + \sum_{n=1}^{+\infty} \sum_{k=1}^{n} \frac{1}{n(n+1)} \\ &= 1 + \sum_{k=1}^{+\infty} \sum_{n=k}^{+\infty} \frac{1}{n(n+1)} \quad \textit{(par interversion des sommations)} \\ &= 1 + \sum_{k=1}^{+\infty} \underbrace{\sum_{n=k}^{+\infty} \left(\frac{1}{n} - \frac{1}{n+1} \right)}_{=1/k} = 1 + \sum_{k=1}^{+\infty} \frac{1}{k} = 1 + S, \end{aligned}$$

ce qui est manifestement absurde.

L'exercice suivant propose d'établir un développement asymptotique de la somme partielle de la série harmonique et présente à cette occasion la constante d'Euler.

3.18. Développement asymptotique de la série harmonique

On pose $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ pour $n \geqslant 1$.

1. Soit $u_n = H_n - \ln n$ et $v_n = u_n - \frac{1}{n}$. Démontrer que ces suites sont adjacentes et convergent vers une constante réelle strictement positive γ.

2. Déterminer un développement asymptotique de H_n comprenant quatre termes.

3. Avec le formalisme moderne.

3. On pose $k_n = \min\{k \in \mathbb{N},\ \mathrm{H}_k \geqslant n\}$. Déterminer $\lim\limits_{n\to+\infty} \dfrac{k_{n+1}}{k_n}$.
(École polytechnique)

Solution.

1. La différence $u_n - v_n = \dfrac{1}{n}$ est positive et converge vers 0. La suite $(u_n)_{n\geqslant 1}$ est décroissante puisque

$$u_n - u_{n+1} = -\frac{1}{n+1} - \ln n + \ln(n+1) = -\frac{1}{n+1} - \ln\left(1 - \frac{1}{n+1}\right) \geqslant 0,$$

en vertu de l'inégalité $\ln(1+x) \leqslant x$ pour $x > -1$. D'autre part, cette même inégalité assure la croissance de $(v_n)_{n\geqslant 1}$:

$$v_{n+1} - v_n = \frac{1}{n} - \ln(n+1) + \ln n = \frac{1}{n} - \ln\left(1 + \frac{1}{n}\right) \geqslant 0.$$

Les deux suites $(u_n)_{n\geqslant 1}$ et $(v_n)_{n\geqslant 1}$ sont donc adjacentes, et elles convergent vers un réel γ. Comme $v_2 = 1 - \ln 2 > 0$, on a $\gamma > 0$.

La limite γ est la constante d'Euler dont 0,577215 *est une valeur approchée par défaut. On sait peu de choses de γ et, notamment, on ignore encore si γ est rationnel ou non.*

2. On a prouvé que $\mathrm{H}_n = \ln n + \gamma + o(1)$ pour n tendant vers $+\infty$.

• Posons $t_n = u_n - \gamma$ $(n \geqslant 1)$. On emploie une méthode classique qui consiste, pour obtenir un équivalent de t_n, à chercher un équivalent de $t_n - t_{n-1}$, puis à «sommer» l'équivalent obtenu. On a pour n tendant vers l'infini,

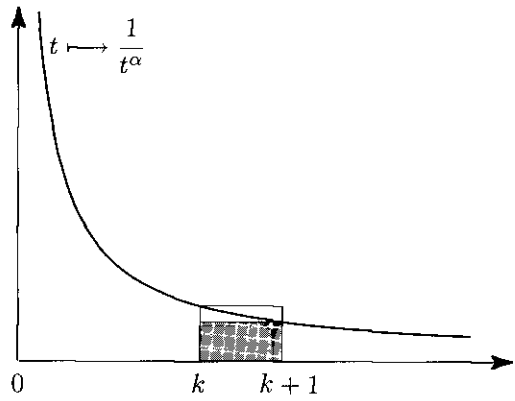
$$t_n - t_{n-1} = \ln\left(1 - \frac{1}{n}\right) + \frac{1}{n} \underset{n\to\infty}{\sim} -\frac{1}{2n^2}.$$

La série $\sum(t_k - t_{k-1})$ converge. Le théorème de sommation des équivalents nous donne

$$\sum_{k=n+1}^{+\infty} (t_k - t_{k-1}) = -t_n \underset{n\to\infty}{\sim} -\frac{1}{2} \sum_{k=n+1}^{+\infty} \frac{1}{k^2} \underset{n\to\infty}{\sim} -\frac{1}{2n},$$

l'équivalent du reste de la série de Riemann s'obtenant à l'aide d'une simple comparaison série-intégrale : si $\alpha > 1$, la fonction $t \longmapsto \dfrac{1}{t^\alpha}$ est décroissante et intégrable sur $[1, +\infty[$, si bien que pour $k \geqslant 2$,

$$\frac{1}{(k+1)^\alpha} \leqslant \int_k^{k+1} \frac{\mathrm{d}t}{t^\alpha} \leqslant \frac{1}{k^\alpha} \leqslant \int_{k-1}^{k} \frac{\mathrm{d}t}{t^\alpha}.$$



En sommant cela entre $n+1$ et N, puis en faisant tendre N vers l'infini, on obtient

$$\int_n^{+\infty} \frac{\mathrm{d}t}{t^\alpha} \leqslant \sum_{k=n+1}^{+\infty} \frac{1}{k^\alpha} \leqslant \int_{n+1}^{+\infty} \frac{\mathrm{d}t}{t^\alpha}.$$

Comme les membres de gauche et de droite sont tous deux équivalents à $\dfrac{1}{\alpha-1}\dfrac{1}{n^{\alpha-1}}$, le théorème d'encadrement assure que

$$\sum_{k=n+1}^{+\infty} \frac{1}{k^\alpha} \underset{n\to\infty}{\sim} \frac{1}{\alpha-1}\frac{1}{n^{\alpha-1}}.$$

Le cas $\alpha = 2$ donne l'équivalent annoncé. On a donc déjà

$$\mathrm{H}_n = \ln n + \gamma + \frac{1}{2n} + o\left(\frac{1}{n}\right).$$

- On pose $w_n = u_n - \gamma - \dfrac{1}{2n}$ pour tout $n \geqslant 1$, suite qui converge vers 0. La somme $\displaystyle\sum_{k=n+1}^{+\infty} (w_k - w_{k-1})$ vaut $-w_n$ et son terme général s'écrit

$$w_n - w_{n-1} = \ln\left(1 - \frac{1}{n}\right) + \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n-2}.$$

Pour n tendant vers l'infini, on a

$$\begin{aligned} w_n - w_{n-1} &= -\frac{1}{n} - \frac{1}{2n^2} - \frac{1}{3n^2} + \frac{1}{n} - \frac{1}{2n} + \frac{1}{2n}\frac{1}{1-\frac{1}{n}} + o\left(\frac{1}{n^3}\right) \\ &= -\frac{1}{2n^2} - \frac{1}{3n^2} - \frac{1}{2n} + \frac{1}{2n}\left(1 + \frac{1}{n} + \frac{1}{n^2}\right) + o\left(\frac{1}{n^3}\right) \\ &= -\frac{1}{3n^3} + \frac{1}{2n^3} + o\left(\frac{1}{n^3}\right) \\ &= \frac{1}{6n^3} + o\left(\frac{1}{n^3}\right) \underset{n\to\infty}{\sim} \frac{1}{6n^3}. \end{aligned}$$

Le théorème de sommation des équivalents donne

$$-w_n \underset{n\to\infty}{\sim} \sum_{k=n+1}^{+\infty} \frac{1}{6k^3} \underset{n\to\infty}{\sim} \frac{1}{12n^2}.$$

Ainsi, on obtient le développement asymptotique

$$\boxed{H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + o\left(\frac{1}{n^2}\right).}$$

3. Pour estimer k_n on va utiliser le début du développement asymptotique de H_n. On sait que $H_n = \ln n + \gamma + \varepsilon_n$ où (ε_n) tend vers 0. Par définition de k_n, on a

$$\ln k_n + \gamma + \varepsilon_{k_n} \geqslant n$$

et

$$\ln(k_n - 1) + \gamma + \varepsilon_{k_n - 1} < n.$$

Il en résulte, en passant à l'exponentielle, que

$$e^n e^{-\gamma - \varepsilon_{k_n - 1}} + 1 > k_n \geqslant e^n e^{-\gamma - \varepsilon_{k_n}}.$$

On a donc $k_n \sim e^n e^{-\gamma}$ et $\boxed{\lim_{n\to+\infty} \frac{k_{n+1}}{k_n} = e}$. ◁

Les sommes de séries que l'on peut calculer explicitement sont assez rares. C'est le cas dans les deux exercices qui suivent. On va y retrouver la série harmonique.

3.19. Calcul de la somme d'une série (1)

Établir la convergence et calculer la somme

$$\sum_{n=0}^{+\infty} \left(\frac{1}{4n+1} - \frac{3}{4n+2} + \frac{1}{4n+3} + \frac{1}{4n+4} \right).$$

(École polytechnique)

Solution.

Notons tout d'abord qu'en vertu du théorème de comparaison des séries à termes positifs, la série converge : en réduisant au même dénominateur on voit que le terme général est

$$\frac{-16n^2 - 12n + 1}{4(n+1)(2n+1)(4n+1)(4n+3)} \sim -\frac{1}{8n^2}.$$

Pour calculer la somme on utilise le développement asymptotique de $H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + \gamma + o(1)$ (voir l'exercice 3.18). Il vient :

$$\begin{aligned} S_n &= \sum_{k=0}^{n} \left(\frac{1}{4k+1} - \frac{3}{4k+2} + \frac{1}{4k+3} + \frac{1}{4k+4} \right) \\ &= \sum_{k=0}^{n} \left(\frac{1}{4k+1} + \frac{1}{4k+2} + \frac{1}{4k+3} + \frac{1}{4k+4} \right) - \sum_{k=0}^{n} \frac{4}{4k+2} \\ &= H_{4n+4} - 2 \sum_{k=0}^{n} \frac{1}{2k+1} = H_{4n+4} - 2\Big(H_{2n+1} - \frac{1}{2} H_n\Big) \\ &= \ln(4n+4) + \ln n + 2\gamma - 2(\ln(2n+1) + \gamma) + o(1). \end{aligned}$$

On a donc $S_n = \ln \dfrac{4n(n+1)}{(2n+1)^2} + o(1)$ ce qui tend vers 0 lorsque n tend vers $+\infty$. Par conséquent la somme de la série est nulle. ◁

L'exemple suivant est nettement plus ardu.

3.20. Calcul de la somme d'une série (2)

Établir la convergence et calculer la somme de la série

$$\sum_{n=1}^{+\infty} n \sum_{p=2^{n-1}}^{2^n-1} \frac{1}{p(2p+1)(2p+2)}.$$

(École polytechnique)

▷ **Solution.**

• Pour $n \geqslant 1$, notons $u_n = \sum_{p=1}^{n-1} \frac{1}{p(2p+1)(2p+2)}$. Le terme général de la série étudiée s'écrit $n(u_{2^n} - u_{2^{n-1}})$. À l'aide d'une comparaison série-intégrale, majorons grossièrement le terme général pour vérifier rapidement la convergence. Comme $\dfrac{1}{p^3} \leqslant \displaystyle\int_{p-1}^{p} \frac{dt}{t^3}$ par décroissance de $t \longmapsto \dfrac{1}{t^3}$ sur $\mathbb{R}_+^*$, on peut majorer ainsi $u_{2^n} - u_{2^{n-1}}$:

$$u_{2^n} - u_{2^{n-1}} \leqslant \sum_{k=2^{n-1}}^{2^n-1} \frac{1}{p^3} \leqslant \sum_{k=2^{n-1}}^{2^n-1} \int_{p-1}^{p} \frac{dt}{t^3} \leqslant \int_{2^{n-1}-1}^{2^n-1} \frac{dt}{t^3} \leqslant \int_{2^{n-2}}^{+\infty} \frac{dt}{t^3},$$

et on obtient

$$n(u_{2^n} - u_{2^{n-1}}) \leqslant \frac{n}{2.(2^{n-2})^2} = \frac{n}{2^{2n-3}} = o\left(\frac{1}{n^2}\right).$$

D'après le théorème de comparaison des séries à termes positifs, la série étudiée converge.

• On va commencer par chercher une expression plus simple de u_n en décomposant en éléments simples la fraction $\dfrac{1}{X(2X+1)(2X+2)}$ sur $\mathbb{C}$. On obtient $\dfrac{1}{2X} + \dfrac{1}{2X+2} - \dfrac{2}{2X+1}$. Pour $n \geqslant 1$, on a alors

$$\begin{aligned}
u_n &= \sum_{p=1}^{n-1} \left(\frac{1}{2p} + \frac{1}{2(p+1)} - \frac{2}{2p+1}\right) \\
&= \frac{1}{2}\sum_{p=1}^{n-1}\frac{1}{p} + \frac{1}{2}\sum_{p=2}^{n}\frac{1}{p} - 2\sum_{p=1}^{n-1}\frac{1}{2p+1} \\
&= \sum_{p=1}^{n-1}\frac{1}{p} + \frac{1}{2n} - \frac{1}{2} - 2\sum_{p=1}^{n-1}\frac{1}{2p+1} \\
&= \sum_{p=1}^{n-1}\frac{1}{p} + \frac{1}{2n} - \frac{1}{2} - 2\left(\sum_{k=1}^{2n-1}\frac{1}{k} - \sum_{k=1}^{n-1}\frac{1}{2k} - 1\right) \\
&= 2\sum_{p=1}^{n-1}\frac{1}{p} + \frac{1}{2n} - \frac{1}{2} - 2\sum_{p=1}^{2n-1}\frac{1}{p} + 2 = -2\sum_{p=n}^{2n-1}\frac{1}{p} + \frac{1}{2n} + \frac{3}{2},
\end{aligned}$$

ce qui donne $u_{2^n} = -2\sum_{p=2^n}^{2^{n+1}-1}\frac{1}{p} + \frac{1}{2^{n+1}} + \frac{3}{2}$ pour tout $n \geqslant 0$.

• Si on note pour $N \geqslant 1$, $S_N = \sum_{n=1}^{N} n(u_{2^n} - u_{2^{n-1}})$ la somme partielle de la série à étudier, on a

$$\begin{aligned}
S_N &= \sum_{n=1}^{N} n(u_{2^n} - u_{2^{n-1}}) = \sum_{n=1}^{N} nu_{2^n} - \sum_{n=0}^{N-1}(n+1)u_{2^n} \\
&= Nu_{2^N} - \sum_{n=0}^{N-1} u_{2^n}.
\end{aligned}$$

Nous allons déterminer la limite de (S_N) en cherchant un développement asymptotique de chacun des termes ci-dessus jusqu'à un terme constant. Commençons par la somme $\sum_{n=0}^{N-1} u_{2^n}$.

$$\begin{aligned}\sum_{n=0}^{N-1} u_{2^n} &= -2\sum_{p=1}^{2^N-1}\frac{1}{p} + \sum_{p=0}^{N-1}\frac{1}{2^{p+1}} + \frac{3}{2}N\\ &= -2H_{2^N-1} + \frac{1}{2}\frac{1-(1/2)^N}{1-(1/2)} + \frac{3}{2}N\\ &= -2H_{2^N-1} + 1 - \left(\frac{1}{2}\right)^N + \frac{3}{2}N,\end{aligned}$$

où $H_n = \sum_{p=1}^{n}\frac{1}{p}$ désigne la somme partielle de la série harmonique. Comme pour n tendant vers l'infini, $H_n = \ln n + \gamma + O\left(\frac{1}{n}\right)$ où γ désigne la constante d'Euler, on obtient le développement asymptotique pour N tendant vers l'infini :

$$\begin{aligned}\sum_{n=0}^{N-1} u_{2^n} &= -2\ln(2^N-1) - 2\gamma + O\left(\frac{1}{2^N-1}\right) + 1 - \left(\frac{1}{2}\right)^N + \frac{3}{2}N\\ &= -2\ln 2^N - 2\ln(1-2^{-N}) - 2\gamma + o(1) + 1 + \frac{3}{2}N\\ &= -(2\ln 2)N + \frac{3}{2}N - 2\gamma + 1 + o(1).\end{aligned}$$

Il reste à regarder le terme Nu_{2^N}. Pour N tendant vers l'infini, on a

$$\begin{aligned}u_{2^N} &= -2H_{2^{N+1}-1} + 2H_{2^N-1} + \frac{1}{2^{N+1}} + \frac{3}{2}\\ &= -2\ln(2^{N+1}-1) + 2\ln(2^N-1) - 2\gamma + 2\gamma + O\left(\frac{1}{2^{N+1}-1}\right)\\ &\quad + \frac{1}{2^{N+1}} + \frac{3}{2} + O\left(\frac{1}{2^{N+1}-1}\right)\end{aligned}$$

$$\begin{aligned}Nu_{2^N} &= -2N(\ln 2^{N+1} - \ln 2^N) - 2N(\ln(1-2^{-N-1}) - \ln(1-2^N))\\ &\quad + O\left(\frac{N}{2^N}\right) + \frac{3}{2}N\\ &= -2N\ln 2 - N\,O\left(\frac{1}{2^N}\right) + O\left(\frac{N}{2^N}\right) + \frac{3}{2}N\\ &= -(2\ln 2)\ln N + \frac{3}{2}N + o(1).\end{aligned}$$

Il en résulte que, pour N tendant vers l'infini,

$$\begin{aligned}S_N &= -(2\ln 2)N + \frac{3}{2}N + o(1) + (2\ln 2)N - \frac{3}{2}N + 2\gamma - 1 + o(1)\\ &= 2\gamma - 1 + o(1) \xrightarrow[N\to\infty]{} 2\gamma - 1.\end{aligned}$$

Conclusion. La série étudiée converge et

$$\boxed{\sum_{n=1}^{+\infty} n \sum_{p=2^{n-1}}^{2^n-1} \frac{1}{p(2p+1)(2p+2)} = 2\gamma - 1.} \quad \lhd$$

L'énoncé suivant se ramène à la recherche d'un développement asymptotique assez précis de la somme partielle d'une série grossièrement divergente.

3.21. Recherche d'un équivalent

Trouver un équivalent de $u_n = \sqrt[n]{\prod_{k=1}^{n} k^k}$ lorsque n tend vers l'infini.

(École normale supérieure)

▷ **Solution.**

Pour n entier naturel non nul, on pose $v_n = \ln u_n = \frac{1}{n} \sum_{k=1}^{n} k \ln k$. On va chercher un développement asymptotique de la somme partielle $\sum_{k=1}^{n} k \ln k$. La fonction $t \longmapsto t \ln t$ est croissante sur $[1, +\infty[$, ce qui invite à utiliser une comparaison série-intégrale. On a

$$\int_1^n t \ln t \mathrm{d}t \leqslant \sum_{k=1}^{n} k \ln k \leqslant \int_1^{n+1} t \ln t \mathrm{d}t.$$

À l'aide d'une intégration par parties, il vient

$$\int_1^n x \ln x \mathrm{d}x = \left[\frac{x^2}{2} \ln x\right]_1^n - \int_1^n \frac{x}{2} \mathrm{d}x = \frac{n^2 \ln n}{2} - \frac{n^2}{4} + \frac{1}{4}.$$

Cette expression prise au rang $n+1$, tout comme celle du rang n, est équivalente à $\frac{n^2 \ln n}{2}$, de sorte que l'inégalité ci-dessus permet d'affirmer que

$$v_n \underset{n\to\infty}{\sim} \frac{1}{n} \int_1^n x \ln x \mathrm{d}x \underset{n\to\infty}{\sim} \frac{n \ln n}{2},$$

ce qui s'écrit $v_n = \frac{n \ln n}{2} + o(n \ln n)$. Malheureusement, la limite du $o(n \ln n)$ n'est pas forcément nulle, et après le passage à l'exponentielle,

cela ne permet pas d'obtenir un équivalent de u_n. Il faut pousser plus loin le développement asymptotique de $\sum_{k=1}^{n} k \ln k$. Posons

$$w_n = \sum_{k=1}^{n} k \ln k - \int_1^n t \ln t \mathrm{d}t = \sum_{k=2}^{n} \int_{k-1}^{k} (k \ln k - x \ln x) \mathrm{d}x.$$

La dérivée de la fonction $x \longmapsto x \ln x$ étant $x \longmapsto 1 + \ln x$ qui est croissante, le théorème des accroissements finis assure que pour $x \in [k-1, k]$ $(k \geqslant 2)$, on a

$$(1 + \ln(k-1))(k-x) \leqslant k \ln k - x \ln x \leqslant (1 + \ln k)(k-x),$$

ce qui donne en intégrant entre $k-1$ et k.

$$\frac{1}{2} + \frac{\ln(k-1)}{2} \leqslant \int_{k-1}^{k} (k \ln k - x \ln x) \mathrm{d}x \leqslant \frac{1}{2} + \frac{\ln k}{2}.$$

En sommant, on obtient un encadrement de w_n,

$$\frac{n-1}{2} + \frac{1}{2} \sum_{k=2}^{n} \ln(k-1) \leqslant w_n \leqslant \frac{n-1}{2} + \frac{1}{2} \sum_{k=2}^{n} \ln k.$$

ce qui s'écrit encore

$$\frac{n}{2} + \frac{1}{2} \ln n! - \frac{\ln n}{2} + O(1) \leqslant w_n \leqslant \frac{n}{2} + \frac{1}{2} \ln n! + O(1).$$

On a, d'après la formule de Stirling,

$$\ln n! = \ln \left(\sqrt{2\pi n} \left(\frac{n}{e} \right)^n (1 + o(1)) \right) = n \ln n - n + \frac{\ln n}{2} + O(1).$$

En reportant dans l'encadrement de w_n, il vient

$$\frac{n \ln n}{2} + O(\ln n) \leqslant w_n \leqslant \frac{n \ln n}{2} + O(\ln n)$$

si bien que $w_n = \dfrac{n \ln n}{2} + O(\ln n)$. On a donc

$$\sum_{k=1}^{n} k \ln k = \frac{n^2 \ln n}{2} - \frac{n^2}{4} + \frac{n \ln n}{2} + O(\ln n),$$

soit en divisant par n,

$$v_n = \frac{n \ln n}{2} - \frac{n}{4} + \frac{\ln n}{2} + O\left(\frac{\ln n}{n} \right)$$

ou plus simplement $v_n = \dfrac{n\ln n}{2} - \dfrac{n}{4} + \dfrac{\ln n}{2} + o(1)$. En passant à l'exponentielle, il vient

$$u_n = \exp\left(\frac{n+1}{2}\ln n - \frac{n}{4} + o(1)\right) = (\sqrt{n})^{n+1}e^{-\frac{n}{4}}e^{o(1)}$$

et donc

$$\boxed{u_n \underset{n\to\infty}{\sim} (\sqrt{n})^{n+1}e^{-\frac{n}{4}}.} \quad \lhd$$

Une autre approche consiste à faire apparaître à partir de v_n une somme de Riemann relative à la fonction $x \longmapsto x\ln x$ sur $[0,1]$ prolongée par continuité en 0. On écrit

$$v_n = \sum_{k=1}^{n} \frac{k}{n}\ln\frac{k}{n} + \sum_{k=1}^{n}\frac{k}{n}\ln n = n\mathrm{S}_n + \frac{n+1}{2}\ln n$$

où $\mathrm{S}_n = \dfrac{1}{n}\sum_{k=1}^{n}\dfrac{k}{n}\ln\dfrac{k}{n}$ est la somme de Riemann en question. Comme la fonction $x \longmapsto x\ln x$ est continue sur $[0,1]$, on sait que S_n converge vers $\int_0^1 t\ln t\,dt = -\dfrac{1}{4}$. Cela ne suffit pas pour conclure car on obtient un développement asymptotique de v_n en $o(n)$. Il faut estimer la différence entre la somme de Riemann et l'intégrale. C'est un exercice assez classique, lorsque la fonction intégrée f est de classe C^1, de montrer que

$$\frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right) - \int_0^1 f = \frac{1}{n}\int_0^1 f' + o\left(\frac{1}{n}\right).$$

On ne peut invoquer ce résultat directement car $x \longmapsto x\ln x$ n'est pas dérivable en 0. Cela dit sa dérivée sur $]0,1]$ est $x \longmapsto 1+\ln x$ et il s'agit d'une fonction intégrable et monotone sur $]0,1]$. On peut prouver que sous ces hypothèses le développement précédent demeure valable. On en déduit que $\mathrm{S}_n = -\dfrac{1}{4} + o\left(\dfrac{1}{n}\right)$, ce qui permet de retrouver l'équivalent de u_n.

Voici maintenant plusieurs exercices en liaison avec l'arithmétique. Si (u_n) est une suite strictement croissante d'entiers $\geqslant 1$, on conçoit que la série $\sum \dfrac{1}{u_n}$ converge si on prend peu d'entiers (par exemple $u_n = 2^n$, $u_n = n^2$) et diverge si on en prend trop (par exemple $u_n = 2n$). L'exercice qui suit montre qu'il y a suffisamment de nombres premiers pour qu'avec cette suite on soit dans le cas de la divergence.

3.22. Série des inverses des entiers premiers

Soit $(p_n)_{n\geqslant 1}$ la suite croissante des nombres premiers. Montrer que la série $\sum \dfrac{1}{p_n}$ diverge.

(École normale supérieure)

▷ **Solution.**

Rappelons que l'ensemble des nombres premiers est infini, ce qui justifie l'existence de la suite $(p_n)_{n\geqslant 1}$. Cette suite d'entiers diverge vers l'infini. Supposons par l'absurde que la série $\sum \dfrac{1}{p_n}$ converge. Il en est alors de même de la série $\sum -\ln\left(1-\dfrac{1}{p_n}\right)$ puisque les termes généraux de ces deux séries sont positifs et équivalents. En passant à l'exponentielle, on en déduit que $u_{\mathrm{N}} = \displaystyle\prod_{n=1}^{\mathrm{N}} \frac{1}{1-p_n^{-1}}$ converge. Or, comme pour tout $\mathrm{M}\geqslant 1$,

$$\frac{1}{1-p_n^{-1}} = \sum_{k=0}^{+\infty} \frac{1}{p_n^k} \geqslant 1 + \frac{1}{p_n} + \frac{1}{p_n^2} + \cdots + \frac{1}{p_n^{\mathrm{M}}},$$

on obtient

$$u_{\mathrm{N}} = \prod_{n=1}^{\mathrm{N}} \sum_{k=0}^{+\infty} \frac{1}{p_n^k} \geqslant \prod_{n=1}^{\mathrm{N}} \left(1 + \frac{1}{p_n} + \frac{1}{p_n^2} + \cdots + \frac{1}{p_n^{\mathrm{M}}}\right).$$

Il nous suffit de choisir M assez grand pour obtenir dans ce produit les inverses de tous les entiers compris entre 1 et p_{N} (ces entiers ont tous leurs facteurs premiers parmi $p_1, \ldots, p_{\mathrm{N}}$). On a donc

$$u_{\mathrm{N}} \geqslant \sum_{i=1}^{p_{\mathrm{N}}} \frac{1}{i}$$

ce qui est absurde puisque le terme de droite tend vers l'infini (la série harmonique diverge et p_{N} tend vers l'infini). ◁

Le célèbre théorème des nombres premiers affirme que $\pi(n) \sim \dfrac{n}{\ln n}$ *où* $\pi(n)$ *est le nombre d'entiers premiers inférieurs à* n. *On peut en déduire un équivalent de* p_n. *En effet, comme* $\pi(p_n) = n$ *on a* $\dfrac{p_n}{\ln p_n} \sim n$, *ou encore* $\dfrac{p_n}{\ln p_n} = n + o(n)$. *On prend le logarithme de cette égalité. Il vient* $\ln p_n - \ln\ln p_n = \ln n + \ln(1+o(1)) = \ln n + o(1)$. *Cela prouve que* $\ln p_n \sim \ln n$, *puis que* $p_n \sim n\ln n$, *équivalent qui éclaire la divergence de la série* $\sum \dfrac{1}{p_n}$.

3.23. Série définie à partir de ppcm

Soit $(u_n)_{n\in\mathbb{N}}$ une suite strictement croissante d'entiers naturels non nuls. On pose $a_n = \text{ppcm}(u_0, u_1, \ldots, u_n)$. Étudier la nature de la série $\sum \dfrac{1}{a_n}$.

(École polytechnique)

Solution.

La suite (a_n) est croissante et diverge vers $+\infty$. Notons $(n_k)_{k\geqslant 0}$ la suite strictement croissante telle que $a_{n_{k+1}} > a_{n_k}$ et (a_i) constante sur l'intervalle $[\![n_k, n_{k+1} - 1]\!]$: les indices n_k donnent les premiers termes des segments sur lesquels la suite (a_n) reste constante. En regroupant les termes égaux, la série positive $\displaystyle\sum_{n\geqslant 0} \frac{1}{a_n}$ est de même nature que la série $\displaystyle\sum_{k\geqslant 0} \frac{n_{k+1} - n_k}{a_{n_k}}$. On va majorer le numérateur et minorer le dénominateur de ce terme général.

Le fait que $a_i = a_{n_k}$ pour $i \in [\![n_k, n_{k+1} - 1]\!]$ implique que u_i divise a_{n_k} (puisque le ppcm reste le même). Comme les termes de la suite (u_n) sont deux à deux distincts, il y a au plus $d(a_{n_k})$ indices i dans $[\![n_k, n_{k+1} - 1]\!]$ (où $d(m)$ désigne le nombre de diviseurs de l'entier m). Autrement dit on a la majoration $n_{k+1} - n_k \leqslant d(a_{n_k})$. De plus, il est clair que pour tout entier $m \in \mathbb{N}^*$ on a la majoration $d(m) \leqslant 2\sqrt{m}$. On a donc $n_{k+1} - n_k \leqslant d(a_{n_k}) \leqslant 2\sqrt{a_{n_k}}$ et par conséquent

$$\frac{n_{k+1} - n_k}{a_{n_k}} \leqslant \frac{2}{\sqrt{a_{n_k}}}.$$

Il reste à minorer a_{n_k}. Mais comme a_{n_k} est un diviseur strict de $a_{n_{k+1}}$ on a $a_{n_{k+1}} \geqslant 2a_{n_k}$. En itérant cela, on a $a_{n_k} \geqslant 2^k a_{n_0}$ pour tout k. Il en découle que

$$\frac{n_{k+1} - n_k}{a_{n_k}} \leqslant \frac{2}{\sqrt{a_{n_k}}} = O\left(\frac{1}{\left(\sqrt{2}\right)^k}\right)$$

ce qui montre la convergence de la série. ◁

3.24. Nature d'une série

Pour $n \geqslant 2$, on note q_n le plus grand diviseur premier de n. Quelle est la nature de la série de terme général $\dfrac{1}{nq_n}$?

(École normale supérieure)

▷ **Solution.**

Notons $(p_k)_{k\geqslant 1}$ la suite croissante des nombres premiers. Il est assez naturel de regrouper les entiers en fonction de la valeur de q_n. Posons donc $\Omega_k = \{n \in \mathbb{N}^*,\ q_n = p_k\}$. Les ensembles $(\Omega_k)_{k\geqslant 1}$ forment une partition de $\mathbb{N}^* \setminus \{1\}$.

Les entiers de Ω_k sont ceux de la forme $p_k(p_1^{\alpha_1}\dots p_k^{\alpha_k})$ avec $(\alpha_1,\dots,\alpha_k) \in \mathbb{N}^k$. On en déduit que la famille $\left(\dfrac{1}{nq_n}\right)_{n\in\Omega_k}$ est sommable et que

$$\sum_{n\in\Omega_k}\frac{1}{nq_n} = \sum_{(\alpha_1,\dots,\alpha_k)\in\mathbb{N}^k}\frac{1}{p_k^2(p_1^{\alpha_1}\dots p_k^{\alpha_k})} = \frac{\mathrm{M}_k}{p_k^2}$$

où, par produit de Cauchy,

$$\mathrm{M}_k = \sum_{(\alpha_1,\dots,\alpha_k)\in\mathbb{N}^k}\frac{1}{p_1^{\alpha_1}\dots p_k^{\alpha_k}} = \prod_{i=1}^{k}\sum_{\alpha_i=0}^{+\infty}\frac{1}{p_i^{\alpha_i}} = \prod_{i=1}^{k}\left(1-\frac{1}{p_i}\right)^{-1}.$$

Par associativité pour les familles positives, la série $\displaystyle\sum_{n\geqslant 2}\frac{1}{nq_n}$ converge si et seulement si la série $\displaystyle\sum_{k\geqslant 1}\frac{\mathrm{M}_k}{p_k^2}$ converge. Montrons que c'est le cas en majorant directement M_k. On passe au logarithme :

$$\begin{aligned}\ln \mathrm{M}_k &= -\sum_{i\leqslant k}\ln\left(1-\frac{1}{p_i}\right) = \sum_{i\leqslant k}\ln\left(\frac{p_i}{p_i-1}\right)\\ &= \sum_{i\leqslant k}\ln\left(1+\frac{1}{p_i-1}\right) \leqslant \sum_{i\leqslant k}\frac{1}{p_i-1}\end{aligned}$$

grâce à la majoration classique $\ln(1+x)\leqslant x$. On a $p_i - 1 \geqslant 2(i-1)$ pour tout $i\geqslant 1$ (car, hormis 2, les nombres premiers sont impairs). On en déduit que $\ln \mathrm{M}_k \leqslant 1 + \dfrac{1}{2}\displaystyle\sum_{i=2}^{k}\frac{1}{i-1} = 1 + \frac{1}{2}\mathrm{H}_{k-1} = O(\ln\sqrt{k})$. On a donc $\mathrm{M}_k = O(\sqrt{k})$ et finalement $\dfrac{\mathrm{M}_k}{p_k^2} = O\left(\dfrac{\sqrt{k}}{p_k^2}\right) = O\left(\dfrac{1}{k^{3/2}}\right)$ et c'est terminé en vertu du théorème de comparaison des séries à termes positifs. ◁

Un théorème de Mertens montre que $\mathrm{M}_k \sim e^{\gamma}\ln p_k$ et cela explique qu'il soit relativement facile d'obtenir une domination de la forme $\mathrm{M}_k = O(p_k^{\alpha})$ avec $\alpha > 0$.

La fonction arithmétique de Möbius a sans doute déjà été rencontrée par le lecteur (par exemple dans l'exercice 4.32. du tome 1 d'algèbre). L'exercice suivant établit la formule d'inversion de Möbius et étudie ensuite l'inversion d'un développement asymptotique.

3.25. Inversion de Möbius

Soit $f : [1, +\infty[\to \mathbb{R}$ continue par morceaux. On lui associe la fonction $\hat{f} : [1, +\infty[\to \mathbb{R}$ définie par $\hat{f}(x) = \sum_{k=1}^{\mathrm{E}(x)} f\left(\frac{x}{k}\right)$ pour $x \geqslant 1$.

1. Montrer que, pour $x \geqslant 1$, $f(x) = \sum_{n=1}^{\mathrm{E}(x)} \mu(n) \hat{f}\left(\frac{x}{n}\right)$, où μ désigne la fonction de Möbius : $\mu(1) = 1$, $\mu(n) = (-1)^r$ si n est le produit de r nombres premiers deux à deux distincts et $\mu(n) = 0$ sinon.

2. On suppose que, pour x tendant vers l'infini,

$$\hat{f}(x) = \mathrm{A}x(\ln x)^2 + \mathrm{B}x \ln x + \mathrm{C}x + O(x^\beta)$$

où $\beta \in]0, 1[$ et A, B, C sont des constantes. Montrer que

$$f(x) = 2\mathrm{A}x \ln x + O(x).$$

(École normale supérieure)

▷ **Solution.**

1. On convient que k désigne un entier strictement positif. Pour $x \geqslant 1$, on notera donc $\sum_{k \leqslant x}$ pour désigner la somme portant sur les entiers naturels non nuls inférieurs à x. On va calculer la somme proposée en remplaçant $\hat{f}$ par son expression. Il vient

$$\sum_{n \leqslant x} \mu(n) \hat{f}\left(\frac{x}{n}\right) = \sum_{n \leqslant x} \mu(n) \sum_{k \leqslant \frac{x}{n}} f\left(\frac{x}{kn}\right) = \sum_{kn \leqslant x} \mu(n) f\left(\frac{x}{kn}\right).$$

Dans cette dernière somme, on regroupe les couples (k, n) selon la valeur m de leur produit. On obtient

$$\sum_{m \leqslant x} f\left(\frac{x}{m}\right) \sum_{n|m} \mu(n).$$

Il ne reste plus qu'à montrer que si on pose $\varepsilon(m) = \sum_{n|m} \mu(n)$, alors $\varepsilon(1) = 1$ et $\varepsilon(m) = 0$ si $m > 1$. Le fait que $\varepsilon(1) = 1$ est clair. Soit $m > 1$ que l'on décompose en facteurs premiers, $m = p_1^{\alpha_1} \dots p_k^{\alpha_k}$. Dans la somme définissant $\varepsilon(m)$ on peut se contenter de prendre les diviseurs sans facteurs carrés (on entend par là les entiers $n \geqslant 1$ qui sont produit de nombres premiers deux à deux distincts). De tels entiers s'écrivent $\prod_{i \in I} p_i$ avec $I \subset [\![1, k]\!]$. On obtient ainsi

$$\varepsilon(m) = \sum_{\mathrm{I}\subset[\![1,k]\!]} \mu\left(\prod_{i\in\mathrm{I}} p_i\right) = \sum_{\mathrm{I}\subset[\![1,k]\!]} (-1)^{|\mathrm{I}|} = 0$$

car il y a autant de parties de $[\![1,k]\!]$ de cardinal pair que de parties de cardinal impair.

On peut aussi, dans le calcul de $\varepsilon(m)$, regrouper les parties I *de* $[\![1,k]\!]$ *selon leur cardinal i. On obtient alors*

$$\sum_{i=0}^{k} \mathrm{C}_k^i (-1)^i = (1-1)^k = 0$$

par la formule du binôme de Newton.

2. Il est naturel de commencer par étudier le développement asymptotique de $\hat{g}$ pour la fonction $g : x \longmapsto x\ln x$. On a

$$\hat{g}(x) = \sum_{k\leqslant x} \frac{x}{k} \ln\frac{x}{k} = x\ln x \sum_{k\leqslant x} \frac{1}{k} - x \sum_{k\leqslant x} \frac{\ln k}{k}.$$

Pour le premier terme de cette différence, on va utiliser le développement asymptotique classique $\displaystyle\sum_{k\leqslant x} \frac{1}{k} = \ln x + \gamma + O(1/x)$ où γ désigne la constante d'Euler (cf. exercice 3.18). Il s'obtient à partir de celui de la somme partielle de la série harmonique. En effet, étant donné que $\mathrm{E}(x) = x + O(1)$, on a

$$\begin{aligned}\sum_{k\leqslant x} \frac{1}{k} &= \ln \mathrm{E}(x) + \gamma + O\left(\frac{1}{\mathrm{E}(x)}\right) = \ln(x + O(1)) + \gamma + O\left(\frac{1}{x}\right)\\ &= \ln x + \ln\left(1 + \left(\frac{1}{x}\right)\right) + \gamma + O(1/x) = \ln x + \gamma + O\left(\frac{1}{x}\right).\end{aligned}$$

Cherchons maintenant un développement asymptotique de $\displaystyle\sum_{k\leqslant x} \frac{\ln k}{k}$. Le calcul de la dérivée de $t \longmapsto \dfrac{\ln t}{t}$ montre la décroissance de cette fonction sur $[e, +\infty[$. On va procéder par une comparaison série-intégrale. Posons, pour $n \geqslant 1$, $u_n = \displaystyle\sum_{k=1}^{n} \frac{\ln k}{k}$. Pour tout $k \geqslant 4 \geqslant e+1$, on a

$$\int_k^{k+1} \frac{\ln t}{t}\mathrm{d}t \leqslant \frac{\ln k}{k} \leqslant \int_{k-1}^{k} \frac{\ln t}{t}\mathrm{d}t,$$

et en sommant de $k = 4$ à $k = n$, on obtient

$$\int_4^{n+1} \frac{\ln t}{t}\mathrm{d}t \leqslant \sum_{k=4}^{n} \frac{\ln k}{k} \leqslant \int_3^{n} \frac{\ln t}{t}\mathrm{d}t.$$

Comme $t \longmapsto \frac{1}{2}\ln^2 t$ est une primitive de $t \longmapsto \frac{\ln t}{t}$, on constate que $\sum_{k=4}^{n} \frac{\ln k}{k}$ est encadré par deux termes équivalents à $\frac{1}{2}\ln^2 n$. Il en résulte que $u_n \underset{n\to\infty}{\sim} \frac{1}{2}\ln^2 n$. Pour pousser le développement asymptotique, on pose $v_n = u_n - \frac{1}{2}\ln^2 n$ et on cherche un équivalent de $v_n - v_{n-1}$. On obtient aisément que $v_n - v_{n-1} \underset{n\to\infty}{\sim} \frac{\ln n}{2n^2}$ qui est le terme général d'une série convergente. On en déduit que v_n tend vers une limite C_1 et qu'en vertu du théorème de sommation des équivalents,

$$C_1 - v_n = \sum_{k=n+1}^{+\infty} v_k - v_{k-1} \underset{n\to\infty}{\sim} \sum_{k=n+1}^{+\infty} \frac{\ln k}{2k^2}.$$

Or, comme $\frac{\ln n}{2n^2} \underset{n\to\infty}{\sim} \frac{\ln(n-1)}{(n-1)} - \frac{\ln n}{n}$, le théorème de sommation des équivalents donne

$$C_1 - v_n \underset{n\to\infty}{\sim} \sum_{k=n+1}^{+\infty} \left(\frac{\ln(k-1)}{(k-1)} - \frac{\ln k}{k}\right) = \frac{\ln n}{n}.$$

Ainsi, on obtient finalement $u_n = \frac{1}{2}\ln^2 n + C_1 + O\left(\frac{\ln n}{n}\right)$.

On peut maintenant obtenir le développement asymptotique de $\sum_{k\leqslant x} \frac{\ln k}{k}$. Si on pose $n = E(x) = x + O(1)$, on a $\frac{\ln x}{x} \underset{n\to\infty}{\sim} \frac{\ln n}{n}$, si bien que

$$\begin{aligned}\sum_{k\leqslant x} \frac{\ln k}{k} = u_n &= \frac{\ln^2(x+O(1))}{2} + C_1 + O\left(\frac{\ln x}{x}\right)\\ &= \frac{\ln^2 x}{2} + C_1 + O\left(\frac{\ln x}{x}\right),\end{aligned}$$

car $\ln^2(x+O(1)) = \ln^2 x + O\left(\frac{\ln x}{x}\right)$. Finalement, on a

$$\hat{g}(x) = \frac{1}{2}x(\ln x)^2 + \gamma x \ln x - C_1 x + O(\ln x).$$

Posons alors $h = f - 2Ag$. Par linéarité, et vu que $\ln x = O(x^\beta)$, on a

$$\hat{h}(x) = C_2 x \ln x + C_3 x + O(x^\beta)$$

(C_2 et C_3 étant des constantes). Or le calcul précédent a aussi montré que

$$\widehat{\mathrm{Id}}(x) = \sum_{k\leqslant x} \frac{x}{k} = x\ln x + \gamma x + O(1).$$

Donc si on pose $k(x) = h(x) - C_2 x$, on a $\hat{k}(x) = C_4 x + O(x^\beta)$, C_4 étant une constante.

Enfin, en posant $l(x) = k(x) - C_4 = f(x) - 2Ax\ln x - C_2 x - C_4$, on obtient

$$\hat{l}(x) = \hat{k}(x) - C_4\,E(x) = C_4(x - E(x)) + O(x^\beta) = O(x^\beta).$$

Utilisons alors la formule d'inversion de Möbius pour la fonction l. On a, en notant $M > 0$ un réel tel que $|\hat{l}(x)| \leqslant Mx^\beta$,

$$|l(x)| \leqslant M \sum_{k<x} \left(\frac{x}{k}\right)^\beta = Mx^\beta \sum_{k\leqslant x} \frac{1}{k^\beta}.$$

Une comparaison série-intégrale classique assure que

$$\sum_{k\leqslant x} \frac{1}{k^\beta} = O\left(\frac{1}{x^{\beta-1}}\right) = O(x^{1-\beta}).$$

Ainsi, $l(x) = O(x)$ et on a donc bien $f(x) = 2Ax\ln x + O(x)$. ◁

La suite τ_n qui donne le nombre de diviseurs de l'entier n a un comportement assez erratique. Elle vaut 2 sur un entier de la forme $n = 2^p - 1$ qui est premier (nombre de Mersenne) puis vaut $p + 1$ sur l'entier suivant qui est une puissance de 2. En effectuant la moyenne de Cesàro de la suite (τ_n), on «lisse» en quelque sorte le comportement de cette suite, et on peut obtenir un équivalent et même un développement asymptotique.

3.26. Nombre moyen de diviseurs des entiers inférieurs à x

Pour $n \in \mathbb{N}^*$, on note τ_n le nombre de diviseurs de n dans $\mathbb{N}$. Si $x \in [1, +\infty[$, on pose $F(x) = \sum_{1\leqslant n\leqslant x} \tau_n$.

1. Trouver un équivalent simple de $F(x)$ lorsque x tend vers $+\infty$.

2. Démontrer que lorsque x tend vers $+\infty$

$$F(x) = x\ln x + (2\gamma - 1)x + O(\sqrt{x}).$$

(École polytechnique)

Solution.

1. Soit $x \geqslant 1$. Les indices sous les signes de sommations désignent toujours des entiers naturels non nuls. On a

$$\begin{aligned}
\mathrm{F}(x) &= \sum_{n\leqslant x}\sum_{d\mid n} 1 = \sum_{\substack{n\leqslant x\\ d\mid n}} 1 \quad \text{(par associativité)}\\
&= \sum_{\substack{d,d'\geqslant 1\\ dd'\leqslant x}} 1 \quad \text{(par changement d'indices } d' = \frac{n}{d}\text{)}\\
&= \sum_{d\leqslant x}\sum_{d'\leqslant \mathrm{E}\left(\frac{x}{d}\right)} 1 \quad \text{(par associativité)}\\
&= \sum_{d\leqslant x} \mathrm{E}\left(\frac{x}{d}\right).
\end{aligned}$$

Pour tout $1 \leqslant d \leqslant x$, on a $\dfrac{x}{d} - 1 \leqslant \mathrm{E}\left(\dfrac{x}{d}\right) \leqslant \dfrac{x}{d}$, ce qui donne l'encadrement suivant

$$\sum_{d\leqslant x}\left(\frac{x}{d} - 1\right) \leqslant \mathrm{F}(x) \leqslant \sum_{d\leqslant x}\frac{x}{d}$$

ou encore

$$x\sum_{d=1}^{\mathrm{E}(x)}\frac{1}{d} - \mathrm{E}(x) \leqslant \mathrm{F}(x) \leqslant x\sum_{d=1}^{\mathrm{E}(x)}\frac{1}{d}.$$

Or $\displaystyle\sum_{d=1}^{\mathrm{E}(x)}\frac{1}{d} \underset{x\to\infty}{\sim} \ln \mathrm{E}(x) \underset{x\to\infty}{\sim} \ln x$. On en déduit que

$$\boxed{\mathrm{F}(x) \underset{x\to\infty}{\sim} x\ln x.}$$

2. Il ne sert à rien de pousser le développement de $\displaystyle\sum_{d=1}^{n}\frac{1}{d}$ un cran plus loin car dans l'encadrement de $\mathrm{F}(x)$ ci-dessus l'écart entre le majorant et le minorant est en $O(x)$. Pour avoir un développement asymptotique de F en $O(\sqrt{x})$ on va utiliser une autre expression de F. L'idée est de couper la somme $\mathrm{F}(x) = \displaystyle\sum_{dd'\leqslant x} 1$ selon la position de d et d' par rapport à $\sqrt{x}$. Si $dd' \leqslant x$, on a soit $d \leqslant \sqrt{x}$ soit $d' \leqslant \sqrt{x}$. Par conséquent, on a

$$\mathrm{F}(x) = 2\sum_{\substack{d\leqslant\sqrt{x}\\ dd'\leqslant x}} 1 - \sum_{d,d'\leqslant\sqrt{x}} 1,$$

ce qui donne

$$\mathrm{F}(x) = 2\left(\sum_{d\leqslant\sqrt{x}}\mathrm{E}\left(\frac{x}{d}\right)\right) - \mathrm{E}(\sqrt{x})^2.$$

On a une somme du même type que précédemment, mais avec un nombre de termes de l'ordre de $\sqrt{x}$. On va pouvoir en déduire le développement de F avec la précision souhaitée encore à l'aide de l'encadrement $u-1 \leqslant \mathrm{E}(u) \leqslant u$ pour $u \in \mathbb{R}$. Il vient, pour x tendant vers l'infini

$$\mathrm{F}(x) = 2\sum_{d\leqslant\sqrt{x}} \frac{x}{d} + O(\sqrt{x}) - (x + O(\sqrt{x}))$$

et

$$\begin{aligned}\sum_{d=1}^{\mathrm{E}(\sqrt{x})} \frac{x}{d} &= x\Big(\ln \mathrm{E}(\sqrt{x}) + \gamma + O\Big(\frac{1}{\sqrt{x}}\Big)\Big)\\ &= x\Big(\ln(\sqrt{x} + O(1)) + \gamma + O\Big(\frac{1}{\sqrt{x}}\Big)\Big).\end{aligned}$$

Or, $\ln(\sqrt{x} + O(1)) = \ln\sqrt{x} + \ln\Big(1 + O\Big(\frac{1}{\sqrt{x}}\Big)\Big) = \frac{1}{2}\ln x + O\Big(\frac{1}{\sqrt{x}}\Big)$.
Donc,

$$\sum_{d=1}^{\mathrm{E}(\sqrt{x})} \frac{x}{d} = \frac{1}{2}x\ln x + \gamma x + O(\sqrt{x}).$$

Par conséquent,

$$\boxed{\mathrm{F}(x) = x\ln x + (2\gamma - 1)x + O(\sqrt{x}).}\ \triangleleft$$

En moyenne les entiers ont donc assez peu de diviseurs : ainsi, le nombre moyen de diviseurs d'un entier compris entre 1 et 10^6 (un million) est approximativement 14. Il ne faudrait toutefois pas croire que le nombre de diviseurs de n est «souvent» de l'ordre de $\ln n$. Le bon ordre de grandeur est $(\ln n)^{\ln 2}$ qui est nettement plus petit. Pour de plus amples renseignements sur le comportement asymptotique de la suite (τ_n) le lecteur pourra consulter HARDY (G.H.) & WRIGHT (E.M.), An introduction to the theory of numbers, *Oxford.*

Les exercices ci-après concernent des séries réelles qui ne sont plus nécessairement positives et des séries complexes. Très souvent l'étude de la convergence d'une série $\sum u_n$ quelconque commence par l'étude de la convergence absolue, c'est-à-dire de la convergence de la série positive $\sum |u_n|$, pour laquelle on peut utiliser les théorèmes de comparaison. La convergence absolue entraîne la convergence de la série : l'inégalité triangulaire assure que pour n et p entiers naturels,

$$|u_{n+1} + u_{n+2} + \cdots + u_{n+p}| \leqslant |u_{n+1}| + |u_{n+2}| + \cdots + |u_{n+p}|.$$

Si $\sum |u_n|$ converge, elle vérifie le critère de Cauchy (pour n assez grand, la tranche $|u_{n+1}| + |u_{n+2}| + \cdots + |u_{n+p}|$ peut être rendue plus petite que

0 *arbitrairement fixé) et par l'inégalité ci-dessus,* $\sum u_n$ *vérifie ce même critère. Comme* $\mathbb{R}$ *et* $\mathbb{C}$ *sont complets, la série* $\sum u_n$ *converge.*

3.27. Nature de $\sum \sin\left(\pi(2+\sqrt{3})^n\right)$

Quelle est la nature de la série $\sum \sin\left(\pi(2+\sqrt{3})^n\right)$?

(École polytechnique)

▷ **Solution.**

Posons $u_n = \sin\left(\pi(2+\sqrt{3})^n\right)$. Déterminer si u_n converge vers 0, revient à savoir si la suite $(2+\sqrt{3})^n$, qui diverge vers l'infini, a tendance ou non à se rapprocher de valeurs entières. Ici, cela est facile à étudier car le nombre $2+\sqrt{3}$ est un entier algébrique (racine de l'équation $x^2 - 4x + 1 = 0$) dont le conjugué $2-\sqrt{3}$ est strictement inférieur à 1 (le lecteur qui a traité l'exercice 1.12 du chapitre 1 aura reconnu un nombre de Pisot). En effet, on a pour tout n,

$$(2+\sqrt{3})^n + (2-\sqrt{3})^n = 2\sum_{0\leqslant 2k\leqslant n} C_n^{2k} 2^{n-2k}(\sqrt{3})^{2k} = 2\sum_{0\leqslant 2k\leqslant n} C_n^{2k} 2^{n-2k} 3^k,$$

qui est un entier. Ainsi, on a

$$|u_n| = \left|\sin\left(\pi(2+\sqrt{3})^n\right)\right| = \left|\sin\left(\pi(2-\sqrt{3})^n\right)\right| \leqslant \pi(2-\sqrt{3})^n.$$

Or, $\pi(2-\sqrt{3})^n$ est le terme d'une série géométrique convergente, donc la série $\sum u_n$ est absolument convergente. Il en résulte que la série $\sum \sin\left(\pi(2+\sqrt{3})^n\right)$ converge. ◁

3.28. Suite (a_n) telle $\sum a_n^k = 0$ pour tout entier k non nul

Soit $\sum a_n$ une série complexe absolument convergente. On suppose que pour tout entier $k \in \mathbb{N}^*$, $\sum_{n=0}^{+\infty} a_n^k = 0$. Montrer que la suite (a_n) est nulle.

(École normale supérieure)

▷ **Solution.**

On raisonne par l'absurde et on suppose que les termes de la suite (a_n) ne sont pas tous nuls. L'idée de la démonstration est de se ramener à une somme finie, seuls comptant les termes de module maximum. Posons

$M = \sup_{n\in\mathbb{N}} |a_n| > 0$. Cette borne supérieure est atteinte. En effet, la suite (a_n) converge vers 0, puisque la série $\sum a_n$ converge et il existe un entier naturel n_0 tel que $|a_n| \leqslant \dfrac{M}{2}$ pour $n \geqslant n_0$. Ceci entraîne que $M = \max_{0\leqslant n\leqslant n_0-1} |a_n|$. Quitte à diviser tous les termes de la suite par M, on se ramène à une suite dont tous les termes sont de module inférieur ou égal à 1, qui possède au moins un terme de module 1 et qui vérifie la même hypothèse que la suite initiale. On note que $I = \{n \in \mathbb{N}, |a_n| = 1\}$ est fini, puisque la série $\sum a_n$ converge.

Si $|a_n| < 1$, la limite de a_n^k quand k tend vers l'infini est nulle. Cela va nous permettre d'éliminer les termes a_n lorsque n n'est pas dans I. Plus précisément, montrons que $\lim_{k\to+\infty} s_k = 0$, où $s_k = \sum_{n\notin I} a_n^k$. Soit $\varepsilon > 0$ et $n_1 \in \mathbb{N}$ tel que $\sum_{n=n_1}^{+\infty} |a_n| \leqslant \varepsilon$. On a alors, pour $n \geqslant n_1$,

$$\begin{aligned} |s_k| &\leqslant \sum_{n\notin I} |a_n^k| \leqslant \sum_{\substack{n\notin I\\ n<n_1}} |a_n^k| + \sum_{\substack{n\notin I\\ n\geqslant n_1}} |a_n^k| \\ &\leqslant \sum_{\substack{n\notin I\\ n<n_1}} |a_n^k| + \sum_{n\geqslant n_1} |a_n| \leqslant \sum_{\substack{n\notin I\\ n<n_1}} |a_n^k| + \varepsilon. \end{aligned}$$

La somme finie $\sum_{\substack{n\notin I\\ n<n_1}} |a_n^k|$ tend vers 0 lorsque k tend vers $+\infty$, car les $|a_n|$ qui interviennent sont de module strictement inférieur à 1, et on aura donc $|s_k| \leqslant 2\varepsilon$ pour k assez grand. Ceci démontre que $\lim_{k\to+\infty} s_k = 0$ et donc, compte tenu de l'hypothèse, que $\lim_{k\to\infty} \sum_{n\in I} a_n^k = 0$. En notant $\lambda_1, \lambda_2, \ldots, \lambda_p$ les valeurs distinctes de a_n, pour $n \in I$, on obtient $\lim_{k\to+\infty} \sum_{i=1}^{p} n_i \lambda_i^k = 0$, où $n_1, n_2, \ldots, n_p$ sont des entiers naturels tous non nuls. La contradiction voulue résulte du lemme suivant.

Lemme. *Soit $p \in \mathbb{N}^*$, $\lambda_1, \lambda_2, \ldots, \lambda_p$ des nombres complexes distincts de module 1 et $\alpha_1, \alpha_2, \ldots, \alpha_p$ des nombres complexes. On suppose que $\lim_{k\to+\infty} \sum_{i=1}^{p} \alpha_i \lambda_i^k = 0$. On a alors $\alpha_1 = \alpha_2 = \cdots = \alpha_p = 0$.*

Démonstration. On raisonne par récurrence sur p, le résultat étant évident pour $p = 1$. On suppose que le résultat est vérifié au rang $p-1$. On considère $\lambda_1, \lambda_2, \ldots, \lambda_p$, de module 1, $\alpha_1, \alpha_2, \ldots, \alpha_p$ et, pour tout $k \in \mathbb{N}$, $u_k = \sum_{i=1}^{p} \alpha_i \lambda_i^k$, tels que $\lim_{k\to+\infty} u_k = 0$. On a alors, pour tout

$k \in \mathbb{N}$, $u_{k+1} - \lambda_p u_k = \sum_{i=1}^{p-1} \alpha_i(\lambda_i - \lambda_p)\lambda_i^k$ et $\lim_{k\to+\infty} u_{k+1} - \lambda_p u_k = 0$. De l'hypothèse de récurrence, on déduit, pour $1 \leqslant i \leqslant p-1$, que $\alpha_i(\lambda_i - \lambda_p) = 0$ et donc $\alpha_i = 0$, puisque les λ_i sont distincts. Le cas $p - 1$ donne alors $\alpha_p = 0$. ◁

3.29. Construction d'une série complexe

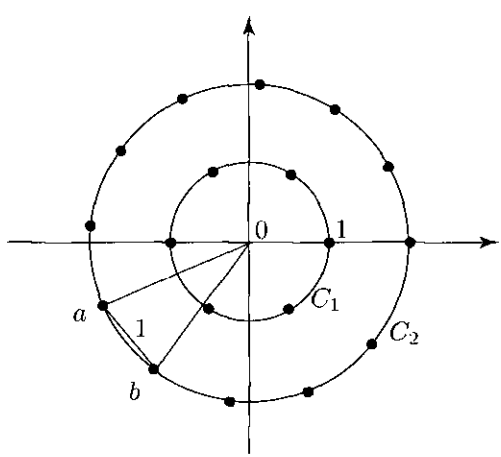
Construire une suite $(z_n) \in \mathbb{C}^{\mathbb{N}}$ telle que $\sum \frac{1}{|z_n|^2}$ diverge et vérifiant $|z_p - z_q| \geqslant 1$ pour tout couple (p, q) d'entiers distincts.

(École normale supérieure)

Solution.

On voit que les deux contraintes sont antagonistes : pour rendre la série divergente on souhaite choisir les z_n avec des « petits » modules, mais la seconde propriété ne permet que d'en prendre un nombre fini dans un compact donné. On est rapidement conduit à l'idée suivante : choisir les z_n sur les cercles C_k de centre 0 et de rayon k tout en respectant la seconde condition.

Soit $k \in \mathbb{N}^*$. Si a, b sont deux points de C_k tels que $|a-b| = 1$, l'angle en 0 du triangle $(0ab)$ vaut $2\arcsin\frac{1}{2k}$. Il en résulte qu'on peut choisir $n_k = \mathrm{E}\left(\frac{\pi}{\arcsin\frac{1}{2k}}\right)$ points sur C_k écartés deux à deux d'au moins 1.



On définit alors la suite (z_n) en prenant $z_1, z_2, \ldots, z_{n_1}$ sur C_1, puis $z_{n_1+1}, \ldots, z_{n_1+n_2}$ sur C_2, et ainsi de suite. La seconde contrainte est alors

respectée, et en sommant par tranches, on voit que la série $\sum \frac{1}{|z_n|^2}$ est de même nature que $\sum \frac{n_k}{k^2}$: en effet, pour n compris entre $n_1+\cdots+n_k+1$ et $n_1+\cdots+n_k+n_{k+1}$, $\frac{1}{|z_n|^2}$ vaut $\frac{1}{k^2}$. Comme $n_k \sim 2k\pi$, la série $\sum \frac{1}{|z_n|^2}$ est bien divergente. ◁

Le lecteur prouvera facilement qu'il est impossible de trouver une suite réelle vérifiant les mêmes hypothèses.

Lorsqu'une série n'est pas absolument convergente, elle peut être cependant convergente : on dit alors qu'elle est semi-convergente. *Pour étudier de telles séries, on peut envisager les techniques suivantes : un développement asymptotique du terme général, une application du théorème spécial des séries alternées, ou encore une transformation d'Abel.*

3.30. Développement asymptotique du terme général

Préciser la nature des séries suivantes :

1. $\sum (-1)^n \left(e - \left(1+\frac{1}{n}\right)^n\right)$;
2. $\sum \sin \sqrt{1+n^2\pi^2}$;
3. $\sum \sin(n!\pi e)$.

(École polytechnique)

▷ **Solution.**

1. Notons pour $n \geqslant 1$, $u_n = (-1)^n \left(e - \left(1+\frac{1}{n}\right)^n\right)$ et faisons-en un développement asymptotique. Pour n tendant vers $+\infty$,

$$\begin{aligned} u_n &= (-1)^n \left(e - e^{n\ln\left(1+\frac{1}{n}\right)}\right) \\ &= (-1)^n \left(e - e^{n\left(\frac{1}{n} - \frac{1}{2n^2} + O\left(\frac{1}{n^3}\right)\right)}\right) \\ &= (-1)^n e \left(1 - e^{-\frac{1}{2n} + O\left(\frac{1}{n^2}\right)}\right) \\ &= (-1)^n e \left(1 - \left(1 - \frac{1}{2n} + O\left(\frac{1}{n^2}\right)\right)\right) \\ &= \frac{(-1)^n e}{2n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

Le terme $O\left(\frac{1}{n^2}\right)$ est le terme général d'une série absolument convergente en vertu du théorème de comparaison des séries à termes positifs. Quant à $\frac{(-1)^n e}{2n}$, c'est le terme général d'une série semi-convergente

d'après le théorème spécial des séries alternées. La série $\sum u_n$ est donc semi-convergente.

Pour cette série, on pouvait directement appliquer le théorème spécial des séries alternées, car il est facile de voir que $\left(1+\frac{1}{n}\right)^n$ *converge vers* e *en croissant.*

2. Pour n tendant vers l'infini, on peut écrire

$$\begin{aligned}\sin\sqrt{1+n^2\pi^2} &= \sin\left(n\pi\sqrt{1+\frac{1}{n^2\pi^2}}\right)\\ &= \sin n\pi\left(1+\frac{1}{2n^2\pi^2}+O\left(\frac{1}{n^3}\right)\right)\\ &= \sin\left(n\pi+\frac{1}{2n\pi}+O\left(\frac{1}{n^2}\right)\right)\\ &= (-1)^n\sin\left(\frac{1}{2n\pi}+O\left(\frac{1}{n^2}\right)\right)\\ &= (-1)^n\left(\frac{1}{2n\pi}+O\left(\frac{1}{n^2}\right)\right)=\frac{(-1)^n}{2n\pi}+O\left(\frac{1}{n^2}\right).\end{aligned}$$

Le terme général de la série étudiée apparaît comme la somme du terme général d'une série alternée semi-convergente et du terme général d'une série absolument convergente : il en résulte que la série $\sum\sin\sqrt{1+n^2\pi^2}$ est semi-convergente.

3. Pour $n\geqslant 1$, on peut écrire

$$\begin{aligned}n!e &= n!\left(\sum_{k=0}^{+\infty}\frac{1}{k!}\right)\\ &= \underbrace{n!\left(\sum_{k=0}^{n}\frac{1}{k!}\right)}_{=p_n}+\frac{1}{n+1}+\frac{1}{(n+1)(n+2)}+\sum_{k=n+3}^{+\infty}\frac{n!}{k!}.\end{aligned}$$

Or, l'entier $p_n=n!+\cdots+n(n-1)+n+1$ est de même parité que $n+1$. D'autre part, on a pour $k\geqslant n+3$, $\dfrac{n!}{k!}\leqslant\dfrac{1}{k(k-1)(k-2)}\underset{k\to+\infty}{\sim}\dfrac{1}{k^3}$. La somme $\displaystyle\sum_{k=n+3}^{+\infty}\frac{1}{k(k-1)(k-2)}$ est donc équivalente à $\displaystyle\sum_{k=n+3}^{+\infty}\frac{1}{k^3}$, elle même équivalente à $\dfrac{1}{2n^2}$ (voir exercice 3.18). On en déduit que

$$\begin{aligned}n!e &= p_n+\frac{1}{n+1}+\frac{1}{(n+1)(n+2)}+O\left(\frac{1}{n^2}\right)\\ &= p_n+\frac{1}{n+1}+O\left(\frac{1}{n^2}\right)=p_n+\frac{1}{n}+O\left(\frac{1}{n^2}\right).\end{aligned}$$

Écrivons un développement asymptotique du terme général de la série étudiée

$$\sin(n!e\pi) = \sin\left(p_n\pi + \frac{\pi}{n} + O\left(\frac{1}{n^2}\right)\right)$$
$$= (-1)^{n+1}\sin\left(\frac{\pi}{n} + O\left(\frac{1}{n^2}\right)\right) = \frac{(-1)^{n+1}\pi}{n} + O\left(\frac{1}{n^2}\right).$$

On conclut comme dans les questions précédentes que la série $\sum\sin(n!\pi e)$ est semi-convergente. ◁

3.31. Autour des séries semi-convergentes

Soit $\mathcal{C}$ l'ensemble des séries convergentes à termes réels. Existe-t-il une série divergente à termes réels non nuls $\sum b_n$, telle que

$$\inf_{(a_n)\in\mathcal{C}} \sup_{n\geqslant 0}\left|1 - \frac{a_n}{b_n}\right| = 0 \ ?$$

(École polytechnique)

▷ **Solution.**

L'exercice demande de chercher une série divergente $\sum b_n$, telle que l'on puisse trouver des séries convergentes $\sum a_n$ avec la suite (a_n) aussi « proche » qu'on veut de la suite (b_n), cette « proximité » étant mesurée par $\sup_{n\geqslant 0}\left|1 - \frac{a_n}{b_n}\right|$.

Montrons pour commencer qu'une série divergente $\sum b_n$ à termes strictement positifs (par exemple $\sum \frac{1}{n}$) ne peut pas convenir. En effet, soit $(a_n) \in \mathbb{R}^{\mathbb{N}}$ telle que $\sup_{n\geqslant 0}\left|1 - \frac{a_n}{b_n}\right| \leqslant \frac{1}{2}$. Alors pour $n \in \mathbb{N}$, $1 - \frac{a_n}{b_n} \leqslant \left|1 - \frac{a_n}{b_n}\right| \leqslant \frac{1}{2}$, de sorte que $a_n \geqslant \frac{1}{2}b_n$. La série $\sum a_n$ est alors divergente.

Le lecteur sait que le théorème sur les séries à termes généraux équivalents est faux si l'une des séries n'est pas supposée de signe constant (au moins à partir d'un certain rang). Prenons donc une suite b_n telle que $\sum b_n$ diverge et pour laquelle il existe une suite c_n équivalente à b_n mais telle que $\sum c_n$ converge. De tels exemples abondent : on peut prendre $b_n = \frac{(-1)^n}{\sqrt{n}} + \frac{1}{n}$ et $c_n = \frac{(-1)^n}{\sqrt{n}}$.

Soit $\varepsilon > 0$. Il existe un rang N tel que pour $n \geqslant \mathrm{N}$, $\left|1 - \frac{c_n}{b_n}\right| \leqslant \varepsilon$. On

prend alors la suite (a_n) définie par : $a_n = b_n$ pour $n < \mathrm{N}$ et $a_n = c_n$ pour $n \geqslant \mathrm{N}$. Il est clair que $\sum a_n$ converge et que $\sup_{n\geqslant 0}\left|1 - \dfrac{a_n}{b_n}\right| \leqslant \varepsilon$. Cette construction étant possible pour tout $\varepsilon > 0$, on a bien la propriété souhaitée. ◁

Le théorème spécial des séries alternées fournit un critère simple et efficace pour prouver la convergence de nombreuses séries : si $(a_n)_{n\in\mathbb{N}}$ est une suite décroissante de réels positifs tendant vers 0, la série alternée $\sum(-1)^n a_n$ converge.

3.32. Calcul de la somme d'une série alternée

1. Montrer que la suite $u_n = \sum_{p=1}^{n} \dfrac{\ln p}{p} - \dfrac{1}{2}\ln^2 n$ converge.

2. En déduire la valeur de la somme $\sum_{n=1}^{+\infty}(-1)^n \dfrac{\ln n}{n}$.

(École polytechnique)

▷ **Solution.**

1. La fonction $t \longmapsto \dfrac{\ln^2 t}{2}$ est une primitive de $t \longmapsto \dfrac{\ln t}{t}$. La fonction $t \longmapsto \dfrac{\ln t}{t}$ décroît sur $[3, +\infty[$, ce qui nous oriente vers la comparaison série-intégrale. On peut écrire pour $n \geqslant 3$,

$$\begin{aligned} u_{n+1} - u_n &= \frac{\ln(n+1)}{n+1} - \frac{1}{2}(\ln^2(n+1) - \ln^2 n) \\ &= \frac{\ln(n+1)}{n+1} - \int_n^{n+1} \frac{\ln t}{t}\mathrm{d}t \leqslant 0. \end{aligned}$$

La suite (u_n) décroît à partir du rang 3. Par ailleurs, on a pour $n \geqslant 3$,

$$\sum_{p=3}^{n} \frac{\ln p}{p} \geqslant \sum_{p=3}^{n} \int_p^{p+1} \frac{\ln t}{t}\mathrm{d}t = \int_3^{n+1} \frac{\ln t}{t}\mathrm{d}t = \frac{1}{2}(\ln^2(n+1) - \ln^2 3).$$

On en déduit que

$$u_n \geqslant \frac{\ln 2}{2} + \frac{1}{2}(\ln^2(n+1) - \ln^2 n - \ln^2 3) \geqslant \frac{1}{2}(\ln 2 - \ln^2 3).$$

La suite (u_n) est décroissante (pour $n \geqslant 3$) et minorée. Elle converge. On notera l sa limite.

2. La série de terme général $(-1)^n \dfrac{\ln n}{n}$ converge car elle vérifie le critère spécial de convergence des séries alternée. En effet la suite de terme général $\dfrac{\ln n}{n}$ est décroissante pour $n \geqslant 3$ et tend vers 0.

Pour calculer sa somme, on part d'une somme partielle qu'on transforme astucieusement pour faire apparaître la suite de la question précédente. On peut écrire, pour $N \in \mathbb{N}^*$,

$$\begin{aligned}\sum_{n=1}^{2N}(-1)^n\frac{\ln n}{n} &= 2\sum_{n=1}^{N}\frac{\ln(2n)}{2n} - \sum_{n=1}^{2N}\frac{\ln n}{n}\\ &= \ln 2\left(\sum_{n=1}^{N}\frac{1}{n}\right) + \sum_{n=1}^{N}\frac{\ln n}{n} - \sum_{n=1}^{2N}\frac{\ln n}{n}\\ &= \ln 2\left(\sum_{n=1}^{N}\frac{1}{n}\right) + u_N + \frac{1}{2}\ln^2 N - u_{2N} - \frac{1}{2}\ln^2(2N).\end{aligned}$$

En remarquant que

$$\frac{1}{2}\ln^2(2N) - \frac{1}{2}\ln^2 N = \frac{1}{2}(\ln(2N) - \ln N)(\ln(2N) + \ln N) = \frac{\ln^2 2}{2} + \ln 2\,\ln N,$$

on peut simplifier cette expression :

$$\sum_{n=1}^{2N}(-1)^n\frac{\ln n}{n} = \ln 2\left(\sum_{n=1}^{N}\frac{1}{n} - \ln N\right) + u_N - u_{2N} - \frac{\ln^2 2}{2}.$$

Nous savons que $\lim\limits_{N\to+\infty} u_N - u_{2N} = l - l = 0$. La limite de $\sum\limits_{n=1}^{N}\frac{1}{n} - \ln N$, lorsque N tend vers l'infini, est la constante d'Euler γ. On obtient finalement

$$\boxed{\sum_{n=1}^{+\infty}(-1)^n\frac{\ln n}{n} = \gamma\ln 2 - \frac{\ln^2 2}{2}.} \quad \triangleleft$$

Il est également important de retenir que, lorsque $(a_n)_{n\in\mathbb{N}}$ est une suite décroissante de réels, tendant vers 0, on a une majoration simple des restes de la série alternée $\sum(-1)^n a_n$: pour tout $N \geqslant 0$,

$$\left|\sum_{n=N}^{+\infty}(-1)^n a_n\right| \leqslant a_N.$$

3.33. Transformation d'Euler

Pour toute suite $(u_n)_{n\in\mathbb{N}}$, on définit la suite $(\Delta u_n)_{n\in\mathbb{N}}$ par $\Delta u_n = u_n - u_{n+1}$. Soit $(u_n)_{n\in\mathbb{N}}$ une suite de réels positifs ou nuls, décroissante et convergente vers 0. On pose $S = \sum\limits_{n=0}^{+\infty}(-1)^n u_n$.

1. Montrer que, pour tout $p \geqslant 1$,

$$S = \frac{1}{2}\left(\sum_{k=0}^{p-1} \frac{1}{2^k}\Delta^k u_0\right) + \frac{1}{2^p}\sum_{n=0}^{+\infty}(-1)^n \Delta^p u_n.$$

2. Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ de classe $\mathcal{C}^\infty$ telle que, pour tout $p \in \mathbb{N}$ et tout $x \geqslant 0$, $(-1)^p f^{(p)}(x) \geqslant 0$. On suppose que $\lim\limits_{n\to+\infty} f(n) = 0$. Montrer que

$$\sum_{n=0}^{+\infty}(-1)^n f(n) = \frac{1}{2}f(0) + \frac{1}{4}\Delta f(0) + \cdots + \frac{1}{2^p}\Delta^{p-1} f(0) + R_p$$

avec $|R_p| \leqslant \dfrac{1}{2^p}|\Delta^p f(0)|$. La série $\sum\left(\dfrac{1}{2}\right)^{k+1}\Delta^k f(0)$ est appelée *transformée d'Euler* de la série $\sum(-1)^n f(n)$.

(**École polytechnique**)

Solution.

1. La série $\sum(-1)^n u_n$ converge d'après le théorème spécial des séries alternées. Pour $N \in \mathbb{N}^*$, on a

$$\begin{aligned}\sum_{n=1}^{N}(-1)^{n-1}\Delta u_{n-1} &= \sum_{n=1}^{N}(-1)^{n-1}u_{n-1} - \sum_{n=1}^{N}(-1)^{n-1}u_n \\ &= \sum_{n=0}^{N-1}(-1)^n u_n + \sum_{n=1}^{N}(-1)^n u_n.\end{aligned}$$

On en déduit que la série $\sum(-1)^{n-1}\Delta u_{n-1}$ converge et que $\sum\limits_{n=1}^{+\infty}(-1)^{n-1}\Delta u_{n-1} = 2S - u_0$. Ceci peut s'écrire

$$S = \frac{1}{2}u_0 + \frac{1}{2}\sum_{n=0}^{+\infty}(-1)^n \Delta u_n,$$

ce qui est l'égalité voulue pour $p = 1$. On va raisonner par récurrence sur p et noter S_p la quantité $\dfrac{1}{2}\left(\sum\limits_{k=0}^{p-1}\dfrac{1}{2^k}\Delta^k u_0\right) + \dfrac{1}{2^p}\sum\limits_{n=0}^{+\infty}(-1)^n\Delta^p u_n$ lorsque celle-ci existe. On vient de prouver l'existence de S_1, qui est égal à S. Si $\sum(-1)^n\Delta^p u_n$ converge et si $S_p = S$, alors pour $N \in \mathbb{N}^*$, on peut écrire

$$\sum_{n=0}^{N}(-1)^n\Delta^{p+1}u_n = \sum_{n=0}^{N}(-1)^n(\Delta^p u_n - \Delta^p u_{n+1})$$
$$= \sum_{n=0}^{N}(-1)^n\Delta^p u_n + \sum_{n=1}^{N+1}(-1)^n\Delta^p u_n.$$

On en déduit que $\sum(-1)^n\Delta^{p+1}u_n$ converge et que

$$\sum_{n=0}^{+\infty}(-1)^n\Delta^{p+1}u_n = 2\sum_{n=0}^{+\infty}(-1)^n\Delta^p u_n - \Delta^p u_0.$$

On a donc

$$S_{p+1} = \frac{1}{2}\left(\sum_{k=0}^{p}\frac{1}{2^k}\Delta^k u_0\right) + \frac{1}{2^p}\sum_{n=0}^{+\infty}(-1)^n\Delta^p u_n - \frac{1}{2^{p+1}}\Delta^p u_0$$
$$= \frac{1}{2}\left(\sum_{k=0}^{p-1}\frac{1}{2^k}\Delta^k u_0\right) + \frac{1}{2^p}\sum_{n=0}^{+\infty}(-1)^n\Delta^p u_n = S_p = S.$$

2. On utilise la question 1. On pose $u_n = f(n)$. Par hypothèse, les fonctions f et $-f'$ sont positives sur $\mathbb{R}_+$. La fonction f est donc positive et décroissante et il en est de même de la suite (u_n). Enfin, par hypothèse, la suite (u_n) converge vers 0. Le résultat de la question précédente s'applique et on a, pour $p \geqslant 1$,

$$\sum_{n=0}^{+\infty}(-1)^n f(n) = \sum_{k=0}^{p-1}\left(\frac{1}{2}\right)^{k+1}\Delta^k f(0) + \frac{1}{2^p}\sum_{n=0}^{+\infty}(-1)^n\Delta^p f(n),$$

ce qui est résultat demandé avec $R_p = \dfrac{1}{2^p}\displaystyle\sum_{n=0}^{+\infty}(-1)^n\Delta^p f(n)$. Il faut démontrer, pour conclure, que $\left|\displaystyle\sum_{n=0}^{+\infty}(-1)^n\Delta^p f(n)\right| \leqslant |\Delta^p f(0)|$. Il s'agit en fait de majorer la somme de la série par son premier terme, propriété des séries alternées vérifiant le critère spécial de convergence.

Montrons donc que la série $\sum(-1)^n\Delta^p f(n)$ vérifie le critère spécial de convergence des séries alternées. Pour cela, on définit, pour toute fonction $f : \mathbb{R}_+ \to \mathbb{R}$, la fonction Δf par $\Delta f(x) = f(x) - f(x+1)$ et on montre que, pour $p \in \mathbb{N}$, la fonction $\Delta^p f$ est positive sur $\mathbb{R}_+$. On raisonne par récurrence sur p et on utilise le fait que $-f'$ vérifie les mêmes hypothèses que f. La propriété est vraie pour $p = 0$, par hypothèse. Si on suppose qu'elle est réalisée au rang p pour toute fonction $\mathcal{C}^\infty$ telle que $(-1)^n f^{(n)} \geqslant 0$ pour tout $n \in \mathbb{N}$, alors c'est vrai pour $-f'$. On a donc $\Delta^p(-f') \geqslant 0$ sur $\mathbb{R}_+$. Mais il est clair que $\Delta^p(-f') = -(\Delta^p f)'$. La fonction $\Delta^p f$ est donc décroissante sur $\mathbb{R}_+$ et par définition, $\Delta^{p+1}f$ est positive. Cela termine la récurrence.

La suite $(\Delta^p f(n))_{n\in\mathbb{N}}$ est donc à termes positifs. Il résulte de la démonstration précédente qu'on a aussi $\Delta^p f$ décroissante pour $p \in \mathbb{N}$. Ainsi, la suite $(\Delta^p f(n))_{n\in\mathbb{N}}$ décroît. Enfin, son terme général tend vers 0, puisqu'il résulte de ce qui précède que $\sum(-1)^n \Delta^p f(n)$ converge. On a donc l'inégalité voulue.

On peut aussi noter que la suite $(\Delta^p f(0))_{p\in\mathbb{N}^*}$ est décroissante, car $\Delta^{p+1} f(0) = \Delta^p f(0) - \Delta^p f(1) \leqslant \Delta^p f(0)$. On a donc $|\mathrm{R}_p| \leqslant \frac{1}{2^p}|f(0)|$, d'où l'on tire $\lim\limits_{p\to+\infty} \mathrm{R}_p = 0$ et

$$\sum_{n=0}^{+\infty}(-1)^n f(n) = \sum_{k=0}^{+\infty}\left(\frac{1}{2}\right)^{k+1}\Delta^k f(0). \quad \triangleleft$$

La transformation d'Euler est une méthode classique d'accélération de convergence des séries alternées. Si on considère par exemple la série harmonique alternée, ce qui revient à prendre $f : x \longmapsto \dfrac{1}{x+1}$, qui a les propriétés voulues, $\displaystyle\sum_{k=0}^{29}\left(\frac{1}{2}\right)^{k+1}\Delta^k f(0)$ $(p = 30)$ donne une valeur approchée de $\ln 2$ à 10^{-10} près, alors qu'il faut additionner un milliard de termes de la série initiale pour arriver au même résultat.

Dans l'exercice suivant, on envisage un regroupement de termes par paquets pour déterminer la nature d'une série. C'est plutôt délicat car la série n'est pas absolument convergente.

3.34. Sommation par paquets

1. Montrer que l'équation (E) : $x^3 - x - 1 = 0$ admet une racine réelle α et deux racines complexes conjuguées β et γ. Montrer que pour tout $n \geqslant 0$, $\mathrm{S}_n = \alpha^n + \beta^n + \gamma^n$ est entier.

2. Étudier les séries $\displaystyle\sum \frac{\sin\left(\alpha^n \frac{\pi}{2}\right)}{n}$ et $\displaystyle\sum \frac{\cos\left(\alpha^n \frac{\pi}{2}\right)}{n}$.

(École polytechnique)

▷ **Solution.**

1. L'étude de la fonction $x \longmapsto x^3 - x - 1$ montre que le polynôme $\mathrm{X}^3 - \mathrm{X} - 1$ possède une seule racine réelle α et que $\alpha > 1$. Les deux autres racines β et γ sont donc complexes conjuguées. De plus, on a $|\beta| = |\gamma| < 1$, puisque $\alpha\beta\gamma = 1$.

Si x est une racine de $\mathrm{X}^3 - \mathrm{X} - 1$, on a, pour $n \in \mathbb{N}$, $x^{n+3} = x^{n+1} + x^n$. On en déduit que, pour tout $n \in \mathbb{N}$, $\mathrm{S}_{n+3} = \mathrm{S}_{n+1} + \mathrm{S}_n$. Sachant que

$S_0 = 3$, $S_1 = 0$ et $S_2 = S_1^2 - 2(\alpha\beta + \beta\gamma + \gamma\alpha) = 0 - 2(-1) = 2$, on en déduit, par une récurrence immédiate, que S_n est entier, pour tout $n \in \mathbb{N}$.

Le lecteur aura sans doute reconnu les sommes de Newton relatives au polynôme à coefficients entiers $X^3 - X - 1$. La relation de récurrence obtenue ici est un cas particulier des célèbres formules de Newton (démontrées dans le cas général dans le tome 1 d'algèbre, exercice 5.26). On peut aussi remarquer que α est un nombre de Pisot (voir l'exercice 1.12).

2. On a $S_n \underset{n\to\infty}{\sim} \alpha^n$, d'où l'idée de comparer les séries proposées à $\sum \dfrac{\sin\left(S_n \dfrac{\pi}{2}\right)}{n}$ et $\sum \dfrac{\cos\left(S_n \dfrac{\pi}{2}\right)}{n}$. L'intérêt de remplacer α^n par S_n provient de ce que S_n est entier.

• On a, pour tout $n \in \mathbb{N}^*$,

$$\begin{aligned} u_n &= \frac{1}{n}\left(\sin\left(S_n\frac{\pi}{2}\right) - \sin\left(\alpha^n\frac{\pi}{2}\right)\right) \\ &= \frac{2}{n}\sin\left((\beta^n+\gamma^n)\frac{\pi}{4}\right)\cos\left((S_n+\alpha^n)\frac{\pi}{4}\right), \end{aligned}$$

$$\text{et}\quad |u_n| \leqslant \frac{2}{n}\left|\sin\left((\beta^n+\gamma^n)\frac{\pi}{4}\right)\right| \leqslant \frac{\pi}{2n}(|\beta|^n + |\gamma|^n).$$

La série $\sum u_n$ est donc convergente puisque $|\beta| = |\gamma| < 1$. Les séries $\sum \dfrac{\sin\left(\alpha^n \dfrac{\pi}{2}\right)}{n}$ et $\sum \dfrac{\sin\left(S_n \dfrac{\pi}{2}\right)}{n}$ sont donc de même nature. On montre de la même manière que les séries $\sum \dfrac{\cos\left(\alpha^n \dfrac{\pi}{2}\right)}{n}$ et $\sum \dfrac{\cos\left(S_n \dfrac{\pi}{2}\right)}{n}$ sont de même nature.

• Les valeurs de $\sin\left(S_n \dfrac{\pi}{2}\right)$ et $\sin\left(S_n \dfrac{\pi}{2}\right)$ ne dépendent que du reste r_n de S_n modulo 4. Calculons r_n pour les premières valeurs de n, en utilisant la relation de récurrence $S_{n+3} = S_{n+1} + S_n$. On obtient : $r_0 = 3$, $r_1 = 0$, $r_2 = 2$, $r_3 = 3$, $r_4 = 2$, $r_5 = 1$, $r_6 = 1$, $r_7 = 3$, $r_8 = 2$, $r_9 = 0$, $r_{10} = 1$, $r_{11} = 2$, $r_{12} = 1$, $r_{13} = 3$, $r_{14} = 3$, $r_{15} = 0$, $r_{16} = 2$.

On remarque que : $r_{n+14} = r_n$ pour $0 \leqslant n \leqslant 2$. Une récurrence immédiate permet de démontrer que, pour tout $n \in \mathbb{N}$, on a $r_{n+14} = r_n$: la suite (r_n) est 14-périodique.

• On va regrouper les termes par paquets de 14 termes et se ramener à la nature de la série $\sum a_n$ avec $a_n = \sum_{k=0}^{13} \dfrac{\sin\left(S_{14n+k} \dfrac{\pi}{2}\right)}{14n+k}$. Soit $N \geqslant 1$ et $n_0 = E(N/14)$. On peut écrire

$$\sum_{n=14}^{N} \frac{\sin\left(S_n \frac{\pi}{2}\right)}{n} = \left(\sum_{n=1}^{n_0} a_n\right) - \delta_N \text{ avec } \delta_N = \sum_{n=N+1}^{14n_0+13} \frac{\sin\left(S_n \frac{\pi}{2}\right)}{n}.$$

Or, la suite δ_N converge vers 0 puisque

$$|\delta_N| \leqslant \sum_{n=14n_0+1}^{14n_0+13} \frac{1}{n} \leqslant \frac{13}{14n_0} \xrightarrow[N\to+\infty]{} 0.$$

On en déduit que les séries $\sum \frac{\sin\left(S_n \frac{\pi}{2}\right)}{n}$ et $\sum a_n$ sont de même nature.

Des calculs précédents, on tire

$$a_n = -\frac{1}{14n} - \frac{1}{14n+3} + \frac{1}{14n+5} + \frac{1}{14n+6} - \frac{1}{14n+7} + \frac{1}{14n+10} + \frac{1}{14n+12} - \frac{1}{14n+13}.$$

On a, pour tout $k \in [\![0, 13]\!]$, quand n tend vers $+\infty$,

$$\frac{1}{14n+k} = \frac{1}{14n}\left(1 + \frac{k}{14n}\right)^{-1} = \frac{1}{14n} + O\left(\frac{1}{n^2}\right).$$

On en déduit que $a_n = O\left(\frac{1}{n^2}\right)$, car dans la somme définissant a_n, il y autant de signes $+$ que de signes $-$. Il en résulte que la série $\sum \frac{\sin\left(a^n \frac{\pi}{2}\right)}{n}$ converge. Le même raisonnement montre que la série de terme général $\frac{\cos\left(S_n \frac{\pi}{2}\right)}{n}$ a même nature que la série de terme général $b_n = \sum_{i=0}^{13} \frac{\cos\left(S_{14n+i} \frac{\pi}{2}\right)}{14n+i}$. Cette fois-ci, on trouve

$$\begin{aligned} b_n &= \frac{1}{14n+1} - \frac{1}{14n+2} - \frac{1}{14n+4} - \frac{1}{14n+8} + \frac{1}{14n+9} - \frac{1}{14n+11} \\ &= -\frac{1}{7n} + O\left(\frac{1}{n^2}\right). \end{aligned}$$

On en déduit la divergence de la série $\sum \frac{\cos\left(a^n \frac{\pi}{2}\right)}{n}$. ◁

Les exercices suivants exploitent la technique des transformations d'Abel, souvent présentée comme une intégration par parties discrète. Plus précisément, considérons deux suites $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ et des entiers $M \leqslant N$. Notons pour $n \geqslant 0$, $A_n = \sum_{k=0}^{n} a_k$. On peut alors écrire $a_n = A_n - A_{n-1}$ (avec la convention $A_{-1} = 0$) et

$$\sum_{n=M}^{N} a_n b_n = \sum_{n=M}^{N} (A_n - A_{n-1}) b_n = \sum_{n=M}^{N} A_n b_n - \sum_{n=M}^{N} A_{n-1} b_n$$
$$= \sum_{n=M}^{N} A_n b_n - \sum_{n=M-1}^{N-1} A_n b_{n+1}$$
$$= A_N b_{N+1} - A_{M-1} b_M - \sum_{n=M}^{N} A_n (b_{n+1} - b_n).$$

La suite A_n joue le rôle de la «primitive» de a_n et $b_{n+1} - b_n$ celui de la dérivée de b_n. Lorsque la série $\sum a_n$ converge, on peut écrire $a_n = R_n - R_{n+1}$ avec $R_n = \sum_{k=n}^{+\infty} a_n$. On arrive alors à

$$\sum_{n=M}^{N} a_n b_n = R_M b_{M-1} - R_{N+1} B_N + \sum_{n=M}^{N} R_n (b_n - b_{n-1}).$$

Cette transformation fut utilisée par Abel en 1826 pour donner un exemple de série de fonctions continues dont la somme n'est pas continue[4]*, à savoir $\sum \dfrac{\sin nx}{n}$.*

3.35. Exemple de transformation d'Abel

Soit $(a_n)_{n\geqslant 0}$ une suite de réels positifs qui est croissante et tend vers $+\infty$. Soit $(x_n)_{n\geqslant 0}$ une suite de nombres complexes telle que $\sum_{n=0}^{+\infty} \dfrac{x_n}{a_n} = l$. Montrer que $\dfrac{1}{a_n} \sum_{k=0}^{n} x_k$ tend vers 0 lorsque n tend vers l'infini.

(École normale supérieure)

▷ **Solution.**

Considérons le reste de la série $\sum \dfrac{x_n}{a_n}$, $R_n = \sum_{k=n}^{+\infty} \dfrac{x_n}{a_n}$ défini pour $n \geqslant 0$. Exprimons x_n en fonction des R_n. On a $R_n - R_{n+1} = \dfrac{x_n}{a_n}$ ce qui donne $x_n = a_n(R_n - R_{n+1})$. Posons pour n entier naturel, $U_n = \dfrac{1}{a_n} \sum_{k=0}^{n} x_k$ et opérons une transformation d'Abel sur le numérateur :

4. Dans les travaux de Cauchy, on pouvait lire à l'époque que la somme d'une série de fonctions continues est encore continue.

$$U_n = \frac{\sum_{k=0}^{n} a_k(R_k - R_{k+1})}{a_n} = \frac{\sum_{k=0}^{n} a_k R_k - \sum_{k=1}^{n+1} a_{k-1} R_k}{a_n}$$

$$= \frac{a_0 R_0 - a_n R_{n+1} + \sum_{k=1}^{n} (a_k - a_{k-1}) R_k}{a_n}$$

$$= \frac{a_0 R_0}{a_n} - R_{n+1} + \frac{\sum_{k=1}^{n} (a_k - a_{k-1}) R_k}{a_n}.$$

Pour montrer que U_n tend vers 0, il suffit de prouver que le quotient

$$\frac{\sum_{k=1}^{n} (a_k - a_{k-1}) R_k}{a_n}$$

tend vers 0. Soit $\varepsilon > 0$ et un entier n_0 tel que si $n \geqslant n_0$, $|R_n| \leqslant \varepsilon$. On majore le quotient précédent pour $n \geqslant n_0$ en coupant la somme en deux :

$$\left| \frac{\sum_{k=1}^{n} (a_k - a_{k-1}) R_k}{a_n} \right| \leqslant \frac{\sum_{k=1}^{n_0} (a_k - a_{k-1}) |R_k|}{a_n} + \frac{\sum_{k=n_0+1}^{n} (a_k - a_{k-1}) \varepsilon}{a_n}$$

$$\leqslant \frac{\sum_{k=1}^{n_0} (a_k - a_{k-1}) |R_k|}{a_n} + \frac{a_n - a_{n_0}}{a_n} \varepsilon \xrightarrow[n \to +\infty]{} \varepsilon.$$

Donc pour n assez grand, $\left| \frac{\sum_{k=1}^{n} (a_k - a_{k-1}) R_k}{a_n} \right| \leqslant 2\varepsilon$. On conclut que U_n converge vers 0. ◁

3.36. Étude asymptotique à l'aide de transformations d'Abel

Soit $\sum a_n$ une série de réels convergente.

1. Montrer que pour n tendant vers l'infini, $\sum_{k=1}^{n} k a_k = o(n)$.

On suppose de plus qu'il existe $c > 0$ tel que pour tout $n \geqslant 1$, $na_n > -c$.

2. Montrer que pour n tendant vers l'infini,

$$\sum_{k=1}^{n} k|a_k| \leqslant 2cn + o(n).$$

3. On pose, pour $n \geqslant 1$, $t_n = \sum_{k=n}^{+\infty} \frac{|a_k|}{k}$. Justifier l'existence de t_n et montrer que pour n tendant vers l'infini, $t_n = O\left(\frac{1}{n}\right)$.

(École polytechnique)

▷ **Solution.**

1. Posons pour $n \geqslant 1$, $\mathrm{R}_n = \sum_{k=n}^{+\infty} a_k$. On a $a_n = \mathrm{R}_n - \mathrm{R}_{n+1}$. Effectuons une transformation d'Abel :

$$\begin{aligned}\sum_{k=1}^{n} ka_k &= \sum_{k=1}^{n} k(\mathrm{R}_k - \mathrm{R}_{k+1}) = \sum_{k=1}^{n} k\mathrm{R}_k - \sum_{k=1}^{n} k\mathrm{R}_{k+1} \\ &= \sum_{k=1}^{n} k\mathrm{R}_k - \sum_{k=2}^{n+1} (k-1)\mathrm{R}_k = \sum_{k=1}^{n} \mathrm{R}_k - n\mathrm{R}_{n+1}.\end{aligned}$$

Par conséquent, on a

$$\frac{1}{n}\sum_{k=1}^{n} ka_k = \frac{1}{n}\sum_{k=1}^{n} \mathrm{R}_k - \mathrm{R}_{n+1} \xrightarrow[n\to\infty]{} 0 - 0 = 0,$$

en vertu du théorème de Cesàro (car $\lim_{n\to+\infty} \mathrm{R}_n = 0$). On conclut que

$$\sum_{k=1}^{n} ka_k = o(n).$$

2. Pour $n \geqslant 1$, on a

$$\begin{aligned}\sum_{k=1}^{n} k|a_k| &= \sum_{k=1}^{n} k(|a_k| - a_k) + \sum_{k=1}^{n} ka_k = \sum_{k=1}^{n} k(|a_k| - a_k) + o(n) \\ &\leqslant \sum_{k=1}^{n} 2c + o(n) \quad \text{car } k(|a_k| - a_k) = \begin{cases} -2ka_k \leqslant 2c & \text{si } a_k < 0 \\ 0 \leqslant 2c & \text{si } a_k \geqslant 0 \end{cases} \\ &\leqslant 2cn + o(n).\end{aligned}$$

3. Posons pour $n \geqslant 1$, $\mathrm{A}_n = \sum_{k=1}^{n} k|a_k|$. On a $n|a_n| = \mathrm{A}_n - \mathrm{A}_{n-1}$. Prenons $\mathrm{N} \geqslant n$ et lançons-nous dans une nouvelle transformation d'Abel :

$$\begin{aligned}
\sum_{k=n}^{\mathrm{N}} \frac{|a_k|}{k} &= \sum_{k=n}^{\mathrm{N}} \frac{\mathrm{A}_k - \mathrm{A}_{k-1}}{k^2} = \sum_{k=n}^{\mathrm{N}} \frac{\mathrm{A}_k}{k^2} - \sum_{k=n}^{\mathrm{N}} \frac{\mathrm{A}_{k-1}}{k^2} \\
&= \sum_{k=n}^{\mathrm{N}} \frac{\mathrm{A}_k}{k^2} - \sum_{k=n-1}^{\mathrm{N}-1} \frac{\mathrm{A}_k}{(k+1)^2} \\
&= \sum_{k=n}^{\mathrm{N}} \mathrm{A}_k \left(\frac{1}{k^2} - \frac{1}{(k+1)^2} \right) - \frac{\mathrm{A}_{n-1}}{n^2} + \frac{\mathrm{A}_{\mathrm{N}}}{(\mathrm{N}+1)^2} \\
&= \sum_{k=n}^{\mathrm{N}} \frac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2} - \frac{\mathrm{A}_{n-1}}{n^2} + \frac{\mathrm{A}_{\mathrm{N}}}{(\mathrm{N}+1)^2} \\
&= \sum_{k=n}^{\mathrm{N}} \frac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2} - \frac{\mathrm{A}_{n-1}}{n^2} + O\left(\frac{1}{\mathrm{N}}\right) \text{ d'après 2.}
\end{aligned}$$

Comme $\dfrac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2} = O\left(\dfrac{k\mathrm{A}_k}{k^4}\right) = O\left(\dfrac{1}{k^2}\right)$ pour $k \geqslant 1$, la série $\sum \dfrac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2}$ est absolument convergente. La somme $\sum_{k=n}^{\mathrm{N}} \dfrac{|a_k|}{k}$ admet donc une limite lorsque N tend vers l'infini. Ainsi t_n existe et on a

$$t_n = \sum_{k=n}^{+\infty} \frac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2} - \frac{\mathrm{A}_{n-1}}{n^2}.$$

On a, lorsque n tend vers l'infini, $\dfrac{\mathrm{A}_{n-1}}{n^2} = O\left(\dfrac{n}{n^2}\right) = O\left(\dfrac{1}{n}\right)$. De plus, il existe $\mathrm{M} \geqslant 0$ tel que si $k \geqslant 1$, $\dfrac{(2k+1)\mathrm{A}_k}{k^2(k+1)^2} \leqslant \dfrac{\mathrm{M}}{k^2}$. Par conséquent,

$$t_n \leqslant \mathrm{M} \sum_{k=n}^{+\infty} \frac{1}{k^2} + O\left(\frac{1}{n}\right).$$

Classiquement, pour n tendant vers l'infini, $\sum_{k=n}^{+\infty} \dfrac{1}{k^2} \sim \dfrac{1}{n}$ (voir exercice 3.18) et finalement, $t_n = O\left(\dfrac{1}{n}\right)$. ◁

3.37. Séries de Hardy

Déterminer la nature de la série $\sum \dfrac{\sin(\pi\sqrt{n})}{n^\alpha}$ en traitant successivement les cas suivants :

1. $\alpha > 1$;

2. $\dfrac{1}{2} < \alpha \leqslant 1$;

3. $\alpha = \dfrac{1}{2}$ (on effectuera un développement asymptotique de la différence $e^{i\pi\sqrt{n+1}} - e^{i\pi\sqrt{n}}$) ;

4. $\alpha < \dfrac{1}{2}$.

(École normale supérieure)

▷ **Solution.**

1. Pour $\alpha > 1$, le théorème de comparaison des séries à termes positifs assure la convergence absolue de la série.

2. Nous allons montrer qu'il y a convergence de la série lorsque $\alpha > \dfrac{1}{2}$ en faisant une comparaison série-intégrale. Étudions l'intégrale

$$\int_1^{+\infty} \frac{\sin \pi\sqrt{t}}{t^\alpha}\mathrm{d}t = \lim_{x\to+\infty} \int_1^x \frac{\sin \pi\sqrt{t}}{t^\alpha}\mathrm{d}t.$$

Pour $x \geqslant 1$, en faisant le changement de variables $u = \sqrt{t}$, il vient

$$\begin{aligned}\int_1^x \frac{\sin \pi\sqrt{t}}{t^\alpha}\mathrm{d}t &= 2\int_1^{\sqrt{x}} \frac{\sin \pi u}{u^{2\alpha-1}}\mathrm{d}u \\ &= 2\left[-\frac{\cos \pi u}{\pi u^{2\alpha-1}}\right]_1^{\sqrt{x}} - \frac{2(2\alpha-1)}{\pi}\int_1^{\sqrt{x}} \frac{\cos \pi u}{u^{2\alpha}}\mathrm{d}u \\ &\xrightarrow[x\to+\infty]{} -\frac{2}{\pi} - \frac{2(2\alpha-1)}{\pi}\int_{[1,+\infty[} \frac{\cos \pi u}{u^{2\alpha}}\mathrm{d}u,\end{aligned}$$

car la fonction $u \longmapsto \dfrac{\cos \pi u}{u^{2\alpha}}$ est intégrable sur $[1, +\infty[$ d'après le théorème de comparaison des fonctions positives intégrables (elle est majorée en valeur absolue par $u \longmapsto \dfrac{1}{u^{2\alpha}}$ et $2\alpha > 1$).

Posons pour $n \geqslant 1$, $u_n = \dfrac{\sin(\pi\sqrt{n})}{n^\alpha}$ et $v_n = \displaystyle\int_n^{n+1} \frac{\sin \pi\sqrt{t}}{t^\alpha}\mathrm{d}t$. D'après ce qui précède, la série de terme général v_n converge. Pour prouver que la série $\sum u_n$ converge, il suffit donc d'établir la convergence de la série de terme général $u_n - v_n$. Or, si $\varphi(t) = \dfrac{\sin \pi\sqrt{t}}{t^\alpha}$ pour $t \geqslant 1$,

$$|u_n - v_n| = \left|\int_n^{n+1} (\varphi(t) - \varphi(n))\mathrm{d}t\right| \leqslant \int_n^{n+1} |\varphi(t) - \varphi(n)|\mathrm{d}t.$$

Comme φ est de classe $\mathcal{C}^\infty$ et

$$\varphi'(t) = \frac{\frac{\pi}{2\sqrt{t}}\cos(\pi\sqrt{t})t^\alpha - \alpha t^{\alpha-1}\sin(\pi\sqrt{t})}{t^{2\alpha}} = O\left(\frac{1}{t^{\alpha+\frac{1}{2}}}\right),$$

il existe $K \geqslant 0$ tel que pour tout $t \geqslant 1$, $|\varphi'(t)| \leqslant \frac{K}{t^{\alpha+\frac{1}{2}}}$. D'après l'inégalité des accroissements finis,

$$|u_n - v_n| \leqslant \int_n^{n+1} \frac{K dt}{n^{\alpha+\frac{1}{2}}} = \frac{K}{n^{\alpha+\frac{1}{2}}}.$$

Donc $\sum |u_n - v_n|$ converge d'après le théorème de comparaison des séries à termes positifs. Il s'ensuit que $\sum(u_n - v_n)$ converge et $\sum u_n$ également.

3. Traitons le cas $\alpha = \frac{1}{2}$. Déterminons un développement asymptotique de la différence $d_n = e^{i\pi\sqrt{n+1}} - e^{i\pi\sqrt{n}}$:

$$\begin{aligned} d_n &= e^{i\pi\sqrt{n}}\left(e^{i\pi(\sqrt{n+1}-\sqrt{n})} - 1\right) = e^{i\pi\sqrt{n}}\left(e^{i\pi\sqrt{n}(\sqrt{1+1/n}-1)} - 1\right) \\ &= e^{i\pi\sqrt{n}}\left(e^{i\pi\sqrt{n}\left(\frac{1}{2n}+O\left(\frac{1}{n^2}\right)\right)} - 1\right) = e^{i\pi\sqrt{n}}\left(e^{i\pi\left(\frac{1}{2\sqrt{n}}+O\left(\frac{1}{n^{3/2}}\right)\right)} - 1\right) \\ &= e^{i\pi\sqrt{n}}\left(\frac{i\pi}{2\sqrt{n}} - \frac{\pi^2}{8n} + O\left(\frac{1}{n^{3/2}}\right)\right) \\ &= \frac{i\pi e^{i\pi\sqrt{n}}}{2\sqrt{n}} - \frac{\pi^2 e^{i\pi\sqrt{n}}}{8n} + O\left(\frac{1}{n^{3/2}}\right). \end{aligned}$$

En prenant la partie réelle, on obtient alors

$$\cos(\pi\sqrt{n+1}) - \cos(\pi\sqrt{n}) = -\frac{\pi\sin(\pi\sqrt{n})}{2\sqrt{n}} - \frac{\pi^2\cos(\pi\sqrt{n})}{8n} + O\left(\frac{1}{n^{3/2}}\right).$$

La série de terme général $O\left(\frac{1}{n^{3/2}}\right)$ est absolument convergente. Comme précédemment pour le sinus, on prouve que la série de terme général $\frac{\cos(\pi\sqrt{n})}{n}$ converge, si bien que la série de terme général $\frac{\sin(\pi\sqrt{n})}{\sqrt{n}}$ est de même nature que celle de terme général $\cos(\pi\sqrt{n+1}) - \cos(\pi\sqrt{n})$. Or cette dernière diverge, puisque la suite $(\cos\pi\sqrt{n})_{n\geqslant 0}$ ne converge pas : en effet, la suite extraite $(\cos\pi\sqrt{n^2}) = (\cos\pi n) = ((-1)^n)$ est divergente. Par conséquent, la série diverge pour la valeur $\alpha = \frac{1}{2}$.

4. Venons-en maintenant au cas $\alpha < \frac{1}{2}$ et démontrons pas l'absurde qu'il y a divergence de la série. Supposons que la série de terme général

$\dfrac{\sin(\pi\sqrt{n})}{n^\alpha}$ converge et notons $\mathrm{A}_n = \sum_{k=1}^{n} \dfrac{\sin(\pi\sqrt{k})}{k^\alpha}$. Nous allons montrer que la série $\sum \dfrac{\sin(\pi\sqrt{n})}{\sqrt{n}}$ convergerait à l'aide d'une transformation d'Abel, ce qui fournira la contradiction souhaitée :

$$\begin{aligned}\sum_{k=1}^{n} \frac{\sin(\pi\sqrt{k})}{\sqrt{k}} &= \sum_{k=1}^{n} \frac{\sin(\pi\sqrt{k})}{k^\alpha} \frac{1}{k^{\frac{1}{2}-\alpha}} = \sum_{k=1}^{n} (\mathrm{A}_k - \mathrm{A}_{k-1}) \frac{1}{k^{\frac{1}{2}-\alpha}} \\ &= \sum_{k=1}^{n} \frac{\mathrm{A}_k}{k^{\frac{1}{2}-\alpha}} - \sum_{k=0}^{n-1} \frac{\mathrm{A}_k}{(k+1)^{\frac{1}{2}-\alpha}} \text{ avec } \mathrm{A}_0 = 0, \\ &= \frac{\mathrm{A}_n}{n^{\frac{1}{2}-\alpha}} + \sum_{k=1}^{n} \mathrm{A}_k \left(\frac{1}{k^{\frac{1}{2}-\alpha}} - \frac{1}{(k+1)^{\frac{1}{2}-\alpha}} \right).\end{aligned}$$

La suite A_n est bornée, donc le premier terme converge. Quant au second, c'est la somme partielle d'une série absolument convergente puisque

$$\left| \mathrm{A}_k \left(\frac{1}{k^{\frac{1}{2}-\alpha}} - \frac{1}{(k+1)^{\frac{1}{2}-\alpha}} \right) \right| \leqslant \sup_{n\in\mathbb{N}} |\mathrm{A}_n| \left(\frac{1}{k^{\frac{1}{2}-\alpha}} - \frac{1}{(k+1)^{\frac{1}{2}-\alpha}} \right).$$

La série de terme général $\dfrac{\sin \pi\sqrt{n}}{\sqrt{n}}$ se retrouve être convergente : c'est une contradiction.

Conclusion. La série $\sum \dfrac{\sin(\pi\sqrt{n})}{n^\alpha}$ converge si, et seulement si, on a $\alpha > \dfrac{1}{2}$. ◁

Le lecteur courageux pourra étudier plus généralement les séries de la forme $\sum \dfrac{\sin(\pi n^\beta)}{n^\alpha}$ *avec* $\alpha, \beta > 0$.

3.38. Recherche d'un équivalent (1)

Déterminer un équivalent de $u_n = \sum_{k=1}^{+\infty} (-1)^{k-1} \mathrm{E}\left(\dfrac{n}{k}\right)$.

(**École normale supérieure, École polytechnique**)

▷ **Solution.**

Observons pour commencer que la somme définissant u_n est finie. Imaginons un instant que l'on enlève la partie entière. On aurait la suite $n \sum_{k=1}^{n} \dfrac{(-1)^{k-1}}{k}$ qui est équivalente à $n \ln 2$. En effet, c'est un résultat

classique[5] que $\sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} = \ln 2$. On va essayer de formaliser cette idée.

Posons $v_n = \frac{u_n}{n}$. On a $v_n = \sum_{k=1}^{+\infty} f(n,k)$ où $f(n,k) = (-1)^{k-1}\frac{1}{n}\mathrm{E}\left(\frac{n}{k}\right)$.

Pour tout k, $f(n,k)$ tend vers $f(k) = \frac{(-1)^{k-1}}{k}$ lorsque n tend vers $+\infty$. On est donc face à un problème d'interversion de limites. La clef en est la majoration uniforme en n des restes des séries $\sum_k f(n,k)$. Comme la série $\sum_k f(n,k)$ est alternée et $|f(n,k)|$ décroît avec k, on a, pour $\mathrm{N} \geqslant 1$,

$$\left|\sum_{k=\mathrm{N}}^{+\infty}(-1)^{k-1}\frac{1}{n}\mathrm{E}\left(\frac{n}{k}\right)\right| \leqslant \frac{1}{n}\mathrm{E}\left(\frac{n}{\mathrm{N}}\right) \leqslant \frac{1}{\mathrm{N}},$$

et de même $\left|\sum_{k=\mathrm{N}}^{+\infty}(-1)^{k-1}\frac{1}{k}\right| \leqslant \frac{1}{\mathrm{N}}$. Soit $\varepsilon > 0$ et N entier naturel tel que $\frac{1}{\mathrm{N}} \leqslant \varepsilon$. On peut écrire

$$\left|\sum_{k=1}^{+\infty} f(n,k) - \sum_{k=1}^{+\infty} f(k)\right| \leqslant \left|\sum_{k=1}^{\mathrm{N}-1} f(n,k) - \sum_{k=1}^{\mathrm{N}-1} f(k)\right| + \left|\sum_{k=\mathrm{N}}^{+\infty} f(n,k)\right| + \left|\sum_{k=\mathrm{N}}^{+\infty} f(k)\right|$$

$$\leqslant \left|\sum_{k=1}^{\mathrm{N}-1} f(n,k) - \sum_{k=1}^{\mathrm{N}-1} f(k)\right| + 2\varepsilon.$$

Comme $\left|\sum_{k=1}^{\mathrm{N}-1} f(n,k) - \sum_{k=1}^{\mathrm{N}-1} f(k)\right|$ converge vers 0, pour n assez grand, $\left|\sum_{k=1}^{+\infty} f(n,k) - \sum_{k=1}^{+\infty} f(k)\right| \leqslant 3\varepsilon$. On obtient donc $\lim_{n\to+\infty} v_n = \ln 2$ soit

$$\boxed{u_n \underset{n\to\infty}{\sim} n\ln 2.} \quad \lhd$$

Notons que le théorème général d'interversion de limites (qui n'est pas au programme) permet de conclure dès lors que l'on a la majoration uniforme des restes.

5. L'identité $\ln(1+x) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} x^k$ est valable pour $x \in]-1,1]$ en vertu du théorème d'Abel-Dirichlet.

3.39. Recherche d'un équivalent (2)

Pour $\alpha > 0, \alpha \neq 1$, déterminer un équivalent de

$$u_n = \sum_{k=0}^{n} \frac{(n\alpha)^k}{k!} \quad \text{et} \quad v_n = \sum_{k=n+1}^{+\infty} \frac{(n\alpha)^k}{k!}.$$

(École polytechnique)

▷ **Solution.**

On remarque que $u_n + v_n = e^{n\alpha}$, ce qui permet d'espérer que déterminer un des équivalents suffira. On va être amené à distinguer les cas $\alpha > 1$ et $\alpha < 1$.

Pour $\alpha > 1$, les termes de la somme définissant u_n croissent et le terme prépondérant est le dernier. En mettant celui-ci en facteur, on obtient

$$u_n = \frac{(n\alpha)^n}{n!} \sum_{k=0}^{n} \frac{n!}{k!} \frac{1}{(n\alpha)^{n-k}} = \frac{(n\alpha)^n}{n!} \sum_{k=0}^{n} \frac{n!}{(n-k)!} \frac{1}{(n\alpha)^k}.$$

On remarque que $\lim_{n\to+\infty} \dfrac{n!}{(n-k)!} \dfrac{1}{(n\alpha)^k} = \dfrac{1}{\alpha^k}$, puisque $\dfrac{n!}{(n-k)!} \underset{n\to\infty}{\sim} n^k$. On aimerait démontrer que

$$\lim_{n\to+\infty} \sum_{k=0}^{n} \frac{n!}{(n-k)!} \frac{1}{(n\alpha)^k} = \sum_{k=0}^{+\infty} \frac{1}{\alpha^k} = \frac{\alpha}{\alpha - 1}.$$

Il s'agit d'intervertir une sommation et une limite. Posons, pour n et k dans $\mathbb{N}$,

$$u_{n,k} = \frac{n!}{(n-k)!} \frac{1}{(n\alpha)^k} \text{ si } k \leqslant n \quad \text{et} \quad u_{n,k} = 0 \text{ si } k > n.$$

On a, pour tout $(k, n) \in \mathbb{N}^2$, $0 \leqslant u_{n,k} \leqslant \dfrac{1}{\alpha^k}$, ce qui montre que la série $\sum_k u_{n,k}$ converge uniformément[6] par rapport à n.

Si on se donne $\varepsilon > 0$, on peut déterminer n_0 tel que $\sum_{k=n_0}^{+\infty} \dfrac{1}{\alpha^k} \leqslant \varepsilon$. On a alors, pour $n \geqslant n_0$,

6. Ici encore, on aurait pu utiliser le théorème général (hors programme) d'interversion des limites : comme les séries $\sum_k u_{n,k}$ convergent uniformément en n, l'interversion est justifiée.

$$\left|\sum_{k=0}^{n}\frac{n!}{(n-k)!}\frac{1}{(n\alpha)^k}-\frac{\alpha}{\alpha-1}\right| \leqslant \sum_{k=0}^{\infty}\left|u_{n,k}-\frac{1}{\alpha^k}\right|$$
$$\leqslant \sum_{k=0}^{n_0-1}\left|u_{n,k}-\frac{1}{\alpha^k}\right|+2\sum_{k=n_0}^{+\infty}\frac{1}{\alpha^k}$$
$$\leqslant \sum_{k=0}^{n_0-1}\left|u_{n,k}-\frac{1}{\alpha^k}\right|+2\varepsilon.$$

La dernière somme tendant vers 0 quand n tend vers $+\infty$, ceci est inférieur à 3ε pour n assez grand. On a donc la limite de $\sum_{k=0}^{n}\frac{n!}{(n-k)!}\frac{1}{(n\alpha)^k}$ qui vaut $\frac{\alpha}{\alpha-1}$, d'où l'on déduit

$$u_n \underset{n\to\infty}{\sim} \frac{\alpha}{\alpha-1}\frac{(n\alpha)^n}{n!}.$$

En utilisant la formule de Stirling, $n! \sim \left(\frac{n}{e}\right)^n\sqrt{2\pi n}$, on peut encore écrire

$$\boxed{u_n \underset{n\to\infty}{\sim} \frac{\alpha}{\alpha-1}\frac{(e\alpha)^n}{\sqrt{2\pi n}}.}$$

Pour trouver un équivalent de v_n montrons que u_n est négligeable devant $u_n+v_n=e^{n\alpha}$. On a $\frac{u_n}{e^{n\alpha}} \sim \frac{\alpha}{\alpha-1}\left(\frac{e\alpha}{e^\alpha}\right)^n\frac{1}{\sqrt{2\pi n}}$. De l'inégalité $e^x \geqslant x+1$, on déduit $e^\alpha = ee^{\alpha-1} \geqslant e\alpha$ et donc $\frac{\alpha}{\alpha-1}\left(\frac{e\alpha}{e^\alpha}\right)^n\frac{1}{\sqrt{2\pi n}} \leqslant \frac{\alpha}{\alpha-1}\frac{1}{\sqrt{2\pi n}}$. Ceci implique $\lim_{n\to+\infty}\frac{u_n}{e^{n\alpha}}=0$. Puisque u_n est négligeable devant $e^{n\alpha}$, on a

$$\boxed{v_n \underset{n\to\infty}{\sim} e^{n\alpha}.}$$

Dans le cas $\alpha>1$, la méthode précédente ne donne rien. En effet, les termes de la somme définissant u_n décroissent, le terme prépondérant est le premier qui vaut 1 et chaque terme tend vers 0. Par contre, on peut l'appliquer à v_n. On obtient, en mettant $\frac{(n\alpha)^n}{n!}$ en facteur,

$$v_n = \frac{(n\alpha)^n}{n!}\sum_{k=n+1}^{+\infty}\frac{(n\alpha)^{k-n}n!}{k!} = \frac{(n\alpha)^n}{n!}\sum_{k=1}^{+\infty}\frac{n!(n\alpha)^k}{(n+k)!}.$$

Cette fois-ci, on pose, pour tout $k\in\mathbb{N}^*$,

$$u_{n,k}=\frac{n!(n\alpha)^k}{(n+k)!}.$$

On a, pour tout $k \in \mathbb{N}^*$, $\lim_{n\to+\infty} u_{n,k} = \alpha^k$ et, d'autre part, pour tout $(n,k) \in \mathbb{N}\times\mathbb{N}^*$, $|u_{n,k}| \leqslant \alpha^k$. La série $\sum_{k\in\mathbb{N}^*} u_{n,k}$ convergeant uniformément par rapport à n, on peut comme précédemment intervertir sommation et limite, ce qui donne

$$\lim_{n\to+\infty} \sum_{k=1}^{+\infty} \frac{n!(n\alpha)^k}{(n+k)!} = \sum_{k=1}^{+\infty} \alpha^k = \frac{\alpha}{1-\alpha}$$

et donc

$$v_n \underset{n\to\infty}{\sim} \frac{\alpha}{1-\alpha}\frac{(n\alpha)^n}{n!}.$$

Avec l'équivalent de Stirling, cela peut s'écrire encore

$$\boxed{v_n \underset{n\to\infty}{\sim} \frac{\alpha}{1-\alpha}\frac{(e\alpha)^n}{\sqrt{2\pi n}}.}$$

Cette fois, on trouve que v_n est négligeable devant $e^{n\alpha}$ et donc que

$$\boxed{u_n \underset{n\to\infty}{\sim} e^{n\alpha}.} \quad \lhd$$

Le cas $\alpha = 1$ est plus difficile et nécessite l'utilisation de la méthode de Laplace. Le lecteur pourra trouver la solution dans CHAMBERT-LOIR (A.), FERMIGIER (S.), MAILLOT (V.), Exercices de mathématiques pour l'agrégation. *Masson 1994, exercices 6-8, p. 117-119.*

L'exercice qui suit est consacré au produit de Cauchy. On rappelle que le produit de Cauchy de deux séries $\sum u_n$ et $\sum v_n$ (réelles ou complexes) est la série $\sum w_n$ où $w_n = \sum_{k=0}^{n} u_k v_{n-k}$. Si les séries $\sum u_n$ et $\sum v_n$ sont absolument convergentes, il en est de même de $\sum w_n$ et

$$\mathrm{W} = \sum_{n=0}^{+\infty} w_n = \sum_{n=0}^{+\infty} u_n \sum_{n=0}^{+\infty} v_n = \mathrm{UV}$$

Le produit de Cauchy de deux séries convergentes peut diverger. Toutefois, l'exercice suivant montre qu'au sens de Cesàro (c'est-à-dire en moyenne), il converge toujours vers UV. *En particulier, si $\sum w_n$ converge, sa somme ne peut être que le produit* UV *en vertu du théorème de Cesàro.*

3.40. Convergence en moyenne d'un produit de Cauchy

Soit $\sum a_n$ et $\sum b_n$ deux séries de nombres complexes convergentes. On pose pour $n \in \mathbb{N}$, $c_n = \sum_{k=0}^{n} a_k b_{n-k}$, $S_n = \sum_{k=0}^{n} c_k$ et $P_n = \frac{1}{n+1} \sum_{k=0}^{n} S_k$. Étudier la suite $(P_n)_{n\geqslant 0}$.

(École polytechnique)

Solution.

La série $\sum c_n$ est le produit de Cauchy des séries $\sum a_n$ et $\sum b_n$. Comme rappelé ci-dessus, si $\sum a_n$ et $\sum b_n$ sont absolument convergentes, il en va de même de la série $\sum c_n$ et dans ces conditions, on a

$$\left(\sum_{n=0}^{+\infty} a_n\right)\left(\sum_{n=0}^{+\infty} b_n\right) = \sum_{n=0}^{+\infty} c_n.$$

Dans ce cas, $(S_n)_{n\in\mathbb{N}}$ converge vers le produit des sommes des séries et il en va de même de $(P_n)_{n\in\mathbb{N}}$, d'après le théorème de Cesàro.

Nous allons voir dans cet exercice que la suite $(P_n)_{n\in\mathbb{N}}$ converge encore vers $\left(\sum_{n=0}^{+\infty} a_n\right)\left(\sum_{n=0}^{+\infty} b_n\right)$ si les séries $\sum a_n$ et $\sum b_n$ sont seulement convergentes.

Notons pour $n \in \mathbb{N}$, $A_n = \sum_{k=0}^{n} a_k$ et $B_n = \sum_{k=0}^{n} b_k$. Par hypothèse, on a

$$\lim_{n\to+\infty} A_n = \sum_{n=0}^{+\infty} a_n = A \quad \text{et} \quad \lim_{n\to+\infty} B_n = \sum_{n=0}^{+\infty} b_n = B.$$

Pour un entier naturel n, on peut écrire,

$$S_n = \sum_{p=0}^{n}\sum_{k=0}^{p} a_k b_{p-k} = \sum_{k=0}^{n}\sum_{p=k}^{n} a_k b_{p-k} = \sum_{k=0}^{n} a_k \sum_{i=0}^{n-k} b_i = \sum_{k=0}^{n} a_k B_{n-k}$$
$$= \sum_{l=0}^{n} a_{n-l} B_l$$

Exprimons maintenant P_N en fonction des termes des suites (A_n) et (B_n). On obtient, pour $N \in \mathbb{N}$,

$$\begin{aligned} P_N &= \frac{1}{N+1}\sum_{n=0}^{N} S_n = \frac{1}{N+1}\sum_{n=0}^{N}\sum_{k=0}^{n} a_{n-k}B_k \\ &= \frac{1}{N+1}\sum_{k=0}^{N}\sum_{n=k}^{N} a_{n-k}B_k = \frac{1}{N+1}\sum_{k=0}^{N} B_k \sum_{n=k}^{N} a_{n-k} \\ &= \frac{1}{N+1}\sum_{k=0}^{N} A_{N-k}B_k. \end{aligned}$$

Montrons que $\lim_{N\to+\infty} P_N = AB$. Nous avons

$$\begin{aligned} |P_N - AB| &= \left|\frac{1}{N+1}\left(\sum_{k=0}^{N}(A_{N-k}B_k - AB)\right)\right| \\ &= \frac{1}{N+1}\left|\sum_{k=0}^{N}((A_{N-k}-A)B_k + A(B_k - B))\right| \\ &\leqslant \frac{1}{N+1}\sum_{k=0}^{N}(|A_{N-k}-A||B_k| + |A||B_k - B|) \end{aligned}$$

La suite $(B_n)_{n\in\mathbb{N}}$ est bornée car convergente. Soit $M \geqslant 0$ tel que pour tout $k \geqslant 0$, $|B_k| \leqslant M$. On obtient la majoration

$$\begin{aligned} |P_N - AB| &\leqslant \frac{M}{N+1}\sum_{k=0}^{N}|A_{N-k}-A| + \frac{|A|}{N+1}\sum_{k=0}^{N}|B_k - B| \\ &= \frac{M}{N+1}\sum_{k=0}^{N}|A_k - A| + \frac{|A|}{N+1}\sum_{k=0}^{N}|B_k - B|. \end{aligned}$$

Puisque $\lim_{N\to+\infty}|A_N - A| = \lim_{N\to+\infty}|B_N - B| = 0$, on obtient par théorème de Cesàro $\lim_{N\to+\infty}\frac{1}{N+1}\sum_{k=0}^{N}|A_k - A| = \frac{1}{N+1}\sum_{k=0}^{N}|B_k - B| = 0$ et donc $\lim_{N\to+\infty} P_N = AB$. On conclut que

$$\boxed{\lim_{N\to+\infty} P_N = \left(\sum_{n=0}^{+\infty} a_n\right)\left(\sum_{n=0}^{+\infty} b_n\right).} \quad \triangleleft$$

Les exercices suivants sont consacrés aux familles sommables et aux séries doubles. Rappelons quelques résultats utiles : une famille $(u_i)_{i\in I}$ de réels positifs est sommable s'il existe $M \geqslant 0$ tel que, pour tout $J \subset I$ fini, $\sum_{i\in J} u_i \leqslant M$. La somme $\sum_{i\in I} u_i$ est alors la borne supérieure des sommes

finies $\sum_{i\in J} u_i$ *où* J *décrit l'ensemble des parties finies de* I. *Dans ces conditions, les* i *pour lesquels* $u_i \neq 0$ *forment un ensemble dénombrable et toute sous-famille est encore sommable.*

Le théorème d'associativité, ou encore théorème de sommation par paquets, constitue une proposition essentielle. Il affirme que si $(u_i)_{i\in I}$ *est une famille de réels positifs, et* $(I_k)_{k\in K}$ *une partition de* I, *on a l'équivalence des conditions suivantes :*

(i) $\forall k \in K$, $(u_i)_{i\in I_k}$ *est sommable et* $\left(\sum_{i\in I_k} u_i\right)_{k\in K}$ *est sommable ;*

(ii) $(u_i)_{i\in I}$ *est sommable.*

Dans ces conditions, $\sum_{i\in I} u_i = \sum_{k\in K} \sum_{i\in I_k} u_i$.

Une famille $(u_i)_{i\in I}$ *de nombres réels ou complexes est dite sommable si la famille* $(|u_i|)_{i\in I}$ *l'est. Dans ces conditions (avec* I *dénombrable), pour toute suite croissante* $(J_n)_{n\geqslant 0}$ *de parties finies de* I *recouvrant* I, $\sum_{i\in J_n} u_i$ *tend, quand* n *tend vers* $+\infty$, *vers un réel* L *indépendant de la suite* $(J_n)_{n\in\mathbb{N}}$ *choisie. Par définition,* L *est la somme des* u_i, *notée* $\sum_{i\in I} u_i$.

Si $(u_i)_{i\in I}$ *est une famille sommable,* $(I_k)_{k\geqslant 0}$ *une partition de* I, *on a encore la sommabilité de la famille* $\left(\sum_{i\in I_k} u_i\right)_{k\in K}$ *et*

$$\sum_{i\in I} u_i = \sum_{k\in K} \sum_{i\in I_k} u_i.$$

Ce théorème s'exprime ainsi pour les séries doubles : si $(u_{p,q})_{p,q\geqslant 0}$ *est une famille de nombres réels ou complexes, la famille est sommable dès lors qu'une des séries suivantes est définie et convergente :*

$$\sum_{p\in\mathbb{N}} \sum_{q\in\mathbb{N}} |u_{p,q}| \quad \text{ou} \quad \sum_{q\in\mathbb{N}} \sum_{p\in\mathbb{N}} |u_{p,q}|,$$

et dans ces conditions, $\sum_{p,q\geqslant 0} u_{p,q} = \sum_{p\in\mathbb{N}} \sum_{q\in\mathbb{N}} u_{p,q} = \sum_{q\in\mathbb{N}} \sum_{p\in\mathbb{N}} u_{p,q}$.

3.41. Application du théorème d'associativité

Établir, pour tout x réel tel que $|x| < 1$, les identités

$$\sum_{n=1}^{+\infty} \frac{x^{2n-1}}{1-x^{2n-1}} = \sum_{n=1}^{+\infty} \frac{x^n}{1-x^{2n}} \quad \text{et} \quad \sum_{n=0}^{+\infty} \frac{x^{2^n}}{1-x^{2^{n+1}}} = \frac{x}{1-x}.$$

(École polytechnique)

▷ **Solution.**

Prenons $x \in]-1,1[$.

• On peut écrire de manière formelle (les justifications viendront ensuite)

$$\begin{aligned}\sum_{n=1}^{+\infty} \frac{x^{2n-1}}{1-x^{2n-1}} &= \sum_{n=1}^{+\infty}\sum_{m=1}^{+\infty} x^{(2n-1)m} \text{ (somme d'une série géométrique),}\\ &= \sum_{m=1}^{+\infty}\sum_{n=1}^{+\infty} x^{(2n-1)m} = \sum_{m=1}^{+\infty}\sum_{n=0}^{+\infty} x^{(2n+1)m}\\ &= \sum_{m=1}^{+\infty} x^m \left(\sum_{n=0}^{+\infty} (x^{2m})^n\right)\\ &= \sum_{m=1}^{+\infty} \frac{x^m}{1-x^{2m}}.\end{aligned}$$

Ces égalités, et la convergence des séries en question, seront justifiées par le théorème d'associativité dès lors que nous aurons prouvé la sommabilité de la famille $(x^{(2n+1)m})_{\substack{n\geqslant 0\\ m\geqslant 1}}$. Comme $|x| < 1$, la série $\sum_n |x|^{(2n+1)m}$ converge et

$$\sum_{n=0}^{+\infty} |x|^{(2n+1)m} = \frac{|x|^m}{1-|x|^{2m}} \underset{m\to+\infty}{\sim} |x|^m.$$

Ainsi, d'après le théorème de comparaison des séries à termes positifs, la série $\sum_m \left(\sum_{n=0}^{+\infty} |x|^{(2n+1)m}\right)$ converge et le théorème d'associativité assure la sommabilité de la famille $(x^{(2n+1)m})_{\substack{n\geqslant 0\\ m\geqslant 1}}$.

On conclut que les séries $\sum \frac{x^{2n-1}}{1-x^{2n-1}}$ et $\sum \frac{x^n}{1-x^{2n}}$ convergent et que

$$\boxed{\sum_{n=1}^{+\infty} \frac{x^{2n-1}}{1-x^{2n-1}} = \sum_{n=1}^{+\infty} \frac{x^n}{1-x^{2n}}.}$$

• Procédons de même par un calcul formel pour la seconde identité :

$$\sum_{n=0}^{+\infty} \frac{x^{2^n}}{1-x^{2^{n+1}}} = \sum_{n=0}^{+\infty}\sum_{m=0}^{+\infty} (x^{2^{n+1}})^m = \sum_{n=0}^{+\infty}\sum_{m=0}^{+\infty} x^{2^n(2m+1)}.$$

En remarquant que tout entier $k \geqslant 1$ s'écrit de manière unique $k = 2^n(2m+1)$ avec $n \geqslant 0$ et $m \geqslant 0$, par associativité formelle, il vient

$$\sum_{n=0}^{+\infty} \frac{x^{2^n}}{1-x^{2^{n+1}}} = \sum_{k=1}^{+\infty} x^k = \frac{x}{1-x}.$$

La convergence des séries en question et les égalités sont justifiées par le théorème d'associativité, puisque la famille $(x^k)_{k\geqslant 1}$ est sommable ($|x|<1$). On conclut

$$\boxed{\sum_{n=0}^{+\infty}\frac{x^{2^n}}{1-x^{2^{n+1}}}=\frac{x}{1-x}\cdot}\ \triangleleft$$

3.42. Étude de sommabilité

Soit a et b des réels strictement positifs. Montrer que la famille $\left(\dfrac{1}{a^m+b^n}\right)_{m,n\geqslant 0}$ est sommable si, et seulement si, $a>1$ et $b>1$.

(École polytechnique)

▷ **Solution.**

Supposons $a\leqslant 1$. Alors, la sous-famille $\left(\dfrac{1}{a^m+b^0}\right)_{m\geqslant 0}$ n'est pas sommable puisque la série de terme général $\dfrac{1}{a^m+1}$ est grossièrement divergente. Il s'ensuit que la famille $\left(\dfrac{1}{a^m+b^n}\right)_{m,n\geqslant 0}$ ne peut être sommable. Il en va de même si $b\leqslant 1$.

Supposons $a<1$ et $b<1$. On va utiliser le théorème d'associativité en partitionnant les indices (m,n) selon la valeur de $m+n=p$. La famille étudiée est donc sommable si la série

$$\sum_p\Bigg(\sum_{m+n=p}\frac{1}{a^m+b^n}\Bigg)\ \text{converge.}$$

Notons $c=\min(a,b)>1$. On a

$$\sum_{m+n=p}\frac{1}{a^m+b^n}\leqslant\sum_{k=0}^{p}\frac{1}{c^k+c^{p-k}}\leqslant\sum_{k=0}^{p}\frac{1}{c^{p/2}}=\frac{(p+1)}{\sqrt{c^p}}=o\Big(\frac{1}{p^2}\Big),$$

car l'un des deux termes de la somme c^k+c^{p-k} est plus grand que $c^{p/2}$. D'après le théorème de comparaison des séries à termes positifs, la série $\sum_p\Big(\sum_{m+n=p}\dfrac{1}{a^m+b^n}\Big)$ converge et on conclut que la famille $\left(\dfrac{1}{a^m+b^n}\right)_{m,n\geqslant 0}$ est sommable. ◁

3.43. Convergence et somme d'une série double

Établir la convergence et calculer $\sum_{n=1}^{+\infty}\sum_{m=1}^{+\infty}\frac{1}{m^2n+n^2m+2mn}$.
(École polytechnique)

▷ **Solution.**

Pour n fixé, $n \geqslant 1$, on a $\frac{1}{m^2n+n^2m+2mn} \underset{n\to\infty}{\sim} \frac{1}{m^2n}$ de sorte que la série $\sum_m \frac{1}{m^2n+n^2m+2mn}$ converge. Pour calculer sa somme, on décompose la fraction rationnelle en éléments simples :

$$\frac{1}{m^2n+n^2n+2mn} = \frac{1}{mn(m+n+2)} = \frac{1}{n(n+2)}\left(\frac{1}{m}-\frac{1}{m+n+2}\right)$$

et on obtient $\sum_{m=1}^{+\infty}\frac{1}{m^2n+n^2n+2mn} = \frac{1}{n(n+2)}\sum_{m=1}^{+\infty}\left(\frac{1}{m}-\frac{1}{m+n+2}\right)$.

Posons et calculons $u_n = \sum_{m=1}^{+\infty}\left(\frac{1}{m}-\frac{1}{m+n+2}\right)$:

$$\begin{aligned} u_n &= \sum_{m=1}^{+\infty}\left(\frac{1}{m}-\frac{1}{m+n+2}\right) = \lim_{N\to+\infty}\sum_{m=1}^{N}\left(\frac{1}{m}-\frac{1}{m+n+2}\right) \\ &= \lim_{N\to+\infty}\left(\sum_{m=1}^{n+2}\frac{1}{m}-\sum_{m=N+1}^{N+n+2}\frac{1}{m}\right) = \sum_{m=1}^{n+2}\frac{1}{m}. \end{aligned}$$

Classiquement, on a $u_n \underset{n\to\infty}{\sim} \ln n$ et donc $\frac{1}{n(n+2)}u_n \underset{n\to\infty}{\sim} \frac{\ln n}{n^2}$. Grâce au théorème de comparaison des séries à termes positifs, on en déduit que la série double étudiée converge.

Remarquons que d'après le théorème d'associativité, la famille positive $\left(\frac{1}{m^2n+n^2m+2mn}\right)_{n,m\geqslant 1}$ *est sommable.*

Calculons la valeur de sa somme, *i.e.* la somme de la série $\sum\frac{1}{n(n+2)}u_n$. Pour $N \geqslant 1$,

$$\begin{aligned} \sum_{n=1}^{N}\frac{1}{n(n+2)}u_n &= \frac{1}{2}\sum_{n=1}^{N}\left(\frac{u_n}{n}-\frac{u_n}{n+2}\right) = \frac{1}{2}\left(\sum_{n=1}^{N}\frac{u_n}{n}-\sum_{n=3}^{N+2}\frac{u_{n-2}}{n}\right) \\ &= \frac{1}{2}\left(u_1+\frac{u_2}{2}+\sum_{n=3}^{N+2}\frac{u_n-u_{n-2}}{n}-\frac{u_{N+1}}{N+1}-\frac{u_{N+2}}{N+2}\right). \end{aligned}$$

Simplifions l'expression de la dernière somme en décomposant la fraction rationnelle en éléments simples :

$$\sum_{n=3}^{N+2} \frac{u_n - u_{n-2}}{n} = \sum_{n=3}^{N+2} \frac{1}{n}\left(\frac{1}{n+2} + \frac{1}{n+1}\right)$$
$$= \sum_{n=3}^{N+2} \left(\frac{3}{2n} - \frac{1}{n+1} - \frac{1}{2(n+2)}\right)$$
$$= \frac{3}{2}\sum_{n=3}^{N+2} \frac{1}{n} - \sum_{n=4}^{N+3} \frac{1}{n} - \frac{1}{2}\sum_{n=5}^{N+4} \frac{1}{n}$$
$$= \frac{3}{2}\left(\frac{1}{3} + \frac{1}{4}\right) - \frac{1}{4} - \frac{3}{2(N+3)} - \frac{1}{2(N+4)}.$$

On en déduit que $\lim\limits_{N\to+\infty} \sum\limits_{n=3}^{N+2} \dfrac{u_n - u_{n-2}}{n} = \dfrac{5}{8}$.

Sachant que $\dfrac{u_N}{N} \underset{N\to\infty}{\sim} \dfrac{\ln N}{N}$, on obtient $\lim\limits_{N\to+\infty} \dfrac{u_N}{N} = 0$ et donc $\sum\limits_{n=1}^{+\infty} \dfrac{1}{n(n+2)} u_n = \dfrac{1}{2}\left(u_1 + \dfrac{u_2}{2} + \dfrac{5}{8}\right)$. On calcule $u_1 = \dfrac{11}{6}$ et $u_2 = \dfrac{25}{12}$ et on conclut $\sum\limits_{n=1}^{+\infty} \dfrac{1}{n(n+2)} u_n = \dfrac{7}{4}$.

Conclusion. La série double $\sum\limits_{m,n} \dfrac{1}{m^2 n + n^2 m + 2mn}$ converge et sa somme est $\dfrac{7}{4}$. ◁

3.44. Étude de séries doubles

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de réels positifs ou nuls.

1. On suppose que la série $\sum a_n$ converge. Montrer que $\sum \dfrac{a_n}{n(n+1)}$ converge. Montrer que $\sum\limits_{k=0}^{+\infty} k \sum\limits_{n=k}^{+\infty} \dfrac{a_n}{n(n+1)}$ existe et comparer sa valeur à $\sum\limits_{n=0}^{+\infty} a_n$.

2. On suppose que la série $\sum \sqrt{n} a_n$ converge et on pose pour $n \geqslant 0$, $w_n = \sum\limits_{p=n}^{+\infty} a_p^2$. Montrer l'existence de w_n et la convergence de la série $\sum \sqrt{\dfrac{w_n}{n}}$.

(École polytechnique)

▷ **Solution.**

1. Pour $n \geqslant 1$, $\dfrac{a_n}{n(n+1)} \leqslant a_n$. D'après le théorème de comparaison des séries à termes positifs, $\sum \dfrac{a_n}{n(n+1)}$ converge.

Posons pour tout $(k,n) \in \mathbb{N}^{*2}$, $u_{k,n} = \begin{cases} \dfrac{ka_n}{n(n+1)} \text{ si } n \geqslant k \\ 0 \text{ si } n < k \end{cases}$. Nous avons donc à montrer la convergence de la série $\sum\limits_{k=1}^{+\infty} \sum\limits_{n=1}^{+\infty} u_{k,n}$. Il s'agit ici d'une série double à termes positifs. La convergence de cette série équivaut donc à celle de la série $\sum\limits_{n=1}^{+\infty} \sum\limits_{k=1}^{+\infty} u_{k,n}$ *i.e.* $\sum\limits_{n=1}^{+\infty} \sum\limits_{k=1}^{n} u_{k,n}$. Or, pour tout $n \geqslant 1$,

$$\sum_{k=1}^{n} u_{k,n} = \frac{a_n}{n(n+1)} \sum_{k=1}^{n} k = \frac{a_n}{n(n+1)} \frac{n(n+1)}{2} = \frac{a_n}{2}.$$

Donc, comme par hypothèse, $\sum \dfrac{a_n}{2}$ converge, la famille de réels positifs $(u_{k,n})_{k,n\geqslant 1}$ est sommable et

$$\sum_{k,n\geqslant 1} u_{k,n} = \sum_{n=1}^{+\infty} \sum_{k=1}^{+\infty} u_{k,n} = \sum_{n=1}^{+\infty} \frac{a_n}{2} = \frac{1}{2} \sum_{n=1}^{+\infty} a_n.$$

En intervertissant les sommations, il vient

$$\frac{1}{2} \sum_{n=1}^{+\infty} a_n = \sum_{k=1}^{+\infty} \sum_{n=1}^{+\infty} u_{k,n} = \sum_{k=1}^{+\infty} \sum_{n=k}^{+\infty} \frac{ka_n}{n(n+1)}.$$

On conclut donc que la série $\sum\limits_{k=1}^{+\infty} k \sum\limits_{n=k}^{+\infty} \dfrac{a_n}{n(n+1)}$ converge et

$$\boxed{\sum_{k=0}^{+\infty} k \sum_{n=k}^{+\infty} \frac{a_n}{n(n+1)} = \frac{1}{2} \sum_{n=1}^{+\infty} a_n.}$$

2. • Si $n \geqslant 1$, $a_n \leqslant \sqrt{n} a_n$. D'après le théorème de comparaison des séries à termes positifs, la série des a_n converge. En particulier, a_n tend vers 0 lorsque n tend vers $+\infty$. Donc, pour n assez grand, $0 \leqslant a_n \leqslant 1$. Dans ces conditions, $a_n^2 \leqslant a_n$. D'après le théorème de comparaison des séries à termes positifs, la série des a_n^2 converge. L'existence des w_n est donc prouvée.

• Posons pour $n \geqslant 0$, $b_n = \sqrt{n} a_n$. La série $\sum b_n$ converge par hypothèse. On a $\sqrt{\dfrac{w_n}{n}} = \sqrt{\sum\limits_{k=n}^{+\infty} \dfrac{b_k^2}{kn}}$.

Remarquons que si $(x_k)_{1\leqslant k\leqslant n}$ est une famille de réels positifs, on a

$$x_1^2 + x_2^2 + \cdots + x_n^2 \leqslant (x_1 + x_2 + \cdots + x_n)^2,$$

ce qui s'écrit encore,

$$\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2} \leqslant x_1 + x_2 + \cdots + x_n.$$

Par conséquent, pour tout $N \geqslant n$, on peut écrire

$$\sqrt{\sum_{k=n}^{N} \frac{b_k^2}{kn}} \leqslant \sum_{k=n}^{+\infty} \frac{b_k}{\sqrt{nk}}.$$

En faisant tendre N vers l'infini, il reste $\sqrt{\frac{w_n}{n}} \leqslant \sum_{k=n}^{+\infty} \frac{b_k}{\sqrt{nk}}$. Par conséquent, pour $N \geqslant 1$,

$$\sum_{n=1}^{N} \sqrt{\frac{w_n}{n}} \leqslant \sum_{n=1}^{N} \sum_{k=n}^{+\infty} \frac{b_k}{\sqrt{kn}}$$

Or la famille des $\frac{b_k}{\sqrt{kn}}$ pour $n \in \mathbb{N}$ et $k \geqslant n$ est sommable : en effet, comme cette famille est constitué de réels positifs, il suffit de vérifier que la série obtenue par interversion des sommations, à savoir $\sum_{k=1}^{+\infty} \sum_{n=1}^{k} \frac{b_k}{\sqrt{kn}}$, converge. Or, pour $k \geqslant 1$, $\sum_{n=1}^{k} \frac{b_k}{\sqrt{kn}} = \frac{b_k}{\sqrt{k}} \sum_{n=1}^{k} \frac{1}{\sqrt{n}} \leqslant \frac{b_k}{\sqrt{k}} \int_0^{k-1} \frac{\mathrm{d}t}{\sqrt{t}}$ puisque $\frac{1}{\sqrt{n}} \leqslant \int_{n-1}^{n} \frac{\mathrm{d}t}{\sqrt{t}}$ (la fonction $x \longmapsto 1/\sqrt{x}$ est continue, décroissante et intégrable sur $]0, k-1[$). Par conséquent,

$$\sum_{n=1}^{k} \frac{b_k}{\sqrt{kn}} \leqslant \frac{b_k}{\sqrt{k}} 2\sqrt{k-1} \leqslant 2b_k$$

D'après le théorème de comparaison des séries à termes positifs, la série $\sum_{k=1}^{+\infty} \sum_{n=1}^{k} \frac{b_k}{\sqrt{kn}}$ converge. Il s'ensuit que

$$\sum_{n=1}^{N} \sqrt{\frac{w_n}{n}} \leqslant \sum_{\substack{n\in\mathbb{N}^* \\ k\geqslant n}} \frac{b_k}{\sqrt{kn}} = \sum_{n=1}^{+\infty} \sum_{k=n}^{+\infty} \frac{b_k}{\sqrt{kn}},$$

ce qui assure la convergence de la série de terme général $\sqrt{\frac{w_n}{n}}$. ◁

3.45. Inégalité de Carleman (1923)

Montrer qu'il existe une constante universelle C telle que, pour toute série convergente à termes positifs $\sum a_n$, on ait l'inégalité

$$\sum_{n=1}^{+\infty} \sqrt[n]{a_1 \dots a_n} \leqslant \mathrm{C} \sum_{n=1}^{+\infty} a_n.$$

Quelle est la plus petite valeur possible de C ?

(École normale supérieure)

▷ **Solution.**

• Dans un premier temps, il paraît naturel d'exploiter directement l'inégalité arithmético-géométrique

$$\sqrt[n]{a_1 \dots a_n} \leqslant \frac{a_1 + a_2 + \dots + a_n}{n}.$$

Malheureusement, la famille $u_{k,n}$ définie par $u_{k,n} = \dfrac{a_k}{n}$ pour $1 \leqslant k \leqslant n$ n'est pas en général sommable puisque si $a_k \neq 0$, $\sum\limits_{n=k}^{+\infty} u_{k,n}$ diverge.

• Écrivons plutôt

$$\sqrt[n]{a_1 \dots a_n} = \frac{\sqrt[n]{(1a_1)(2a_2)\dots(na_n)}}{\sqrt[n]{n!}} \leqslant \frac{1}{n\sqrt[n]{n!}} \sum_{k=1}^{n} k a_k.$$

Nous allons user de l'identité $\sum\limits_{n=k}^{+\infty} \dfrac{1}{n(n+1)} = \sum\limits_{n=k}^{+\infty} \left(\dfrac{1}{n} - \dfrac{1}{n+1} \right) = \dfrac{1}{k}$; on peut écrire

$$\sqrt[n]{a_1 \dots a_n} \leqslant \frac{n+1}{\sqrt[n]{n!}} \sum_{k=1}^{n} \frac{k a_k}{n(n+1)}.$$

• La famille $(v_{k,n})_{k,n \geqslant 1}$ définie par $v_{k,n} = \dfrac{k a_k}{n(n+1)}$ pour $1 \leqslant k \leqslant n$ et 0 sinon, est sommable puisqu'elle est positive, que pour k fixé,

$$\sum_{n=1}^{+\infty} v_{k,n} = \sum_{n=k}^{+\infty} v_{k,n} = k a_k \sum_{n=k}^{+\infty} \frac{1}{n(n+1)} = k a_k \frac{1}{k} = a_k$$

et que la série de terme général a_k est convergente.

• La quantité $\dfrac{n+1}{\sqrt[n]{n!}}$ admet une limite en $+\infty$ puisque la formule de Stirling donne

$$\frac{n+1}{\sqrt[n]{n!}} \underset{n \to +\infty}{\sim} n \frac{e}{n} = e.$$

Par conséquent, il existe $C \geqslant 0$ tel que pour tout $n \geqslant 1$, $\dfrac{n+1}{\sqrt[n]{n!}} \leqslant C$. Ainsi, si $N \geqslant 1$, $\sqrt[n]{a_1 \dots a_n} \leqslant \dfrac{n+1}{\sqrt[n]{n!}} \sum_{k=1}^{n} v_{k,n} \leqslant C \sum_{k=1}^{n} v_{k,n}$ et

$$\sum_{n=1}^{N} \sqrt[n]{a_1 \dots a_n} \leqslant \sum_{n=1}^{N} C \sum_{k=1}^{n} v_{k,n} \leqslant C \sum_{(k,n)\in(\mathbb{N}^*)^2} v_{k,n}$$
$$\leqslant C \sum_{k=1}^{+\infty} \sum_{n=1}^{+\infty} v_{k,n} = C \sum_{k=1}^{+\infty} a_k.$$

La série $\sum \sqrt[n]{a_1 \dots a_n}$ étant à termes positifs, elle est convergente et

$$\sum_{n=1}^{+\infty} \sqrt[n]{a_1 \dots a_n} \leqslant C \sum_{n=1}^{+\infty} a_n.$$

• Recherchons maintenant la meilleure constante C. Nous allons prouver qu'en fait $\dfrac{n+1}{\sqrt[n]{n!}} \leqslant e$. Cette inégalité est équivalente à

$$\ln(n+1) \leqslant 1 + \frac{1}{n} \sum_{k=1}^{n} \ln k \quad \text{ou encore à} \quad (n+1)\ln(n+1) - n \leqslant \sum_{k=1}^{n+1} \ln k.$$

Or pour $2 \leqslant k \leqslant n+1$, on a $\ln k \geqslant \int_{k-1}^{k} \ln x \mathrm{d}x$. On en déduit que

$$\sum_{k=1}^{n+1} \ln k = \sum_{k=2}^{n+1} \ln k \geqslant \int_{1}^{n+1} \ln x \mathrm{d}x$$

et comme

$$\int_{1}^{n+1} \ln x \mathrm{d}x = [x(\ln x - 1)]_{1}^{n+1} = (n+1)\ln(n+1) - (n+1) - (-1)$$
$$= (n+1)\ln(n+1) - n,$$

on a bien l'inégalité. La constante $C = e$ convient.

• C'est en fait la meilleure constante possible. Soit C une constante réalisant l'inégalité. Il y a égalité dans l'inégalité arithmético-géométrique lorsque tous les termes sont égaux : $1a_1 = 2a_2 = \dots = na_n$. On aurait envie de choisir $a_k = \dfrac{1}{k}$ mais la série harmonique diverge. Aussi, fixons N un entier naturel non nul et posons $a_k = \dfrac{1}{k}$ pour $1 \leqslant k \leqslant N$ et $a_k = 0$ pour $k > N$. Ainsi,

$$\sum_{n=1}^{N} \frac{1}{\sqrt[n]{n!}} \leqslant C \sum_{n=1}^{N} \frac{1}{n} \underset{N\to\infty}{\sim} C \ln N,$$

tandis que $\dfrac{1}{\sqrt[n]{n!}} \underset{n\to\infty}{\sim} \dfrac{e}{n}$ est le terme général d'une série divergente. Par conséquent, la somme partielle $\displaystyle\sum_{n=1}^{N} \frac{1}{\sqrt[n]{n!}}$ est équivalent à $\displaystyle\sum_{n=1}^{N} \frac{e}{n} \underset{N\to\infty}{\sim} e \ln N$. Nécessairement, $e \leqslant C$.

Conclusion. Pour toute série $\sum a_n$ à termes positifs, on a l'inégalité

$$\sum_{n=1}^{+\infty} \sqrt[n]{a_1 \dots a_n} \leqslant e \sum_{n=1}^{+\infty} a_n,$$

et e est la meilleure constante réalisant cette inégalité. ◁

Bien qu'il n'utilise pas la théorie des familles sommables, nous avons tenu à rapprocher l'exercice suivant de l'inégalité de Carleman : si cette dernière permettait de majorer la somme de la série des moyennes géométriques des n premiers termes d'une suite (a_n), l'inégalité de Hardy s'avère être un résultat analogue sur les moyennes harmoniques ; en effet, si dans l'inégalité de Carleman on remplace les a_n par leurs inverses et les moyennes géométriques par des moyennes harmoniques, on tombe sur l'inégalité de Hardy.

3.46. Inégalité de Hardy

Établir l'existence d'une constante $K > 0$ telle que, pour toute suite $(a_n)_{n\geqslant 1}$ de réels strictement positifs avec $\sum \dfrac{1}{a_n}$ convergente, on ait

$$\sum_{n=1}^{+\infty} \frac{n}{a_1 + \cdots + a_n} \leqslant K \sum_{n=1}^{+\infty} \frac{1}{a_n}.$$

On pourra commencer par établir que

$$\frac{n}{a_1 + \cdots + a_n} \leqslant \frac{4}{n(n+1)^2} \sum_{k=1}^{n} \frac{k^2}{a_k}.$$

Quelle est la meilleure constante K possible ?

(École polytechnique)

▷ **Solution.**

Pour $n \geqslant 1$, posons $u_n = \dfrac{n}{a_1 + \cdots + a_n}$.

• Soit $n \geqslant 1$. D'après l'inégalité de Cauchy-Schwarz, on peut écrire

$$\left(\sum_{k=1}^{n} k\right)^2 = \left(\sum_{k=1}^{n} \frac{k}{\sqrt{a_k}}\sqrt{a_k}\right)^2 \leqslant \left(\sum_{k=1}^{n} \frac{k^2}{a_k}\right)\left(\sum_{k=1}^{n} a_k\right),$$

ce qui donne

$$\frac{1}{a_1 + \cdots + a_n} \leqslant \frac{1}{\left(\sum_{k=1}^{n} k\right)^2}\left(\sum_{k=1}^{n} \frac{k^2}{a_k}\right) = \frac{4}{n^2(n+1)^2}\left(\sum_{k=1}^{n} \frac{k^2}{a_k}\right).$$

Par conséquent, $u_n \leqslant 4\sum_{k=1}^{n} \frac{k^2}{a_k n(n+1)^2}$. On a pour $N \geqslant 1$,

$$\begin{aligned}\sum_{n=1}^{N} u_n &\leqslant 4\sum_{n=1}^{N}\sum_{k=1}^{n} \frac{k^2}{a_k n(n+1)^2} = 4\sum_{k=1}^{N}\sum_{n=k}^{N} \frac{k^2}{a_k n(n+1)^2} \\ &= 4\sum_{k=1}^{N} \frac{k^2}{a_k}\left(\sum_{n=k}^{N} \frac{1}{n(n+1)^2}\right).\end{aligned}$$

Majorons $\sum_{n=k}^{N} \frac{1}{n(n+1)^2}$. Pour $n \geqslant 1$, on a

$$\frac{1}{n^2} - \frac{1}{(n+1)^2} = \frac{2n+1}{n^2(n+1)^2} \geqslant \frac{2n}{n^2(n+1)^2} = \frac{2}{n(n+1)^2},$$

et on en déduit que

$$\sum_{n=k}^{N} \frac{1}{n(n+1)^2} \leqslant \sum_{n=k}^{N} \frac{1}{2}\left(\frac{1}{n^2} - \frac{1}{(n+1)^2}\right) = \frac{1}{2}\left(\frac{1}{k^2} - \frac{1}{(N+1)^2}\right) \leqslant \frac{1}{2k^2}.$$

Il s'ensuit que

$$\sum_{n=1}^{N} u_n \leqslant 4\sum_{k=1}^{N} \frac{k^2}{a_k}\frac{1}{2k^2} \leqslant 2\sum_{k=1}^{N} \frac{1}{a_k} \leqslant 2\sum_{k=1}^{+\infty} \frac{1}{a_k}.$$

La convergence de la série de terme général u_n est donc garantie et

$$\sum_{n=1}^{+\infty} \frac{n}{a_1 + \cdots + a_n} \leqslant 2\sum_{k=1}^{+\infty} \frac{1}{a_k}.$$

La constante $K = 2$ satisfait l'inégalité ; nous allons prouver maintenant que c'est la plus petite possible.

Puisque la moyenne harmonique est inférieure ou égale à la moyenne géométrique, l'exercice précédent (3.45) permet d'établir directement

l'existence de K *(en remarquant que* a_n *est changé en* $\frac{1}{a_n}$*) et de majorer* K *par* e.

• Soit K une constante satisfaisant la condition de l'énoncé. Pour garantir l'égalité dans l'inégalité de Cauchy-Schwarz, il suffirait que pour k entier non nul $\frac{k}{\sqrt{a_k}} = \sqrt{a_k}$ *i.e.* $a_k = k$. Malheureusement, la série des $1/k$ est divergente. L'idée est prendre $a_k = k$ jusqu'à un certain rang, puis de prendre $\frac{1}{a_k}$ très petit au-delà.

Soit $N \geqslant 1$. Pour $n \leqslant N$, on pose $a_n = n$ et pour $n > N$, $a_n = 2^n$. On a alors

$$\sum_{n=1}^{+\infty} \frac{1}{a_n} \leqslant \sum_{n=1}^{N} \frac{1}{n} + \sum_{n=N+1}^{+\infty} \frac{1}{2^n} \leqslant \sum_{n=1}^{N} \frac{1}{n} + \sum_{n=1}^{+\infty} \frac{1}{2^n} = \sum_{n=1}^{N} \frac{1}{n} + 1 \underset{N\to\infty}{\sim} \ln N$$

$$\text{et} \quad \sum_{n=1}^{N} \frac{n}{a_1 + \cdots + a_n} \leqslant \sum_{n=1}^{+\infty} \frac{n}{a_1 + \cdots + a_n} \leqslant K \sum_{n=1}^{+\infty} \frac{1}{a_n}.$$

Puisque $\sum_{n=1}^{N} \frac{n}{a_1 + \cdots + a_n} = \sum_{n=1}^{N} \frac{2}{n+1}$, on en déduit

$$\sum_{n=1}^{N} \frac{2}{n+1} \leqslant K \sum_{n=1}^{+\infty} \frac{1}{a_n}, \text{ c'est-à-dire } \frac{2}{K} \leqslant \frac{\sum_{n=1}^{+\infty} \frac{1}{a_n}}{\sum_{n=1}^{N} \frac{1}{n+1}} \underset{N\to\infty}{\sim} \frac{\ln N}{\ln N} = 1.$$

On en déduit que $2 \leqslant K$.

Conclusion. On a pour toute suite $(a_n)_{n\geqslant 1}$ de réels strictement positifs telle que $\sum_{n=1}^{+\infty} \frac{1}{a_n}$ converge l'inégalité

$$\boxed{\sum_{n=1}^{+\infty} \frac{n}{a_1 + \cdots + a_n} \leqslant 2 \sum_{n=1}^{+\infty} \frac{1}{a_n}.}$$

et 2 est la meilleure constante possible. ◁

On peut se demander si un résultat analogue existe pour la moyenne arithmétique : la réponse est non puisque, si l'on prend $a_n = \frac{1}{n^2}$ *pour* $n \geqslant 1$*, la série* $\sum a_n$ *converge mais la moyenne*

$$\frac{a_1 + \cdots + a_n}{n} \underset{n\to\infty}{\sim} \frac{\pi^2}{6n}$$

n'est pas le terme général d'une série convergente.

Terminons maintenant par deux exercices portant sur l'étude de l'action sur une série d'une permutation de ses termes. Dans le premier exercice, on étudie la situation lorsque la série est absolument convergente : on constate que la nouvelle série est encore convergente et de même somme (notons qu'il s'agit là d'un résultat de cours sur les familles sommables) ; inversement, si $\sum u_{\sigma(n)}$ converge pour toute permutation σ, alors la série $\sum u_n$ est absolument convergente. Dans le second exercice, on prouvera un résultat troublant sur les séries semi-convergentes réelles : par une permutation idoine, il est possible d'obtenir une somme arbitrairement fixée (théorème de Riemann).

3.47. Convergence commutative

1. Soit (a_n) une suite de nombres complexes. Établir l'équivalence

$$\sum |a_n| \text{ converge } \iff \forall \sigma \in \mathcal{S}(\mathbb{N}),\ \sum a_{\sigma(n)} \text{ converge.}$$

2. Soit (c_n) une suite de nombres complexes non nuls. Établir l'équivalence

$$\sum |c_n - 1| \text{ converge } \iff \exists a \in \mathbb{C}^*, \forall \sigma \in \mathcal{S}(\mathbb{N}),\ \lim_{n\to+\infty} \prod_{k=0}^{n} c_{\sigma(k)} = a.$$

(**École normale supérieure**)

▷ **Solution.**

1. • Supposons que $\sum |a_n|$ converge. Soit $\sigma \in \mathcal{S}(\mathbb{N})$. Alors, pour tout N de $\mathbb{N}$, $\sum_{n=0}^{N} |a_{\sigma(n)}| \leqslant \sum_{n=0}^{+\infty} |a_n|$. On en déduit que $\sum |a_{\sigma(n)}|$ converge et donc que $\sum a_{\sigma(n)}$ converge.

Montrons, de plus, ce qui nous sera utile pour la suite de l'exercice, que $\sum_{n=0}^{+\infty} a_{\sigma(n)} = \sum_{n=0}^{+\infty} a_n$. Soit $\varepsilon > 0$ et un entier n_0 tel que $\sum_{n=n_0}^{+\infty} |a_n| \leqslant \varepsilon$. Posons $n_1 = \max_{k\in[\![0,n_0]\!]} \sigma^{-1}(k)$. On a alors $\sigma(n) > n_0$ pour $n > n_1$. Pour $N > n_0, n_1$, on a

$$\left| \sum_{n=0}^{N} a_{\sigma(n)} - \sum_{n=0}^{N} a_n \right| \leqslant 2 \sum_{n=n_0}^{+\infty} |a_n| \leqslant 2\varepsilon,$$

car les termes de la suite (a_n) d'indice inférieur à n_0 sont présents

dans les deux sommes. En faisant tendre N vers $+\infty$, on obtient $\left|\sum_{n=0}^{+\infty} a_{\sigma(n)} - \sum_{n=0}^{+\infty} a_n\right| \leqslant 2\varepsilon$ pour tout $\varepsilon > 0$, d'où l'égalité des sommes.

• Supposons maintenant que $\sum |a_n|$ diverge. Posons, pour tout entier n, $x_n = \mathrm{Re}(a_n)$ et $y_n = \mathrm{Im}(a_n)$. Puisque, pour tout n, $|a_n| = \sqrt{x_n^2 + y_n^2} \leqslant |x_n| + |y_n|$, on en déduit que $\sum |x_n|$ ou $\sum |y_n|$ diverge.

Montrons que si $(u_n)_{n\in\mathbb{N}}$ est une suite réelle telle que $\sum |u_n|$ diverge, alors il existe $\sigma \in \mathcal{S}(\mathbb{N})$ tel que $\sum u_{\sigma(n)}$ diverge. Ceci nous montrera qu'on peut trouver $\sigma \in \mathcal{S}(\mathbb{N})$ tel que $\sum x_{\sigma(n)}$ ou $\sum y_{\sigma(n)}$ diverge. On pourra conclure que $\sum a_{\sigma(n)}$ diverge.

Si (u_n) est de signe constant à partir d'un certain rang, $\sum u_n$ diverge et $\sigma = \mathrm{Id}_{\mathbb{N}}$ convient. Sinon, il y a une infinité de termes strictement positifs et une infinité de termes strictement négatifs. Posons, pour $n \in \mathbb{N}$, $u_n^+ = \max(u_n, 0)$ et $u_n^- = \max(-u_n, 0)$. Puisque que $|u_n| = u_n^+ + u_n^-$, l'une des deux séries $\sum u_n^+$ et $\sum u_n^-$ diverge. Soit $(u_{\varphi(n)})_{n\in\mathbb{N}}$ la suite extraite de $(u_n)_{n\in\mathbb{N}}$ comportant les termes de la suite positifs ou nuls et $(u_{\psi(n)})_{n\in\mathbb{N}}$ la suite extraite de $(u_n)_{n\in\mathbb{N}}$ comportant les termes strictement négatifs. De ce qui précède, on déduit que $\sum u_{\varphi(n)}$ ou $\sum u_{\psi(n)}$ diverge. En effet, pour tout $\mathrm{N} \in \mathbb{N}$, $\sum_{n=0}^{\mathrm{N}} u_{\varphi(n)} = \sum_{n=0}^{\varphi(\mathrm{N})} u_n^+$ et $\sum_{n=0}^{\mathrm{N}} u_{\psi(n)} = -\sum_{n=0}^{\psi(\mathrm{N})} u_n^-$. Supposons que $\sum u_{\varphi(n)}$ diverge. Alors $\lim_{\mathrm{N}\to+\infty} \sum_{n=0}^{\mathrm{N}} u_{\varphi(n)} = +\infty$. On définit une suite d'entiers strictement croissante $(n_k)_{k\in\mathbb{N}^*}$ de la façon suivante :

* n_1 est le plus petit entier tel $u_{\psi(0)} + \sum_{j=0}^{n_1} u_{\varphi(j)} > 1$; n_1 existe car $\lim_{\mathrm{N}\to+\infty} \sum_{n=0}^{\mathrm{N}} u_{\varphi(n)} = +\infty$;

* $n_1, n_2, \ldots, n_k$ étant déterminés, n_{k+1} est le plus petit entier supérieur à n_k tel que $u_{\psi(k)} + \sum_{j=n_k+1}^{n_{k+1}} u_{\varphi(j)} > 1$; n_{k+1} existe pour les mêmes raisons que n_1.

Définissons maintenant σ, élément de $\mathcal{S}(\mathbb{N})$. On pose pour $n \in [\![0, n_1]\!]$, $\sigma(n) = \varphi(n)$ et $\sigma(n_1 + 1) = \psi(0)$, puis pour tout $k \in \mathbb{N}^*$, et pour $n \in [\![n_k + k + 1, n_{k+1} + k]\!]$, $\sigma(n) = \varphi(n - k)$ et $\sigma(n_{k+1} + k + 1) = \psi(k)$.

Pour tout $k \in \mathbb{N}$, σ réalise une bijection de $[\![n_k + k + 1, n_{k+1} + k]\!]$ sur $\varphi([\![n_k + 1, n_{k+1}]\!])$; σ réalise donc une bijection de $\mathbb{N} \setminus \{n_k + k, k \in \mathbb{N}^*\}$ sur $\varphi(\mathbb{N})$. D'autre part, σ réalise une bijection de $\{n_k + k, k \in \mathbb{N}^*\}$ sur $\psi(\mathbb{N})$. Puisque $\{\varphi(\mathbb{N}), \psi(\mathbb{N})\}$ est une partition de $\mathbb{N}$, σ est un élément de $\mathcal{S}(\mathbb{N})$.

Enfin, on a pour tout $k \in \mathbb{N}^*$, $\sum_{n=0}^{n_k+k+1} u_{\sigma(n)} \geqslant k$, donc la série $\sum u_{\sigma(n)}$

diverge. Si c'est $\sum u_{\psi(n)}$ qui diverge, il suffit de considérer la suite $(-u_n)_{n\in\mathbb{N}}$ et de lui appliquer la démonstration qui précède pour montrer l'existence de $\sigma \in \mathcal{S}$ tel que $\sum u_{\sigma(n)}$ diverge.

2. Pour tout $n \in \mathbb{N}$, $c_n \neq 0$. On note ρ_n le module de c_n et θ_n la mesure principale de l'argument de c_n ($\theta_n \in]-\pi, \pi]$). On a alors

$$|c_n - 1|^2 = |\rho_n e^{i\theta_n} - 1|^2 = \rho_n^2 + 1 - 2\rho_n \cos\theta_n$$
$$= (\rho_n - 1)^2 + 2\rho_n(1 - 2\cos\theta_n) = (\rho_n - 1)^2 + 4\rho_n \sin^2\frac{\theta_n}{2}.$$

• Supposons que $\sum |c_n - 1|$ converge. Pour σ élément de $\mathcal{S}(\mathbb{N})$, on pose, pour tout $N \in \mathbb{N}$,

$$P_N = \prod_{n=0}^{N} c_{\sigma(n)} = \prod_{n=0}^{N} \rho_{\sigma(n)} \prod_{n=0}^{N} e^{i\theta_{\sigma(n)}} = \left(\prod_{n=0}^{N} \rho_{\sigma(n)}\right) \exp\left(i \sum_{n=0}^{N} \theta_{\sigma(n)}\right).$$

Il résulte de l'expression de $|c_n - 1|^2$ que, pour tout $n \in \mathbb{N}$,

$$|\rho_n - 1| \leqslant |c_n - 1| \quad \text{et} \quad 2\sqrt{\rho_n}\left|\sin\frac{\theta_n}{2}\right| \leqslant |c_n - 1|.$$

On en déduit que $\sum |\rho_n - 1|$ converge et que $\sum 2\sqrt{\rho_n}\left|\sin\dfrac{\theta_n}{2}\right|$ converge. On a donc $\lim\limits_{n\to+\infty} \rho_n = 1$ et $\lim\limits_{n\to+\infty} \sin\dfrac{\theta_n}{2} = 0$. Sachant que $\theta_n \in]-\pi, \pi]$, on en déduit que $\lim\limits_{n\to+\infty} \theta_n = \lim\limits_{n\to+\infty} 2\arcsin\left(\sin\dfrac{\theta_n}{2}\right) = 0$, puis que $2\sqrt{\rho_n}\left|\sin\dfrac{\theta_n}{2}\right| \underset{n\to\infty}{\sim} \left|\dfrac{\theta_n}{2}\right|$. Ceci montre que $\sum |\theta_n|$ converge.

De la question 1, on déduit que pour tout $\sigma \in \mathcal{S}(\mathbb{N})$, les séries $\sum(\rho_{\sigma(n)} - 1)$ et $\sum \theta_{\sigma(n)}$ convergent et que de plus $\Phi = \sum\limits_{n=0}^{+\infty} \theta_{\sigma(n)}$ est indépendant du choix de σ. On a $\lim\limits_{n\to+\infty} e^{i\left(\sum\limits_{n=0}^{N} \theta_{\sigma(n)}\right)} = e^{i\Phi}$.

Posons $R_N = \prod\limits_{n=0}^{N} \rho_{\sigma(n)}$; alors $\ln R_N = \sum\limits_{n=0}^{N} \ln \rho_{\sigma(n)}$. La suite $(\rho_n)_{n\in\mathbb{N}}$ converge vers 1 ; on en déduit que $|\ln \rho_{\sigma(n)}| \underset{n\to\infty}{\sim} |\rho_{\sigma(n)} - 1|$, puis que $\sum |\ln \rho_{\sigma(n)}|$ converge et que $\sum\limits_{n=0}^{+\infty} \ln \rho_{\sigma(n)} = S$ est indépendant du choix de σ. On obtient $\lim\limits_{N\to+\infty} R_N = e^S$. Finalement $\lim\limits_{n\to+\infty} P_N = e^S e^{i\Phi} = a \neq 0$. De plus a est indépendant de σ.

• Réciproquement, supposons qu'il existe $a \in \mathbb{C}^*$ tel que , pour tout $\sigma \in \mathcal{S}(\mathbb{N})$, $\lim\limits_{n\to+\infty} \prod\limits_{k=0}^{n} c_{\sigma(k)} = a$. Notons R le module de a et Φ la mesure principale de l'argument de a.

Fixons $\sigma \in \mathcal{S}(\mathbb{N})$ et reprenons les notations précédentes. On a $\lim_{N\to+\infty} R_N = \lim_{N\to+\infty} |P_N| = |a| = R$. On en déduit que $\lim_{N\to+\infty} \sum_{n=0}^{N} \ln \rho_{\sigma(n)} = \ln R$. On a donc montré que, pour tout $\sigma \in \mathcal{S}(\mathbb{N})$, $\sum \ln \rho_{\sigma(n)}$ converge. On en déduit que $\sum |\ln \rho_n|$ converge. On a alors $\lim_{n\to+\infty} \rho_n = 1$ et donc $|\rho_n - 1| \underset{n\to\infty}{\sim} |\ln \rho_n|$. On a donc également $\sum |\rho_n - 1|$ convergente.

Toujours pour $\sigma \in \mathcal{S}(\mathbb{N})$ fixé, posons $\Phi_N = \sum_{n=0}^{N} \theta_{\sigma(n)}$ $(N \in \mathbb{N})$. On a

$$\lim_{N\to+\infty} e^{i\Phi_N} = \lim_{n\to+\infty} \frac{P_N}{|P_N|} = \frac{a}{R} = e^{i\Phi} \quad \text{et}$$

$$\lim_{n\to+\infty} e^{i\theta_{\sigma(n)}} = \lim_{n\to+\infty} \frac{e^{i\Phi_n}}{e^{i\Phi_{n-1}}} = 1.$$

Sachant que $|e^{i\theta_{\sigma(n)}} - 1| = 2\sin\dfrac{\theta_{\sigma(n)}}{2}$ et que $\theta_{\sigma(n)} \in]-\pi, \pi]$, on en déduit que $\lim_{n\to+\infty} \theta_{\sigma(n)} = \lim_{n\to+\infty} 2\arcsin\left(\sin\dfrac{\theta_{\sigma(n)}}{2}\right) = 0$.

Nous savons que $\lim_{N\to+\infty} e^{i(\Phi_N - \Phi)} = 1$. Pour tout $N \in \mathbb{N}$, il existe $k_N \in \mathbb{Z}$ tel que $\Phi_N - \Phi - 2k_N\pi \in]-\pi, \pi]$. Sachant que $\lim_{N\to+\infty} e^{i(\Phi_N - \Phi - 2k_N\pi)} = 1$, on montre, comme précédemment, que

$$\lim_{N\to+\infty} \Phi_N - \Phi - 2k_N\pi = 0.$$

On a alors $\lim_{N\to+\infty} \Phi_N - \Phi_{N-1} - 2\pi(k_N - k_{N-1}) = 0$. Sachant que $\lim_{N\to+\infty} \Phi_N - \Phi_{N-1} = \lim_{N\to+\infty} \theta_{\sigma(N)} = 0$, on en déduit que $\lim_{N\to+\infty} k_N - k_{N-1} = 0$. (k_N) étant une suite d'entiers, elle est constante, égale à k, pour N assez grand. On a donc $\lim_{N\to+\infty} \Phi_N - \Phi - 2k\pi = 0$. La suite (Φ_N) converge, c'est-à-dire que $\sum \theta_{\sigma(n)}$ converge. Cela étant vrai pour tout $\sigma \in \mathcal{S}(\mathbb{N})$, on en déduit que $\sum |\theta_n|$ converge.

Reprenant l'égalité $|c_n - 1|^2 = (\rho_n - 1)^2 + 4\rho_n \sin^2\dfrac{\theta_n}{2}$, on en déduit

$$|c_n - 1| \leqslant |\rho_n - 1| + 2\rho_n \left|\sin\frac{\theta_n}{2}\right|.$$

Or $\sum |\rho_n - 1|$ converge et $2\rho_n \left|\sin\dfrac{\theta_n}{2}\right| \underset{n\to\infty}{\sim} |\theta_n|$ donc $\sum \rho_n \left|\sin\dfrac{\theta_n}{2}\right|$ converge, et finalement $\sum |c_n - 1|$ converge. ◁

3.48. Théorème de Riemann

Soit $\sum a_n$ une série réelle semi-convergente et $\alpha \in \mathbb{R}$. Montrer qu'il existe une permutation σ de $\mathbb{N}$ telle que la série $\sum a_{\sigma(n)}$ soit convergente de somme α.

(École polytechnique)

Solution.

• La série $\sum a_n$ n'est pas absolument convergente, donc la famille $(a_n)_{n\geqslant 0}$ n'est pas sommable. On note A (resp. B) l'ensemble des indices $n \geqslant 0$ tels que $a_n \geqslant 0$ (resp. $a_n < 0$). L'ensemble $\mathbb{N}$ est réunion disjointe de A et de B.

La partie A est infinie sinon, à partir d'un certain rang les a_n seraient négatifs et la convergence de la série équivaudrait à l'absolue convergence. Or $\sum a_n$ est semi-convergente. De même, B est infini.

• Les familles $(a_n)_{n\in A}$ et $(a_n)_{n\in B}$ ne sont pas sommables. En effet, supposons par exemple $(a_n)_{n\in A}$ sommable. Alors si on pose pour $n \geqslant 0$ $a'_n = 0$ si $a_n < 0$ et $a'_n = a_n$ si $a_n \geqslant 0$, la série $\sum a'_n$ est absolument convergente et donc convergente. La série $\sum(a_n - a'_n)$ est donc convergente et même absolument convergente puisque $a_n - a'_n$ est de signe constant négatif. Comme $a_n = (a_n - a'_n) + a'_n$, a_n est la somme de termes généraux de deux séries absolument convergentes. Il s'ensuit que $\sum a_n$ est absolument convergente, ce qui est contraire à l'hypothèse.

De même, on montre que $(a_n)_{n\in B}$ n'est pas sommable.

On va construire $\sigma(n)$ par récurrence sur n. On pose $\sigma(0) = 0$ et pour $n \geqslant 1$, deux cas se présentent :

(*i*) si $\sum_{k=0}^{n-1} a_{\sigma(k)} \leqslant \alpha$, on va ajouter un terme positif : on prend pour $\sigma(n)$ le plus petit des entiers k de A distincts de $\sigma(0), \ldots, \sigma(n-1)$ (ce plus petit élément est bien défini).

(*ii*) si $\sum_{k=0}^{n-1} a_{\sigma(k)} > \alpha$, on va ajouter un terme négatif : on prend pour $\sigma(n)$ le plus petit des entiers k de B distincts de $\sigma(0), \ldots, \sigma(n-1)$.

L'application σ est par construction injective. Montrons qu'elle est surjective. Imaginons un instant qu'il existe $N \in \mathbb{N}$ qui ne soit pas dans l'image de σ. Supposons par exemple $N \in A$. Les $\sigma(k)$ dans A sont tous inférieurs à N et ils sont donc en nombre fini. Il existe donc un rang n_0 à partir duquel tous les $\sigma(n)$ sont dans B. Par conséquent, si $n \geqslant n_0$, on obtient

$$\sum_{k=0}^{n-1} a_{\sigma(k)} > \alpha$$

Comme cette série est à termes négatifs (sauf pour un nombre fini

de termes), et que ses sommes partielles sont minorées, elle converge. La famille $(a_{\sigma(n)})_{n\in\mathbb{N}}$ est donc sommable. Mais, à un nombre fini de termes près, la famille des $a_{\sigma(n)}$ est la famille $(a_n)_{n\in B}$ qui, elle, n'est pas sommable. La contradiction permet de conclure que σ est bijective.

• Il reste à prouver que $\sum a_{\sigma(n)}$ converge et est de somme α. Soit $\varepsilon > 0$. Comme la série $\sum a_n$ converge, a_n tend vers 0. On a également $a_{\sigma(n)}$ convergente vers 0. En effet, il existe $n_1 \geqslant 0$ tel que si $n \geqslant n_1$, $|a_n| \leqslant \varepsilon$. Par injectivité de σ, il n'y a qu'un nombre fini de $n \in \mathbb{N}$ tel que $\sigma(n) \leqslant n_1$. Soit donc $n_0 > \max\{k, \sigma(k) \leqslant n_1\}$. Si $n \geqslant n_0$, $\sigma(n) \geqslant n_1$ et $|a_{\sigma(n)}| \leqslant \varepsilon$.

• Il existe un rang $N \geqslant n_0$ tel que $\sigma(N) \in A$ et $\sigma(N+1) \in B$ car on a vu que les $\sigma(n)$ ne pouvaient rester dans A (resp. B).

Posons pour $n \geqslant 0$, $S_n = \sum_{k=0}^{n} a_{\sigma(k)}$. Montrons que si $n \geqslant N$, $|S_n - \alpha| \leqslant \varepsilon$. Si $\sigma(N) \in A$, c'est que $S_{N-1} \leqslant \alpha$ et si $\sigma(N+1) \in B$, c'est que $S_N > \alpha$. Or $|S_N - S_{N-1}| = |a_{\sigma(N)}| \leqslant \varepsilon$. Donc S_N appartient à $[\alpha - \varepsilon, \alpha + \varepsilon]$.

Soit $n > N$. Imaginons que $S_n > \alpha + \varepsilon$. Comme on passe de S_{n-1} à S_n d'un saut de longueur $a_{\sigma(n)}$ inférieur ou égal à ε, on ne peut avoir $S_{n-1} \leqslant \alpha$.

Cela entraîne $a_{\sigma(n)} < 0$ et finalement $S_{n-1} \geqslant S_n$. Mais alors de proche en proche, on obtient $\alpha + \varepsilon < S_n \leqslant S_{n-1} \leqslant S_{n-2} \leqslant \cdots \leqslant S_N$. On obtient une contradiction. Donc $S_n \leqslant \alpha + \varepsilon$.

Imaginons que $S_n < \alpha - \varepsilon$. Comme on passe de S_{n-1} à S_n d'un saut de longueur $a_{\sigma(n)}$ inférieur ou égal à ε, on ne peut avoir $S_{n-1} > \alpha$. Cela entraîne $a_{\sigma(n)} \geqslant 0$ et finalement $S_{n-1} \leqslant S_n$. Mais alors de proche en proche, on obtient $\alpha - \varepsilon > S_n \geqslant S_{n-1} \geqslant S_{n-2} \geqslant \cdots \geqslant S_N$. C'est encore une contradiction et $S_n \geqslant \alpha - \varepsilon$.

Ainsi, pour $n \geqslant N$, $|S_n - \alpha| \leqslant \varepsilon$.

Conclusion. On a donc construit une bijection σ de $\mathbb{N}$ telle que $\sum a_{\sigma(n)}$ converge et

$$\boxed{\sum_{n=0}^{+\infty} a_{\sigma(n)} = \alpha.} \lhd$$

Chapitre 4

Fonctions d'une variable réelle

À partir du XVII^{e} *siècle, le développement du calcul infinitésimal, motivé par de nombreux problèmes de cinématique, de mécanique, ou de calcul des variations, fait de la «fonction» l'objet central des mathématiques modernes, alors que jusque là, le «nombre» était la base de l'édifice mathématique. Nous devons à Bernoulli et Leibniz le terme même de «fonction» : pour Bernoulli (1698), une fonction de la variable x est «une quantité formée d'une manière quelconque à partir de x et de constantes». L'écriture $y = f(x)$ est introduite par Euler en 1734. Les fonctions sont représentées par des courbes dans le plan et Euler se demande si une courbe donnée correspond toujours à une fonction. C'est lui qui distingue les courbes continues, des courbes discontinues, qui sont le plus souvent, à cette époque, des graphes de fonctions continues par morceaux. Cette double conception des fonctions, comme expressions analytiques ou comme graphes du plan, ne sera pas vraiment éclaircie avant le* XIX^{e} *siècle (c'est Dirichlet qui donnera la définition moderne d'une fonction comme correspondance ; il proposera ainsi (en 1837) un exemple de fonction discontinue partout, la fonction χ définie par $\chi(x) = 1$ pour x rationnel et $\chi(x) = 0$ pour x irrationnel). Lagrange, cherchant à établir les fondements de l'Analyse, s'en tient au point de vue formel et refuse de se référer à toute notion de limite. Ces hésitations empêchent les mathématiciens du* XVIII^{e} *siècle de mener jusqu'à leur achèvement certains de leurs travaux, comme l'étude de l'équation des cordes vibrantes. C'est la génération suivante, avec entre autres Gauss, Cauchy, Bolzano et Abel, qui donnera dans la première moitié du* XIX^{e} *siècle un statut rigoureux aux notions de convergence, de continuité... Quant au concept de limite d'une fonction numérique, on doit sans doute sa première définition précise à Weierstrass. Les premiers exercices concernent cette dernière notion.*

4.1. Applications propres

Soit $f : \mathbb{R} \to \mathbb{R}$ continue. Montrer qu'il y a équivalence entre les assertions suivantes :

(*i*) $\lim_{x \to \pm\infty} |f(x)| = +\infty$;

(*ii*) pour tout compact $K \subset \mathbb{R}$, $f^{-1}(K)$ est compact.

(École polytechnique)

▷ **Solution.**

• Supposons que $\lim_{x \to \pm\infty} |f(x)| = +\infty$. Soit K un compact. Comme K est fermé et f continue, $f^{-1}(K)$ est fermé. S'il n'est pas compact, c'est qu'il n'est pas borné. Dans ce cas, il existe une suite $(x_n)_{n\in\mathbb{N}}$ de $f^{-1}(K)$ telle que $\lim_{n \to +\infty} |x_n| = +\infty$. Mais alors $\lim_{n \to +\infty} |f(x_n)| = +\infty$ ce qui est impossible puisque $(f(x_n))_{n\in\mathbb{N}}$ est une suite de K, donc est bornée.

• Réciproquement, supposons (*ii*). Soit A un réel positif. Par hypothèse, $f^{-1}([-A, A])$ est compact. En particulier, il existe $M \geqslant 0$ tel que $f^{-1}([-A, A]) \subset [-M, M]$. Ainsi pour tout réel x vérifiant $|x| > M$, on a $|f(x)| > A$. Cela exprime exactement que $\lim_{x \to \pm\infty} |f(x)| = +\infty$. ◁

Une application qui vérifie ces conditions est dite propre.

4.2. Zéros d'une combinaison linéaire d'exponentielles

Soit $\omega_1, \ldots, \omega_n$ des réels et $a_1, \ldots, a_n$ des complexes. On suppose que la fonction $F : t \in \mathbb{C} \longmapsto \sum_{k=1}^{n} a_k e^{\omega_k t}$ est non nulle. Montrer que l'ensemble des parties réelles des zéros de F est borné.

(École polytechnique)

▷ **Solution.**

En regroupant les différents termes correspondant à une même valeur de ω_k, on peut se ramener au cas où les ω_k sont distincts. On peut même supposer que $\omega_1 < \cdots < \omega_n$ et qu'aucun des a_k n'est nul (par hypothèse, il en existe au moins un non nul).

Si les a_k sont tous réels et si on se restreint à t réel le résultat est assez clair. En effet, quand t tend vers $+\infty$ le terme $a_n e^{\omega_n t}$ domine tous les autres de sorte que $F(t) \underset{+\infty}{\sim} a_n e^{\omega_n t}$. En particulier $F(t) \neq 0$ pour t assez grand. De même on a $F(t) \underset{-\infty}{\sim} a_1 e^{\omega_1 t}$ de sorte que $F(t)$ ne s'annule pas non plus sur un voisinage de $-\infty$.

Dans le cas général, l'idée est exactement la même. Pour $t \in \mathbb{C}$ on pose $x = \mathrm{Re}(t)$. On a alors pour tout $\omega \in \mathbb{R}$, $|e^{\omega t}| = e^{\omega x}$. Si on écrit

$$\frac{\mathrm{F}(t)}{a_n e^{\omega_n t}} = 1 + \sum_{k=1}^{n-1} \frac{a_k}{a_n} e^{(\omega_k - \omega_n)t}.$$

on remarque que

$$\left| \sum_{k=1}^{n-1} \frac{a_k}{a_n} e^{(\omega_k - \omega_n)t} \right| \leqslant \sum_{k=1}^{n-1} \frac{|a_k|}{|a_n|} e^{(\omega_k - \omega_n)x} \xrightarrow[x \to +\infty]{} 0$$

car les réels $\omega_k - \omega_n$ sont tous < 0. La dernière expression est indépendante de la partie imaginaire de t. On peut donc trouver $\mathrm{A} \in \mathbb{R}$ tel que, pour tout $t \in \mathbb{C}$ tel que $x = \mathrm{Re}(t) \geqslant \mathrm{A}$, on ait $\left| \sum_{k=1}^{n-1} \frac{a_k}{a_n} e^{(\omega_k - \omega_n)t} \right| \leqslant \frac{1}{2}$ et donc $\left| \frac{\mathrm{F}(t)}{a_n e^{\omega_n t}} \right| \geqslant \frac{1}{2}$ grâce à l'inégalité triangulaire. *A fortiori*, on a $\mathrm{F}(t) \neq 0$ si $x \geqslant \mathrm{A}$. En mettant $a_1 e^{\omega_1 t}$ en facteur et en faisant tendre x vers $-\infty$, on démontre de même qu'il existe un réel B tel que $\mathrm{F}(t) \neq 0$ si $x \leqslant \mathrm{B}$. Les zéros de la fonction F ont donc une partie réelle qui appartient au segment $[\mathrm{A}, \mathrm{B}]$. ◁

Bien entendu, dans les cas concrets, un des outils essentiels pour calculer les limites est le développement limité. En voici un exemple.

4.3. Un calcul de limite

Calculer la limite lorsque x tend vers 0^+ de $\dfrac{x^{\mathrm{sh}\,x} - (\mathrm{sh}\,x)^x}{(\sin x)^x - x^{\sin x}}$.
(**École polytechnique**)

▷ **Solution.**

Il s'agit d'une forme indéterminée du type $\frac{0}{0}$: les deux termes du numérateur et les deux termes du dénominateur tendent vers 1 quand x tend vers zéro. On va chercher un équivalent du numérateur et du dénominateur. On a $(\mathrm{sh}\,x)^x = \exp(x \ln \mathrm{sh}\,x)$ et

$$\begin{aligned} x \ln \mathrm{sh}(x) &= x \ln(x + o(x)) = x(\ln x + \ln(1 + o(1))) \\ &= x(\ln x + o(1)) = x \ln x + o(x) \xrightarrow[x \to 0^+]{} 0. \end{aligned}$$

Comme $\mathrm{sh}\,x \ln x \underset{0^+}{\sim} x \ln x$ tend également vers 0, on peut écrire pour le numérateur

$$x^{\mathrm{sh}\,x} - (\mathrm{sh}\,x)^x = (\mathrm{sh}\,x)^x (e^{\ln x\,\mathrm{sh}\,x - x \ln \mathrm{sh}\,x} - 1) \underset{0^+}{\sim} \ln x\,\mathrm{sh}\,x - x \ln \mathrm{sh}\,x,$$

car $(\mathrm{sh}\,x)^x$ tend vers 1 par continuité de l'exponentielle et $e^u - 1$ est équivalent à u en 0. De plus, si on pose $u(x) = \ln x\,\mathrm{sh}\,x - x \ln \mathrm{sh}\,x$, on a

$$\begin{aligned} u(x) &= \left(x + \frac{x^3}{6} + o(x^3)\right) \ln x - x \ln \left(x + \frac{x^3}{6} + o(x^3)\right) \\ &= x \ln x + \frac{x^3 \ln x}{6} + o(x^3 \ln x) - x \ln x - x \ln \left(1 + \frac{x^2}{6} + o(x^2)\right) \\ &= \frac{x^3 \ln x}{6} + o(x^3 \ln x). \end{aligned}$$

On procède de même pour le dénominateur. On a également $x \ln \sin x$ et $\sin x \ln x$ qui tendent vers 0 lorsque x tend vers 0^+, si bien que

$$(\sin x)^x - x^{\sin x} \underset{0^+}{\sim} x \ln \sin x - \ln x \sin x = v(x)$$

et

$$\begin{aligned} v(x) &= x \ln \left(x - \frac{x^3}{6} + o(x^3)\right) - \ln x \left(x - \frac{x^3}{6} + o(x^3)\right) \\ &= x \ln x + x \ln \left(1 - \frac{x^2}{6} + o(x^2)\right) - x \ln x + \frac{x^3 \ln x}{6} + o(x^3 \ln x) \\ &= \frac{x^3 \ln x}{6} + o(x^3 \ln x). \end{aligned}$$

On conclut que $\dfrac{x^{\operatorname{sh} x} - (\operatorname{sh} x)^x}{(\sin x)^x - x^{\sin x}} \underset{x \to 0}{\sim} \dfrac{\frac{x^4 \ln x}{6}}{\frac{x^3 \ln x}{6}} = 1$ et donc

$$\boxed{\lim_{x \to 0+} \frac{x^{\operatorname{sh} x} - (\operatorname{sh} x)^x}{(\sin x)^x - x^{\sin x}} = 1.} \quad \triangleleft$$

L'exercice suivant est à rapprocher de l'exercice 2.3 qui concerne les suites sous-additives.

4.4. Fonctions sur-additives

Soit $f : \mathbb{R}_+ \to \mathbb{R}$ telle que

(i) pour tout $(s,t) \in \mathbb{R}_+^2$, $f(s+t) \geqslant f(s) + f(t)$;

(ii) il existe $M \geqslant 0$ tel que pour tout $t \in \mathbb{R}_+$, $|f(t)| \leqslant Mt$.

Montrer l'existence de $\alpha = \lim\limits_{t \to 0} \dfrac{f(t)}{t}$ et de $\beta = \lim\limits_{t \to +\infty} \dfrac{f(t)}{t}$.

Comparer $f(t)$ avec αt et βt pour $t \in \mathbb{R}_+$.

(École polytechnique)

▷ **Solution.**

On remarque que pour tout entier $n \in \mathbb{N}^*$ et tout réel $t > 0$ on a $f(nt) = f(\underbrace{t + \cdots + t}_{n \text{ fois}}) \geqslant nf(t)$ et donc $\dfrac{f(nt)}{nt} \geqslant \dfrac{f(t)}{t}$. Si $\lim\limits_{t \to +\infty} \dfrac{f(t)}{t} = \beta$, alors, en faisant tendre n vers $+\infty$, on obtient, pour tout $t > 0$, $\beta \geqslant \dfrac{f(t)}{t}$ et donc $\beta \geqslant \sup\limits_{t>0} \dfrac{f(t)}{t}$. L'inégalité dans l'autre sens étant évidente, on a nécessairement $\beta = \sup\limits_{t>0} \dfrac{f(t)}{t}$. En partant de l'inégalité $\dfrac{f(t)}{t} \geqslant \dfrac{f\left(\frac{t}{n}\right)}{\frac{t}{n}}$, on montre de même que la limite en 0 ne peut être que $\inf\limits_{t>0} \dfrac{f(t)}{t}$.

Pour tout $t > 0$, on a $\dfrac{|f(t)|}{t} \leqslant M$. Donc $\alpha = \inf\limits_{t>0} \dfrac{f(t)}{t}$ et $\beta = \sup\limits_{t>0} \dfrac{f(t)}{t}$ existent dans $\mathbb{R}$. Montrons que $\lim\limits_{t \to +\infty} \dfrac{f(t)}{t} = \beta$ et que $\lim\limits_{t \to 0} \dfrac{f(t)}{t} = \alpha$.

Soit $\varepsilon > 0$. Il existe $u > 0$ tel que $\beta - \varepsilon \leqslant \dfrac{f(u)}{u} \leqslant \beta$. Montrons que pour t assez grand, $\beta - 2\varepsilon \leqslant \dfrac{f(t)}{t} \leqslant \beta$. L'inégalité de droite est vérifiée par définition de β. Pour tout $t > 0$, écrivons $t = nu + r$ avec $n \in \mathbb{N}$ et $0 \leqslant r < u$ (cette écriture est unique et n est la partie entière de $\dfrac{t}{u}$). On obtient

$$\frac{f(t)}{t} = \frac{f(nu + r)}{t} \geqslant \frac{f(nu) + f(r)}{t} \geqslant \frac{nf(u)}{t} + \frac{f(r)}{t}.$$

On a $\lim\limits_{t \to +\infty} \dfrac{n}{t} = \lim\limits_{t \to +\infty} \dfrac{\mathrm{E}\left(\frac{t}{u}\right)}{t} = \dfrac{1}{u}$ et $\lim\limits_{t \to +\infty} \dfrac{f(r)}{t} = 0$, puisque on a $|f(r)| \leqslant Mr \leqslant Mu$. On en déduit que

$$\lim_{t \to +\infty} \left(\frac{nf(u)}{t} + \frac{f(r)}{t} \right) = \frac{f(u)}{u} \geqslant \beta - \varepsilon.$$

Donc pour t assez grand, $\dfrac{f(t)}{t} \geqslant \dfrac{nf(u)}{t} + \dfrac{f(r)}{t} \geqslant \beta - 2\varepsilon$. D'où le résultat annoncé.

On vient de traiter la limite en $+\infty$. En 0 la démarche est analogue. Soit $\varepsilon > 0$. Il existe $u > 0$ tel que $\alpha \leqslant \dfrac{f(u)}{u} \leqslant \alpha + \varepsilon$. Montrons que pour t assez proche de 0, $\alpha \leqslant \dfrac{f(t)}{t} \leqslant \alpha + 2\varepsilon$. Il n'y a qu'une inégalité à démontrer. Soit $t < u$. Cette fois-ci, on divise u par t : $u = nt + r$ avec $n \in \mathbb{N}$ et $0 \leqslant r < t$. Alors on a $f(u) \geqslant f(nt) + f(r) \geqslant nf(t) + f(r)$ et donc

$$\frac{f(t)}{t} \leqslant \frac{f(u)}{nt} - \frac{f(r)}{nt} \leqslant \frac{f(u)}{u - r} + \frac{Mr}{u - r}.$$

Quand t tend vers 0, r tend vers 0 et

$$\lim_{t\to 0} \frac{f(u)}{u-r} + \frac{Mr}{u-r} = \frac{f(u)}{u} \leqslant \alpha + \varepsilon.$$

On a bien, pour t assez proche de 0,

$$\frac{f(t)}{t} \leqslant \frac{f(u)}{u-r} + \frac{Mr}{u-r} \leqslant \alpha + 2\varepsilon,$$

ce qui est le résultat annoncé.

Enfin, par définition de α et β, on a $\alpha t \leqslant f(t) \leqslant \beta t$ pour tout $t > 0$. ◁

Une fonction réglée est une fonction qui a en tout point une limite à gauche et à droite. Dans l'exercice suivant, on démontre qu'une telle fonction n'a pas trop de discontinuités. Indiquons que la solution fait appel à la propriété de Borel-Lebesgue.

4.5. Points de discontinuité d'une fonction réglée

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction réglée, c'est-à-dire possédant en tout point une limite à droite et une limite à gauche. Montrer que l'ensemble des points de discontinuité de f est au plus dénombrable.

(École polytechnique)

▷ **Solution.**

Nous noterons, pour tout $a \in \mathbb{R}$, $\varphi(a)$ la limite de f à droite en a et $\psi(a)$ la limite à gauche de f en a. Nous commençons par démontrer le résultat sur un intervalle compact I. Cela suffira car $\mathbb{R}$ peut s'écrire comme une réunion dénombrable de segments.

La fonction f est discontinue en a si $\varphi(a) \neq f(a)$ ou $\psi(a) \neq f(a)$. Pour tout $\varepsilon > 0$, considérons

$$D_\varepsilon = \{x \in I, |\varphi(x) - f(x)| \geqslant \varepsilon \text{ ou } |\psi(x) - f(x)| \geqslant \varepsilon\}.$$

Pour tout $a \in I$, il existe $\eta > 0$ tel que

$$\forall x \in]a, a+\eta[,\ |f(x)-\varphi(a)| \leqslant \frac{\varepsilon}{3} \quad \text{et} \quad \forall x \in]a-\eta, a[,\ |f(x)-\psi(a)| \leqslant \frac{\varepsilon}{3}.$$

Si on prend x et x' dans $]a, a+\eta[$, on a donc $|f(x) - f(x')| \leqslant \frac{2\varepsilon}{3}$. On en déduit, en faisant tendre x' vers x par valeurs supérieures ou inférieures que, pour tout $x \in]a, a+\eta[$,

$$|\varphi(x) - f(x)| \leqslant \frac{2\varepsilon}{3} \quad \text{et} \quad |\psi(x) - f(x)| \leqslant \frac{2\varepsilon}{3}.$$

On peut faire exactement le même raisonnement sur $]a-\eta, a[$ et on trouve finalement, en posant $\mathrm{I}_a =]a-\eta, a+\eta[$ que, pour tout $x \in \mathrm{I}_a \setminus \{a\}$, on a

$$|\varphi(x) - f(x)| \leqslant \frac{2\varepsilon}{3} \quad \text{et} \quad |\psi(x) - f(x)| \leqslant \frac{2\varepsilon}{3}.$$

On en déduit que a est le seul point de I_a qui peut appartenir à D_ε.

Pour tout $a \in \mathrm{I}$, on peut trouver un tel intervalle I_a. Les intervalles ouverts I_a (pour a variant dans I) recouvrent manifestement I. D'après le théorème de Borel-Lebesgue, I est recouvert par un nombre fini de tels intervalles. Soit $a_1, a_2, \ldots, a_k$ tels que $\mathrm{I} \subset \bigcup_{i=1}^{k} \mathrm{I}_{a_i}$. D'après ce qui précède D_ε contient au plus $a_1, a_2, \ldots, a_k$. Il est donc fini. En particulier, pour tout $n \in \mathbb{N}^*$, $\mathrm{D}_{\frac{1}{n}}$ est fini. Or il est clair que l'ensemble des points de discontinuité de f sur I est égal à $\bigcup_{n\in\mathbb{N}^*} \mathrm{D}_{\frac{1}{n}}$. On en déduit que cet ensemble est au plus dénombrable, car réunion d'une famille dénombrable d'ensembles finis.

Passons maintenant à l'ensemble des points de discontinuité de f sur $\mathbb{R}$. Notons $\mathcal{D}$ cet ensemble et, pour tout $n \in \mathbb{N}^*$, $\mathcal{D}_n$ l'ensemble des points de discontinuité de f sur $[-n, n]$. Il résulte de ce qui précède que chaque $\mathcal{D}_n$ est au plus dénombrable. On en déduit que $\mathcal{D}$, qui est une réunion dénombrable d'ensembles au plus dénombrables, est également au plus dénombrable, ce qui est le résultat que nous voulions démontrer. ◁

Le lecteur pourra montrer qu'une fonction réglée sur un segment $[a, b]$ est limite uniforme sur $[a, b]$ d'une suite (f_n) de fonctions en escalier (cela apparaît pratiquement dans la solution ci-dessus). Cela donne un autre éclairage du résultat de l'exercice : comme une fonction en escalier n'a qu'un nombre fini de points de discontinuité, il existe un ensemble dénombrable D *tel que toutes les fonctions f_n soient continues sur l'ensemble $[a, b] \setminus \mathrm{D}$. Le théorème de continuité des limites uniformes montre que f est continue sur $[a, b] \setminus \mathrm{D}$.*

Cela permet aussi de construire très facilement la théorie de l'intégrale pour les fonctions réglées. On montre que si (f_n) est une suite de fonctions en escalier qui converge uniformément vers f sur $[a, b]$, alors la suite des intégrales $\int_a^b f_n$ est une suite de Cauchy donc converge. On prouve alors que sa limite ne dépend pas de la suite (f_n) choisie, mais seulement de f et par définition on dit qu'il s'agit de $\int_a^b f$.

Une fonction monotone est bien entendu réglée d'après le théorème des limites monotones. Dans l'énoncé suivant on montre que n'importe quel ensemble dénombrable est l'ensemble des points de discontinuité d'une telle fonction.

4.6. Discontinuités d'une fonction monotone

Soit A une partie dénombrable de $\mathbb{R}$. Montrer l'existence d'une fonction $f : \mathbb{R} \longmapsto \mathbb{R}$, monotone, dont A est l'ensemble des points de discontinuité.

(École polytechnique)

▷ **Solution.**

Soit $n \longmapsto a_n$ une bijection de $\mathbb{N}^*$ sur A. Pour tout $n \in \mathbb{N}^*$, notons f_n la fonction définie par $f_n(x) = \begin{cases} 0 & \text{si } x \leqslant a_n \\ \dfrac{1}{n^2} & \text{si } x > a_n. \end{cases}$ La fonction f_n est croissante et continue en tout point de $\mathbb{R}$ distinct de a_n. Posons pour tout réel x, $f(x) = \sum_{n=1}^{\infty} f_n(x)$.

On a pour tout $n \in \mathbb{N}^*$ et tout réel x, $|f_n(x)| \leqslant \dfrac{1}{n^2}$, donc la série de fonctions $\sum f_n$ converge normalement et f est définie sur $\mathbb{R}$. Chacune des fonctions f_n est croissante donc f est croissante.

Soit $x_0 \in \mathbb{R} \setminus \mathrm{A}$. Montrons que f est continue en x_0. Soit $\varepsilon > 0$ et $\mathrm{N} \in \mathbb{N}$ tel que $\sum_{n=\mathrm{N}+1}^{+\infty} \dfrac{1}{n^2} \leqslant \varepsilon$. On a, pour tout $x \in \mathbb{R}$,

$$|f(x) - f(x_0)| \leqslant \sum_{n=1}^{\mathrm{N}} |f_n(x) - f_n(x_0)| + 2\varepsilon.$$

Soit $\alpha = \min_{1 \leqslant n \leqslant \mathrm{N}} |x_0 - a_n| > 0$. Si $|x - x_0| < \alpha$, le segment $[x_0, x]$ ne contient aucun point a_n pour $1 \leqslant n \leqslant \mathrm{N}$, donc les fonctions $f_1, \ldots, f_{\mathrm{N}}$ sont constantes sur $[x_0, x]$. On a donc, pour $|x - x_0| < \alpha$, $|f(x) - f(x_0)| \leqslant 2\varepsilon$. La fonction f est continue en x_0.

Soit $x_0 \in \mathrm{A}$, $x_0 = a_k$. Montrons que f n'est pas continue en x_0. On définit ε et N comme dans le premier cas, en supposant de plus $\mathrm{N} \geqslant k$. Par inégalité triangulaire, on a, pour tout réel x,

$$|f(x) - f(x_0)| \geqslant \left| \sum_{n=1}^{\mathrm{N}} f_n(x) - f_n(x_0) \right| - 2\varepsilon.$$

Soit $\alpha = \min_{\substack{1 \leqslant n \leqslant \mathrm{N} \\ n \neq k}} |a_n - x_0| > 0$. Si $|x - x_0| \leqslant \alpha$, le segment $[x_0, x]$ ne contient aucun a_n avec $1 \leqslant n \leqslant \mathrm{N}$ et $k \neq n$. On a donc pour $|x - x_0| < \alpha$, $|f(x) - f(x_0)| \geqslant |f_k(x) - f_k(x_0)| - 2\varepsilon$. On a $f_k(x_0) = f_k(a_k) = 0$ et pour $x > x_0$, $f_k(x) = \dfrac{1}{k^2}$. On en déduit que, pour $x_0 < x < x_0 + \alpha$, on a $|f(x) - f(x_0)| \geqslant \dfrac{1}{k^2} - 2\varepsilon$. En choisissant $\varepsilon < \dfrac{1}{2k^2}$, on voit que $f(x) - f(x_0)$

ne tend pas vers 0 quand x tend vers x_0 : la fonction f n'est pas continue en x_0. L'ensemble des points de discontinuité de f est A. $\lhd$

La caractérisation séquentielle de la continuité en un point affirme qu'une fonction f est continue en a si et seulement si pour toute suite x_n qui converge vers a, la suite $f(x_n)$ converge vers $f(a)$. Quelle notion de continuité obtient-on en remplaçant la convergence des suites par la convergence en moyenne ? C'est ce qu'étudie l'exercice suivant.

4.7. Continuité au sens de Cesàro

On dit que $f : \mathbb{R} \to \mathbb{R}$ est continue au sens de Cesàro en a si pour toute suite $(x_n) \in \mathbb{R}^{\mathbb{N}}$ on a

$$\frac{x_1 + x_2 + \cdots + x_n}{n} \to a \Longrightarrow \frac{f(x_1) + f(x_2) + \cdots + f(x_n)}{n} \to f(a).$$

Déterminer les fonctions continues au sens de Cesàro sur $\mathbb{R}$.

(École polytechnique)

$\rhd$ **Solution.**

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue au sens de Cesàro sur $\mathbb{R}$. Quitte à retrancher à f la constante $f(0)$ on peut supposer que $f(0) = 0$. Soit x, y deux réels. Considérons la suite (u_n) qui prend alternativement les valeurs x et y : $u_{2p} = x$ et $u_{2p+1} = y$ pour tout $p \in \mathbb{N}$. Il est aisé de voir que (u_n) converge en moyenne vers $\dfrac{x+y}{2}$ et de même que $\lim\limits_{n\to+\infty} \dfrac{f(u_1) + \cdots + f(u_n)}{n} = \dfrac{f(x) + f(y)}{2}$. Grâce à l'hypothèse, on a donc

$$f\left(\frac{x+y}{2}\right) = \frac{f(x) + f(y)}{2}$$

et cela vaut pour tout couple $(x, y) \in \mathbb{R}^2$. En prenant $y = 0$ on a donc $f\left(\dfrac{x}{2}\right) = \dfrac{f(x)}{2}$ pour tout réel x. Par suite f est additive, c'est-à-dire vérifie $f(x + y) = f(x) + f(y)$ pour tout couple (x, y). Il est classique d'en déduire que f est $\mathbb{Q}$-linéaire. On a donc $f(r) = f(1)r$ pour tout rationnel r. Si x est un réel quelconque on peut trouver une suite de rationnels $(r_n)_{n\geqslant 1}$ qui converge vers x. Par le théorème de Cesàro cette suite converge aussi vers x en moyenne et on a donc

$$f(x) = \lim_{n\to+\infty} \frac{f(r_1) + \cdots + f(r_n)}{n} = \lim_{n\to+\infty} f(1)\frac{r_1 + \cdots + r_n}{n} = f(1)x.$$

On en déduit que f est linéaire. Par conséquent on a montré qu'une fonction continue au sens de Cesàro sur $\mathbb{R}$ est nécessairement une fonction

affine. Réciproquement, il est clair que toutes les fonctions affines sont continues au sens de Cesàro sur $\mathbb{R}$. $\lhd$

Les exercices qui suivent concernent les propriétés globales des fonctions continues. Dans les premiers énoncés interviennent deux propriétés topologiques fondamentales des fonctions continues, à savoir le fait que l'image par une fonction continue d'un compact est un compact (Weierstrass énonce notamment en 1861 qu'une fonction continue sur un segment et à valeurs réelles est bornée et admet un maximum et un minimum) et le théorème des valeurs intermédiaires. Ce dernier résultat fut longtemps considéré comme une évidence, Euler et Gauss l'utilisant sans plus de justifications. C'est Bolzano qui entreprit d'en donner une preuve en 1817.

Les trois questions de l'énoncé suivant sont indépendantes.

4.8. Existence d'un point fixe

1. Soit I un segment de $\mathbb{R}$ et $f : \mathrm{I} \to \mathbb{R}$ une fonction continue. Montrer que si $f(\mathrm{I}) \subset \mathrm{I}$ ou si $\mathrm{I} \subset f(\mathrm{I})$, alors f admet un point fixe dans I.

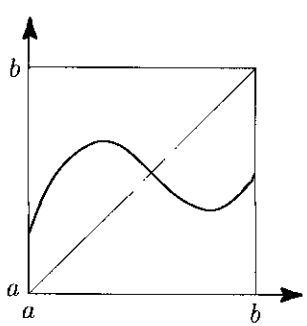
2. Soit $f : [0,1] \to [0,1]$ une fonction croissante. Montrer que f admet un point fixe.

3. Soit $f, g : [0,1] \to [0,1]$ deux fonctions continues vérifiant $f \circ g = g \circ f$. Montrer qu'il existe $x \in [0,1]$ tel que $f(x) = g(x)$.

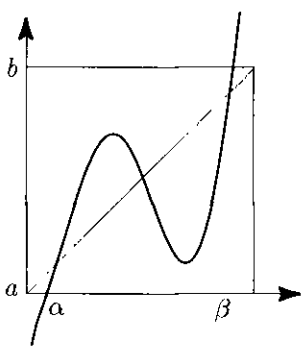
(**École normale supérieure**)

$\rhd$ **Solution.**

1. Posons $\mathrm{I} = [a,b]$ et $g(x) = f(x) - x$. Si $f(\mathrm{I}) \subset \mathrm{I}$ on a $g(a) \geqslant 0$, $g(b) \leqslant 0$ et, comme g est continue, le théorème des valeurs intermédiaires garantit l'existence de $x \in [a,b]$ tel que $g(x) = 0$.



$f(\mathrm{I}) \subset \mathrm{I}$



$\mathrm{I} \subset \mathrm{F}(\mathrm{I})$

Si maintenant $I \subset f(I)$, il existe $(\alpha, \beta) \in I^2$ tel que $f(\alpha) = a$ et $f(\beta) = b$. Alors $g(\alpha) \leqslant 0$ et $g(\beta) \geqslant 0$ de sorte que g s'annule dans ce cas aussi.

2. Posons $A = \{x \in [0,1],\ f(x) \leqslant x\}$. C'est une partie non vide ($1 \in A$) et minorée par 0. Soit $\alpha = \inf A$. On va montrer que $f(\alpha) = \alpha$. Soit $x \in A$. On a $\alpha \leqslant x$ donc $f(\alpha) \leqslant f(x) \leqslant x$ par croissance de f. Il en résulte que $f(\alpha)$ minore A donc que $f(\alpha) \leqslant \alpha$. Toujours par croissance de f on a $f(f(\alpha)) \leqslant f(\alpha)$. Ainsi, $f(\alpha) \in A$ et on a donc $\alpha \leqslant f(\alpha)$. D'où l'égalité.

3. Supposons par l'absurde que $g - f$ ne s'annule pas et par exemple qu'elle est strictement positive sur $[0,1]$. Continue, elle atteint son minimum sur le compact $[0,1]$ et il existe donc $\varepsilon > 0$ tel que pour tout $x \in [0,1]$, $g(x) - f(x) \geqslant \varepsilon$. On a alors pour tout x,

$$g^2(x) = g(g(x)) \geqslant f(g(x)) + \varepsilon = g(f(x)) + \varepsilon \geqslant f^2(x) + 2\varepsilon$$

puis, par une récurrence immédiate, $g^n(x) \geqslant f^n(x) + n\varepsilon$ pour tout $n \geqslant 1$ et tout $x \in [0,1]$. C'est impossible car, si $x \in [0,1]$, les suites $(g^n(x))_{n\geqslant 0}$ et $(f^n(x))_{n\geqslant 0}$ sont bornées. ◁

Le résultat suivant est aussi connu sous le nom de «Universal chord theorem». Il montre que si le graphe d'une fonction continue sur $\mathbb{R}$ admet une corde de longueur 1, alors il admet des cordes de longueur $1/n$ pour tout entier naturel non nul n (une corde étant par définition un segment horizontal qui joint deux points du graphe). Ce résultat est optimal au sens où si $\lambda \in]0,1]$ n'est pas l'inverse d'un entier, il existe des fonctions continues ayant une corde de longueur 1 mais pas de corde de longueur λ.

4.9. Un théorème de Paul Lévy (1934)

Soit $E = \{f \in \mathcal{C}^0([0,1], \mathbb{R}),\ f(0) = f(1)\}$. Si $f \in E$, on pose

$$A(f) = \{\sigma \in \mathbb{R}_+,\ \exists x \in [0,1],\ f(x+\sigma) = f(x)\}$$

1. Montrer que $\bigcap_{f\in E} A(f)$ est compact et que

$$\{0\} \cup \left\{\frac{1}{p},\ p \in \mathbb{N}^*\right\} \subset \bigcap_{f\in E} A(f).$$

2. Montrer que pour tout $f \in E$, il existe $\varepsilon > 0$, tel que

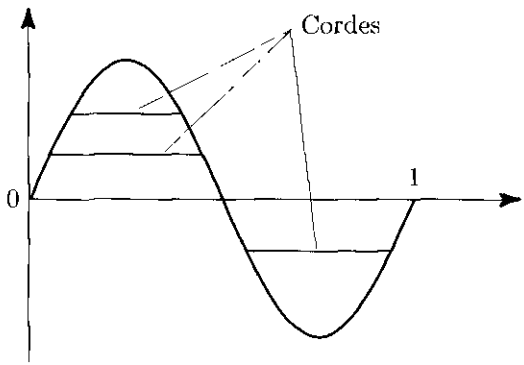
$$[0,\varepsilon] \cup \left\{\frac{1}{p},\ p \in \mathbb{N}^*\right\} \subset A(f).$$

3. Montrer que $\bigcap_{f\in E} A(f) = \{0\} \cup \left\{\frac{1}{p},\ p \in \mathbb{N}^*\right\}$.

(École normale supérieure)

▷ **Solution.**

1. L'ensemble $A(f)$ peut se voir comme l'ensemble des longueurs des «cordes» horizontales du graphe de f. Par exemple, dans le cas de la fonction $f : x \longmapsto \sin 2\pi x$, on a visiblement $A(f) = [0, \frac{1}{2}] \cup \{1\}$:

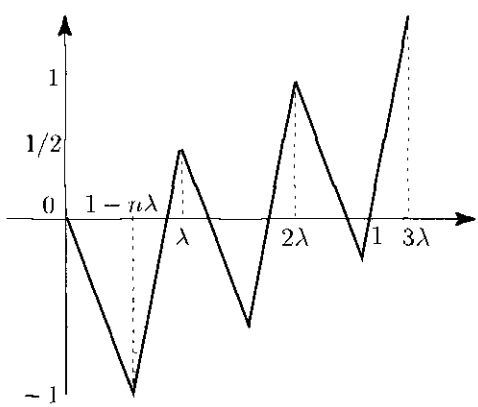


Il est clair que $0 \in A(f)$ pour tout $f \in E$ et que $A(f) \subset [0,1]$. C'est donc une partie bornée de $\mathbb{R}$. Montrons que $A(f)$ est fermé. Soit (σ_n) une suite de points de $A(f)$ qui converge vers σ. Pour chaque entier n, il existe x_n dans $[0,1]$ tel que $f(x_n + \sigma_n) = f(x_n)$. Comme $[0,1]$ est compact, on peut extraire de (x_n) une sous-suite $(x_{\varphi(n)})$ qui converge vers un point $x \in [0,1]$. Par continuité de f, on a $f(x+\sigma) = f(x)$ et $\sigma \in A(f)$. Donc $A(f)$ est un compact de $\mathbb{R}$. Comme l'intersection d'une famille de compacts est encore compacte, $\bigcap_{f\in E} A(f)$ est un compact.

Soit $p \in \mathbb{N}^*$ et f dans E. Montrons que $\frac{1}{p} \in A(f)$. Considérons pour cela la fonction g définie sur $\left[0, 1 - \frac{1}{p}\right]$ par $g(x) = f\left(x + \frac{1}{p}\right) - f(x)$. Comme $g(0) + g\left(\frac{1}{p}\right) + \cdots + g\left(\frac{p-1}{p}\right) = f(1) - f(0) = 0$, la fonction g s'annule nécessairement sur $\left[0, 1 - \frac{1}{p}\right]$ (sinon elle garderait un signe constant et la somme précédente ne serait pas nulle). Donc $\frac{1}{p} \in A(f)$ et on a bien $\{0\} \cup \left\{\frac{1}{p},\ p \in \mathbb{N}^*\right\} \subset \bigcap_{f\in E} A(f)$.

2. Soit $f \in \mathrm{E}$. Quitte à retrancher $f(0)$ à f, on peut supposer que $f(0) = f(1) = 0$. Si f est de signe constant sur $[0,1]$, on voit que $\mathrm{A}(f) = [0,1]$. En effet, si $\sigma \in [0,1]$, la fonction $f(x+\sigma) - f(x)$ prend en 0 et $1-\sigma$ des valeurs de signe opposé, donc s'annule. Le même raisonnement montre que si $\mathrm{I} = [a,b]$ est un intervalle de $[0,1]$ sur lequel f garde un signe constant avec $a < b$ et $f(a) = f(b) = 0$, alors $[0, b-a] \subset \mathrm{A}(f)$ et $\varepsilon = b-a$ répond à la question. Un tel intervalle existe nécessairement. Si f est nulle, c'est clair. Sinon, elle prend par exemple une valeur strictement positive et l'ensemble $\{x \in [0,1],\ f(x) > 0\}$ est un ouvert non vide. En particulier, il existe $\alpha < \beta$ tel que le segment $[\alpha, \beta]$ est entièrement contenu dans cet ensemble. La partie $\{x \in [0,\alpha], f(x) \leqslant 0\}$ (resp. $\{x \in [\beta, 1], f(x) \leqslant 0\}$) est non vide et majorée (resp. minorée). Si a (resp. b) désigne sa borne supérieure (resp. sa borne inférieure), f reste strictement positive sur $]a,b[$. De plus $f(a)$ et $f(b)$ sont nuls : en effet, a et b sont adhérents aux parties $\{x, f(x) > 0\}$ et $\{x, f(x) \leqslant 0\}$ et f est continue. *Pour obtenir l'intervalle $]a,b[$ sur lequel f est strictement positive et $f(a) = f(b) = 0$, on pouvait également considérer une composante connexe de l'ouvert non vide $\{x \in [0,1],\ f(x) > 0\}$ (cf. exercice 1.27 dans lequel cette notion est étudiée).*

3. La première question a montré une des deux inclusions. Pour prouver la seconde, il suffit de montrer que si $\lambda \in]0,1]$ n'est pas l'inverse d'un entier, il existe une fonction f de E telle que $\lambda \notin \mathrm{A}(f)$. Notons n la partie entière de $1/\lambda$. On a $n\lambda < 1 < n\lambda + \lambda$. On commence par se donner f sur l'intervalle $[0,\lambda]$. On choisit une fonction continue quelconque vérifiant $f(0) = 0$, $f(\lambda) = 1/n$ et $f(1 - n\lambda) = -1$. On prolonge alors f à $[0,1]$ en imposant $f(x+\lambda) = f(x) + 1/n$ pour tout x. Voici un exemple avec $n = 2$ où on choisit f affine par morceaux.



Par construction f n'a pas de corde de longueur λ puisque la différence $f(x+\lambda) - f(x)$ vaut toujours $1/n$. Il reste à vérifier deux choses. Tout

d'abord le fait que f est bien continue et enfin que $f(1) = 0$. Le premier point provient du fait que les raccords aux points $k\lambda$ pour $k = 0, 1, \ldots, n$ sont continus car $f(\lambda) = 1/n$. Enfin, $f(1) = f(1 - n\lambda + n\lambda) = f(1 - n\lambda) + 1 = 0$. ◁

Petite application pratique de l'exercice : montrer que, si un candidat à l'X effectue le 2000 m en 10 minutes, il existe un intervalle de temps de 5 minutes pendant lequel il parcourt 1000 m.

Il est naturel de se demander si la propriété des valeurs intermédiaires caractérise les fonctions continues. Ce n'est pas le cas comme le montre l'exemple de la fonction $f : x \mapsto \sin \frac{1}{x}$ prolongée par la valeur 0 en 0. L'image d'un intervalle I qui ne contient pas 0 est un intervalle (car f est continue sur I) et l'image d'un intervalle qui contient 0 et qui n'est pas réduit à $\{0\}$ est le segment $[-1, 1]$. Pourtant f n'est pas continue en 0. Dans l'énoncé suivant on ajoute une hypothèse permettant d'obtenir une réciproque.

4.10. Une réciproque au théorème des valeurs intermédiaires

Soit $f : \mathbb{R} \to \mathbb{R}$ telle que, pour tout segment $[a, b]$ de $\mathbb{R}$, $f([a, b])$ est un segment et telle que $f^{-1}(\{x\})$ est fermé pour tout $x \in \mathbb{R}$. Montrer que f est continue.

(École normale supérieure)

▷ **Solution.**

Soit $a \in \mathbb{R}$. Considérons le segment $\mathrm{K}_n = \left[a - \frac{1}{n}, a + \frac{1}{n}\right]$ pour tout $n \geqslant 1$. Son image directe $f(\mathrm{K}_n)$ s'écrit $[m_n, \mathrm{M}_n]$ avec $m_n \leqslant f(a) \leqslant \mathrm{M}_n$. On a $f(\mathrm{K}_{n+1}) \subset f(\mathrm{K}_n)$ pour tout n donc par le théorème des segments emboîtés, l'intersection $\mathrm{I} = \bigcap_{n \geqslant 1} f(\mathrm{K}_n) = \bigcap_{n \geqslant 1} [m_n, \mathrm{M}_n]$ est un segment $[m, \mathrm{M}]$ qui contient en particulier $f(a)$.

Montrons que I est en fait réduit à $\{f(a)\}$. Soit $y \in \mathrm{I}$. Pour tout $n \geqslant 1$, il existe alors $a_n \in \mathrm{K}_n$ tel que $f(a_n) = y$. La suite $(a_n)_{n \geqslant 1}$ converge vers a et $a_n \in f^{-1}(\{y\})$. Or, la partie $f^{-1}(\{y\})$ étant fermée, a est dans cette partie, autrement dit $f(a) = y$. On a donc $\mathrm{I} = \{f(a)\}$.

Soit $\varepsilon > 0$. Il existe $n_0 \in \mathbb{N}^*$ tel que pour tout $n \geqslant n_0$, $[m_n, \mathrm{M}_n] \subset [f(a) - \varepsilon, f(a) + \varepsilon]$. Dans ces conditions, pour tout $x \in \mathbb{R}$ avec $|x - a| \leqslant \frac{1}{n}$, on a $f(x) \in f(\mathrm{K}_n) = [m_n, \mathrm{M}_n]$ et $|f(x) - f(a)| \leqslant \varepsilon$: la fonction f est bien continue en a. ◁

Dans le contre-exemple que nous avons donné avant l'exercice on voit par exemple que l'ensemble $f^{-1}(\{1\})$ n'est pas fermé.

Si A *et* B *sont deux parties de* $\mathbb{R}$, *un homéomorphisme* $f : \mathrm{A} \to \mathrm{B}$ *est une bijection continue dont la réciproque est également continue. Si l'ensemble de départ est un intervalle* I, $f : \mathrm{I} \to \mathbb{R}$ *réalise un homéomorphisme de* I *sur* $f(\mathrm{I})$ *dès lors que* f *est continue et strictement monotone. Réciproquement, si* f *désigne un homéomorphisme défini sur un intervalle* I, f *est strictement monotone : en effet, l'ensemble* $\mathrm{T} = \{(x, y) \in \mathrm{I}^2,\ x < y\}$ *est convexe dans* $\mathbb{R}^2$ *et en particulier connexe. La fonction* F, *qui à* $(x, y) \in \mathrm{T}$ *associe* $\mathrm{F}(x, y) = \dfrac{f(y) - f(x)}{y - x}$, *est continue et ne s'annule pas par injectivité de* f. *Le théorème des valeurs intermédiaires garantit alors que* F *garde un signe constant, autrement dit, que* f *est strictement monotone.*

Il est aisé de vérifier que l'ensemble des homéomorphismes de $\mathbb{R}$ *(ou plus généralement d'un intervalle quelconque) est un groupe pour la composition. L'exercice suivant se propose d'en chercher les sous-groupes finis.*

4.11. Sous-groupes finis du groupe des homéomorphismes

Trouver les sous-groupes finis du groupe des homéomorphismes de $\mathbb{R}$.

(**École normale supérieure**)

▷ **Solution.**

Soit G un tel sous-groupe et n son cardinal. Si $f \in \mathrm{G}$ on a $f^n = \mathrm{Id}$ en vertu du théorème de Lagrange. Si f est strictement croissante, alors f est forcément l'identité. En effet, s'il existe un réel x tel que $f(x) > x$ alors on a, par croissance de f, $x = f^n(x) > f^{n-1}(x) > \cdots > f(x) > x$ ce qui est absurde. On voit de même qu'il ne peut exister de réel x tel que $f(x) < x$. Ainsi, hormis l'identité, toutes les fonctions de G sont des homéomorphismes strictement décroissants. Or, si $f \in \mathrm{G}$ est strictement décroissante, alors f^2 est strictement croissante donc vaut l'identité. Il en découle que tous les éléments de G sont d'ordre 2. De même si f et g sont deux éléments strictement décroissants de G alors $g \circ f$ est strictement croissante et donc $g = f^{-1} = f$. On en déduit que G contient au plus deux éléments.

Conclusion. La réciproque étant immédiate, on peut conclure que les sous-groupes finis du groupe des homéomorphismes de $\mathbb{R}$ sont $\{\mathrm{Id}\}$ et les groupes à deux éléments $\{\mathrm{Id}, f\}$, où f est une involution strictement décroissante. ◁

Dans l'exercice suivant on démontre que deux homéomorphismes de $[0, 1]$ *ayant pour uniques points fixes* 0 *et* 1 *sont conjugués.*

4.12. Conjugaison de deux homéomorphismes

On note H le groupe des homéomorphismes de $[0,1]$ sur lui-même, c'est-à-dire des bijections continues de $[0,1]$ sur $[0,1]$. Soit f et g deux éléments de H dont les seuls points fixes sont 0 et 1. Montrer qu'il existe $h \in \mathrm{H}$ tel que $f \circ h = h \circ g$.

(**École normale supérieure**)

▷ **Solution.**

• En termes de théorie des groupes, l'énoncé demande de montrer que f et g sont conjugués dans le groupe H. Les homéomorphismes f et g sont strictement monotones et vu les hypothèses, strictement croissants. Comme 0 et 1 sont les seuls points fixes de f, on a soit $f(x) > x$ pour tout $x \in]0,1[$, soit $f(x) < x$ pour tout $x \in]0,1[$. Si on est dans le second cas, on peut poser $\tilde{f}(x) = 1 - f(1-x) = (h^{-1} \circ f \circ h)(x)$ avec $h : x \longmapsto 1-x$. On a alors

$$\tilde{f}(x) = 1 - f(1-x) > 1 - (1-x) = x$$

pour tout $x \in]0,1[$ et $\tilde{f}$ est conjugué à f. Comme la conjugaison est une relation d'équivalence, on peut donc supposer que pour tout $x \in]0,1[$, $f(x) > x$ et $g(x) > x$.

• Choisissons $x_0 = y_0 = \dfrac{1}{2} \in]0,1[$ et posons, pour tout $k \in \mathbb{Z}$, $x_k = f^k(x_0)$ et $y_k = g^k(y_0)$. Les hypothèses sur f et g montrent que les suites $(x_k)_{k\in\mathbb{Z}}$ et $(y_k)_{k\in\mathbb{Z}}$ sont strictement croissantes et vérifient

$$\lim_{k\to-\infty} x_k = 0, \quad \lim_{k\to+\infty} x_k = 1, \quad \lim_{k\to-\infty} y_k = 0, \quad \lim_{k\to+\infty} y_k = 1.$$

Posons $\mathrm{I}_k = [x_k, x_{k+1}[$, $\mathrm{J}_k = [y_k, y_{k+1}[$: f induit un homéomorphisme de I_k sur I_{k+1}, et g induit un homéomorphisme de J_k sur J_{k+1}. Choisissons maintenant un homéomorphisme quelconque (par exemple affine) $h_0 : [y_0, y_1[\to [x_0, x_1[$ et posons, avec des notations un peu abusives, $h_k = f^k \circ h_0 \circ g^{-k}$: il est clair que h_k réalise un homéomorphisme de J_k dans I_k. Soit $h : [0,1] \to [0,1]$ définie par $h(0) = 0$, $h(1) = 1$, et $h(y) = h_k(y)$ si $y \in \mathrm{J}_k$. Alors h est un homéomorphisme strictement croissant de $[0,1]$ dans lui-même et, comme pour tout $k \in \mathbb{Z}$, $f \circ h_k = f^{k+1} \circ h \circ g^{-(k+1)} \circ g = h_{k+1} \circ g$ on a bien $f \circ h = h \circ g$. ◁

L'exercice suivant introduit un outil important dans l'étude des homéomorphismes du cercle, le nombre de rotation de Poincaré. Le lecteur se reportera aux commentaires qui suivent l'exercice.

4.13. Nombre de rotation de Poincaré (1885)

Soit $f : \mathbb{R} \to \mathbb{R}$ une application continue croissante, telle que pour tout x, $f(x+1) = f(x)+1$. On veut prouver que pour tout x, la suite $u_n(x) = \dfrac{f^n(x)}{n}$ admet une limite finie l indépendante de x (f^n désigne la n-ième itérée de f).

1. Soit $(x, y) \in \mathbb{R}^2$. Montrer qu'il existe $k \in \mathbb{N}$ tel que pour tout $n \in \mathbb{N}$, $|f^n(x) - f^n(y)| \leqslant k$. Que peut-on en déduire ?

2. Montrer que pour tout $(n, m) \in \mathbb{N}^2$,

$$f^n(0) + f^m(0) - 1 \leqslant f^{n+m}(0) \leqslant f^n(0) + f^m(0) + 1.$$

3. En déduire que la suite $(u_n(0))_{n\geqslant 0}$ converge et vérifier que sa limite ℓ appartient à $[f(0) - 1, f(0) + 1]$. Conclure.

(**École normale supérieure**)

▷ **Solution.**

1. L'entier k doit convenir pour $n = 0$. Choisissons donc un entier k tel que $|y - x| \leqslant k$. On a $-k + x \leqslant y \leqslant k + x$. En utilisant la croissance de f et la 1-périodicité de $t \longmapsto f(t) - t$, on a

$$-k + f(x) = f(-k + x) \leqslant f(y) \leqslant f(k + x) = f(x) + k,$$

et par une récurrence immédiate $-k + f^n(x) \leqslant f^n(y) \leqslant f^n(x) + k$ pour tout entier naturel n. Il en résulte que $|u_n(x) - u_n(y)| \leqslant \dfrac{k}{n}$ et donc que $\lim\limits_{n\to\infty} |u_n(x) - u_n(y)| = 0$.

Par conséquent, il suffit de montrer que la suite $(u_n(x))_{n\geqslant 0}$ converge pour une valeur particulière de x, par exemple 0.

2. Fixons un entier n, et notons a la partie entière de $f^n(0)$. On a donc $a \leqslant f^n(0) < a + 1$. Comme la fonction f^m est croissante, on obtient $f^m(a) \leqslant f^{n+m}(0) \leqslant f^m(a) + 1$. Or, puisque a est un entier, $f^m(a) = f^m(0) + a$. On obtient donc

$$\begin{aligned} f^m(0) + f^n(0) - 1 \leqslant f^m(0) + a = f^m(a) &\leqslant f^{n+m}(0) \\ &\leqslant f^m(0) + a + 1 \leqslant f^m(0) + f^n(0) + 1. \end{aligned}$$

3. En prenant $m = n$, on en déduit que

$$2f^n(0) - 1 \leqslant f^{2n}(0) \leqslant 2f^n(0) + 1.$$

Une récurrence facile montre que

$$mf^n(0) - (m - 1) \leqslant f^{mn}(0) \leqslant mf^n(0) + (m - 1).$$

En divisant par nm, on en déduit que $|u_{nm}(0) - u_n(0)| \leqslant \dfrac{m-1}{nm} \leqslant \dfrac{1}{n}$. Par symétrie on a aussi $|u_{nm}(0) - u_m(0)| \leqslant \dfrac{1}{n}$ et donc, par inégalité triangulaire, $|u_m(0) - u_n(0)| \leqslant \dfrac{1}{n} + \dfrac{1}{m}$. Il en résulte que la suite $(u_n(0))$ est une suite de Cauchy et donc qu'elle converge. Comme on a, de plus, $|u_m(0) - u_1(0)| \leqslant 1$, on obtient en faisant tendre m vers l'infini.

$$|\ell - f(0)| \leqslant 1.$$

La question 1 montre alors que $u_n(x)$ converge aussi vers ℓ pour tout réel x. ◁

Lorsque l'application f est strictement croissante elle induit un homéomorphisme «croissant» $\tilde{f}$ du cercle unité de $\mathbb{C}$ par

$$\forall \theta \in \mathbb{R}, \quad \tilde{f}(e^{2i\pi\theta}) = e^{2i\pi f(\theta)}$$

Le qualificatif «croissant» traduit le fait que $f(x)$ tourne sur le cercle dans le même sens que x. Par exemple, si on prend pour f une translation $t_a : x \longmapsto x + a$, $\tilde{f}$ est la rotation d'angle $2\pi a$. Inversement, tout homéomorphisme croissant du cercle peut être obtenu de cette manière à partir d'un homéomorphisme strictement croissant $f : \mathbb{R} \to \mathbb{R}$ vérifiant $f(x+1) = f(x) + 1$ pour tout x. Celui-ci est d'ailleurs unique à une constante entière près. Le lecteur pourra prouver ce «théorème de relèvement» à titre d'exercice.

La limite l obtenue dans l'exercice, prise modulo 1, est par définition le nombre de rotation de Poincaré de l'homéomorphisme $\tilde{f}$, introduit par celui-ci en 1885. Par exemple, le nombre de rotation de la rotation $\tilde{t}_a$ est égal à $a \pmod 1$. Notons $r(\tilde{f})$ ce nombre. Le lecteur pourra montrer que $r(\tilde{f}) \equiv 0$ si et seulement si $\tilde{f}$ admet un point fixe, et que $r(\tilde{f}) \in \mathbb{Q}$ si et seulement si $\tilde{f}$ admet un point périodique. Ces deux assertions se vérifient immédiatement dans le cas des rotations.

Les prochains exercices concernent quelques équations fonctionnelles. Il est bien connu que les seuls morphismes continus du groupe $(\mathbb{R}, +)$ dans lui-même sont les homothéties $x \longmapsto ax$. Comme $(\mathbb{R}, +)$ est isomorphe et homéomorphe à $(\mathbb{R}_+^*, \times)$ via le logarithme, on en déduit directement que les morphismes continus de $(\mathbb{R}, +)$ dans $(\mathbb{R}_+^*, \times)$ sont les exponentielles ($x \longmapsto a^x$, $a > 0$), que les morphismes continus de $(\mathbb{R}_+^*, \times)$ dans $(\mathbb{R}_+^*, \times)$ sont les fonctions puissances ($x \longmapsto x^\alpha$, $\alpha \in \mathbb{R}$), que les morphismes continus de $(\mathbb{R}_+^*, \times)$ dans $(\mathbb{R}, +)$ sont les fonctions logarithmes ($x \longmapsto \log_a(x)$ avec $a > 0$).

Il existe des morphismes discontinus de $(\mathbb{R}, +)$ dans lui-même, mais ils sont «indescriptibles» (leur existence résulte de l'axiome du choix). Dès qu'on cherche les morphismes qui ont une propriété raisonnable de

plus (par exemple monotones ou bornés sur un voisinage de 0) on ne trouve plus que les homothéties. C'est la situation de l'exercice suivant, qui s'intéresse aux morphismes qui ont la propriété de respecter l'inverse.

4.14. Morphismes du groupe $(\mathbb{R}, +)$ respectant l'inverse

Que dire d'une application f de $\mathbb{R}$ dans $\mathbb{R}$ vérifiant $f(1) = 1$ et

$$\forall (x, y) \in \mathbb{R}^2, \quad f(x+y) = f(x) + f(y) \text{ et } \forall x \in \mathbb{R}^*, \quad f\left(\frac{1}{x}\right) = \frac{1}{f(x)} ?$$

On pourra commencer par démontrer que f est bornée sur un voisinage de 0 en utilisant la fonction $x \longmapsto x + \dfrac{1}{x}$

(École normale supérieure)

Solution.

Comme f est un morphisme de $(\mathbb{R}, +)$ dans $(\mathbb{R}, +)$, on a classiquement pour tout $x \in \mathbb{R}$ et tout $r \in \mathbb{Q}$, $f(rx) = rf(x)$ (c'est-à-dire que f est $\mathbb{Q}$-linéaire). Redémontrons ce fait. C'est vrai pour $r \in \mathbb{Z}$ puisque f est un morphisme de groupes. Si $r = \dfrac{p}{q} \in \mathbb{Q}$ avec $p \in \mathbb{Z}$ et $q \in \mathbb{N}^*$, on a

$$qf(rx) = f(qrx) = f(px) = pf(x),$$

et donc $f(rx) = rf(x)$. En particulier, pour tout $r \in \mathbb{Q}$, $f(r) = rf(1) = r$.

On va essayer d'exploiter la seconde propriété vérifiée par f pour prouver que f est l'identité. Suivons l'indication de l'énoncé et considérons la fonction $\varphi : x \longmapsto x + \dfrac{1}{x}$. Une rapide étude des variations de φ montre que $\varphi(\mathbb{R}^*) =]-\infty, -2] \cup [2, +\infty[$. Par hypothèse, on a pour tout $x \in \mathbb{R}^*$,

$$\begin{aligned} |f(\varphi(x))| &= \left| f\left(x + \frac{1}{x}\right) \right| = \left| f(x) + f\left(\frac{1}{x}\right) \right| \\ &= \left| f(x) + \frac{1}{f(x)} \right| = |\varphi(f(x))| \geqslant 2. \end{aligned}$$

Comme tout y tel que $|y| \geqslant 2$ peut s'écrire $y = x + \dfrac{1}{x}$, on en déduit que pour tout $y \in \mathbb{R}^*$,

$$|y| \geqslant 2 \Longrightarrow |f(y)| \geqslant 2.$$

En passant à l'inverse, on obtient pour $x \in \mathbb{R}^*$,

$$|x| \leqslant \frac{1}{2} \Longrightarrow \left|\frac{1}{x}\right| \geqslant 2 \Longrightarrow |f(x)| = \frac{1}{\left|f\left(\frac{1}{x}\right)\right|} \leqslant \frac{1}{2}.$$

Autrement dit, f est bornée sur l'intervalle $[-1/2, 1/2]$. Cette information va être suffisante pour montrer que f est l'identité. Comme dans la recherche des morphismes continus, on utilise la densité de $\mathbb{Q}$. Soit x un réel quelconque, et $(r_n)_{n\in\mathbb{N}^*}$ une suite de rationnels telle que, pour tout $n \in \mathbb{N}^*$, $|x - r_n| \leqslant \dfrac{1}{2n}$. On a alors, pour tout $n \in \mathbb{N}^*$,

$$|f(x) - r_n| = |f(x) - f(r_n)| = |f(x - r_n)| = \frac{1}{n}|f(n(x - r_n))| \leqslant \frac{1}{2n}$$

(car $|n(x - r_n)| \leqslant \dfrac{1}{2}$) et donc $|f(x) - x| \leqslant \dfrac{1}{n}$. En passant à la limite, on en déduit que $f(x) = x$.

Conclusion. L'application f est l'identité. ◁

Dans l'exercice suivant on se ramène aux morphismes continus de $(\mathbb{R}, +)$ dans lui-même.

4.15. Une équation fonctionnelle (1)

Déterminer les applications continues $f : \mathbb{R} \to \mathbb{R}$ telles que pour tout $(x, y) \in \mathbb{R}^2$, $f(\sqrt{x^2 + y^2}) = f(x)f(y)$.

(École polytechnique)

▷ **Solution.**

• *Analyse.* La fonction nulle est visiblement solution. On se donne donc dans la suite une solution f non nulle. On observe pour commencer que f est paire. En effet, si y_0 est un réel tel que $f(y_0) \neq 0$, on a l'égalité $f(x) = \dfrac{f(\sqrt{x^2 + y_0^2})}{f(y_0)}$ pour tout x et le second terme définit une fonction paire. On s'intéresse donc à la restriction de f à $\mathbb{R}_+$.

En prenant $x = y = 0$, on obtient $f(0) = f(0)^2$ et donc $f(0) \in \{0, 1\}$. Si $f(0) = 0$, la fonction f est nulle (prendre $y = 0$). On a donc $f(0) = 1$.

On va montrer que f ne s'annule pas. Elle sera donc, par continuité, strictement positive sur $\mathbb{R}$. En prenant $y = x$ dans l'équation vérifiée par f, on obtient $f(\sqrt{2}x) = f(x)^2$ pour tout réel x. Il en résulte que si t est un zéro de f, alors $\dfrac{t}{\sqrt{2}}$ aussi et, en itérant, $\dfrac{t}{\sqrt{2}^n}$ pour tout $n \in \mathbb{N}$. Comme la suite $\dfrac{t}{\sqrt{2}^n}$ tend vers 0 lorsque n tend vers l'infini et comme f est continue, on en déduit que $f(0) = 0$. Cela n'est pas le cas, donc il n'existe pas de réel t tel que $f(t) = 0$.

Si on pose $g(x) = \ln f(\sqrt{x})$ pour $x \geqslant 0$, la fonction g est continue et vérifie l'équation fonctionnelle $g(x^2 + y^2) = g(x^2) + g(y^2)$, soit encore $g(X + Y) = g(X) + g(Y)$ pour $X, Y \geqslant 0$. Il en résulte que g est de la forme

$x \longmapsto ax$. On en déduit que $f(x) = e^{ax^2}$ pour tout $x \geqslant 0$ puis pour tout x réel par parité.

• *Synthèse.* Il est évident, réciproquement, que toutes les fonctions $x \longmapsto e^{ax^2}$ sont solutions du problème posé. ◁

L'équation fonctionnelle qui suit est nettement plus délicate à étudier.

4.16. Une équation fonctionnelle (2)

Trouver toutes les fonctions continues $g : \mathbb{R} \to \mathbb{R}$ telle que $g(g(x)) = 2g(x) - x$ pour tout x.

(École polytechnique)

▷ **Solution.**

• Soit g une fonction répondant au problème. On observe d'abord que g est injective : si $g(x) = g(y)$, on a $g(g(x)) = 2g(x) - x = g(g(y)) = 2g(y) - y$, puis $x = y$. Comme elle est continue sur $\mathbb{R}$, elle est donc strictement monotone (voir les propos précédant l'exercice 4.12). Si g est strictement décroissante, $g \circ g$ est strictement croissante, mais $x \longmapsto 2g(x) - x$ est strictement décroissante, ce qui est absurde. Donc g est strictement croissante sur $\mathbb{R}$. D'après le théorème des limites monotones, elle admet en $+\infty$ (resp. $-\infty$) une limite (finie ou infinie). Soit $l = \lim\limits_{x \to +\infty} g(x)$. Si l est fini, $g(g(x)) - 2g(x)$ tend vers $g(l) - 2l$ ce qui est absurde car cette fonction vaut $-x$. Donc $\lim\limits_{x \to +\infty} g(x) = +\infty$ et de même $\lim\limits_{x \to -\infty} g(x) = -\infty$. Ainsi, g réalise un homéomorphisme strictement croissant de $\mathbb{R}$ sur lui-même.

• Si on considère la fonction h définie par $h(x) = g(x) - x$, l'hypothèse s'écrit : pour tout x, $h(g(x)) = h(x)$. On en déduit que, pour tout $n \in \mathbb{Z}$, $h(g^n(x)) = h(x)$, soit $g^{n+1}(x) - g^n(x) = h(x)$. Ainsi, pour tout $n \in \mathbb{Z}$, $g^n(x) = x + nh(x)$ et $\dfrac{g^n(x)}{n}$ tend vers $h(x)$ lorsque n tend vers l'infini.

• On va alors montrer que la limite de $\dfrac{g^n(x)}{n}$ quand n tend vers l'infini ne dépend pas de x, ce qui prouvera que la fonction h est constante et donc que g est une translation. On suppose h non nulle et on considère $a \in \mathbb{R}$ tel que $h(a) \neq 0$. On traite le cas $h(a) > 0$, l'autre cas étant similaire. La suite $(g^n(a))_{n \in \mathbb{Z}}$ est croissante et on a $\lim\limits_{n \to +\infty} g^n(a) = +\infty$ et $\lim\limits_{n \to -\infty} g^n(a) = -\infty$. Soit alors x un réel quelconque. Il existe un unique entier $p \in \mathbb{Z}$ tel que $g^p(a) \leqslant x < g^{p+1}(a)$. Pour tout n, on a, par croissance de g,

$$\frac{g^{n+p}(a)}{n} \leqslant \frac{g^n(x)}{n} < \frac{g^{n+p+1}(a)}{n}.$$

Par encadrement, on obtient $\lim\limits_{n\to+\infty} \frac{g^n(x)}{n} = \lim\limits_{n\to+\infty} \frac{g^n(a)}{n}$. Cela vaut pour tout x, donc h est constante.

Conclusion. La réciproque étant immédiate, les translations sont les seules solutions de l'équation proposée. ◁

L'énoncé suivant est issu des Olympiades internationales de Mathématiques de 1983. Il s'agit d'une équation fonctionnelle avec très peu d'hypothèses sur la fonction.

4.17. Une équation fonctionnelle (3)

Déterminer les applications $f : \mathbb{R}_+^* \to \mathbb{R}_+^*$ qui tendent vers 0 en $+\infty$ et qui vérifient $f(xf(y)) = yf(x)$ pour tout $(x, y) \in (\mathbb{R}_+^*)^2$.

(École polytechnique)

▷ **Solution.**

• *Analyse.* Soit f une solution. En prenant $y = x$, on a pour tout $x \in \mathbb{R}_+^*$, $f(xf(x)) = xf(x)$. Autrement dit, $xf(x)$ est un point fixe de f. Examinons ce qu'on peut dire de l'ensemble des points fixes de f. Soit $y > 0$ tel que $f(y) = y$. On a alors, pour tout $x \in \mathbb{R}_+^*$, $f(xy) = yf(x)$. Supposons $y \neq 1$. Comme on a également, pour tout $x \in \mathbb{R}_+^*$, $f\left(\frac{x}{y}y\right) = yf\left(\frac{x}{y}\right)$, soit $f\left(\frac{x}{y}\right) = \frac{f(x)}{y}$, on peut supposer $y > 1$. On déduit de la relation précédente que, pour $x > 0$ et $n \in \mathbb{N}$, on a $f(xy^n) = y^n f(x)$. On fait tendre n vers $+\infty$. On obtient $\lim\limits_{n\to+\infty} y^n = +\infty$ et $\lim\limits_{n\to+\infty} f(xy^n) = 0$, par hypothèse. On en déduit que $f(x) = 0$, ce qui est contraire à l'hypothèse. On a donc nécessairement $y = 1$. Le seul point fixe de f est 1. On en déduit que, pour tout $x \in \mathbb{R}_+^*$, on a $xf(x) = 1$, soit $f(x) = \frac{1}{x}$.

• *Synthèse.* On vérifie aisément que la fonction $x \longmapsto \frac{1}{x}$ est solution du problème. ◁

Les deux exercices qui suivent concernent la recherche de racines carrées au sens de la composition.

4.18. Une équation fonctionnelle (4)

Soit a, b deux réels avec $a \neq 0$ et $a \neq 1$. Déterminer les fonctions $f : \mathbb{R} \to \mathbb{R}$ de classe C^1 telles que $f(f(x)) = ax + b$ pour tout x.

(École normale supérieure)

▷ **Solution.**

• *Analyse.* Soit f une solution de l'équation. Comme a est non nul, la fonction affine $\varphi : x \mapsto ax + b$ est bijective et il en est donc de même de f. Comme f est continue, elle est donc strictement monotone sur $\mathbb{R}$. Par conséquent $f \circ f$ est strictement croissante de sorte qu'on doit avoir $a > 0$: il n'y a pas de solution si a est négatif. Appliquons la relation en $f(x)$ pour $x \in \mathbb{R}$. Il vient $f(ax + b) = af(x) + b$ pour tout x. En dérivant cette identité on en déduit que $f'(ax + b) = f'(x)$ pour tout x. On va itérer cela. Supposons d'abord $0 < a < 1$. Soit $x \in \mathbb{R}$ fixé et (u_n) la suite arithmético-géométrique définie par $u_0 = x$ et la relation de récurrence $u_{n+1} = au_n + b$. D'après ce qui précède on a $f'(x) = f'(u_n)$ pour tout entier n. Or la suite (u_n) converge vers l'unique point fixe de $\varphi : x \mapsto ax + b$ à savoir $w = \dfrac{b}{1-a}$. Comme f' est continue on a $f'(x) = f'(w)$ et comme cela vaut pour tout x, f' est constante. Il en découle que f est affine. On a le même résultat dans le cas où $a > 1$ en remplaçant φ par φ^{-1} (comme $f' \circ \varphi = f'$ on a aussi $f' \circ \varphi^{-1} = f'$).

• *Synthèse.* Si on pose $f(x) = \alpha x + \beta$ on a $f(f(x)) = \alpha^2 x + \alpha\beta + \beta$. Pour $a > 0$ on a donc deux solutions obtenues en prenant $\alpha = \pm\sqrt{a}$ et $\beta = \dfrac{b}{\alpha + 1}$ (on a $\alpha \neq -1$ puisque $a \neq 1$). ◁

On notera que pour l'identité on aurait une infinité de solutions.

4.19. Une équation fonctionnelle (5)

Déterminer les fonctions continues $f : \mathbb{R} \mapsto \mathbb{R}$ telles que l'on ait $f(f(x)) = e^x$ pour tout x.

(École normale supérieure)

▷ **Solution.**

• *Analyse.* Soit f une solution. La fonction exp étant injective, il en est de même de f. Celle-ci étant de plus continue, elle est strictement monotone. Ainsi f réalise une bijection de $\mathbb{R}$ sur un intervalle ouvert I. On a par hypothèse $(f \circ f)(\mathbb{R}) = \mathbb{R}_+^*$ et donc $f(\mathrm{I}) = \mathbb{R}_+^*$. On en déduit que $\mathbb{R}_+^* \subset \mathrm{I}$. Ainsi l'une des bornes de I est $+\infty$ et comme $f(\mathrm{I}) \neq \mathbb{R}$ (car sinon $(f \circ f)(\mathbb{R}) = \mathbb{R}$), il existe $a \in \mathbb{R}_-^*$ tel que $\mathrm{I} =]a, +\infty[$. On

a donc $+\infty = \lim_{+\infty} f$ ou $\lim_{-\infty} f$. Comme $f(\mathrm{I}) = \mathbb{R}_+^*$, on a nécessairement $+\infty = \lim_{+\infty} f$, ce qui implique que f est strictement croissante et que $\lim_{-\infty} f = a$. On a ensuite

$$0 = \lim_{-\infty} \exp = \lim_{-\infty} f \circ f = f(a) \quad \text{et} \quad f(0) = (f \circ f)(a) = e^a.$$

On définit la suite d'intervalles (I_n) par $\mathrm{I}_0 =]-\infty, a]$ et pour tout $n \in \mathbb{N}$, $\mathrm{I}_{n+1} = f(\mathrm{I}_n)$. On a alors $\mathrm{I}_1 =]a, 0]$, $\mathrm{I}_2 =]0, e^a]$ et, pour tout $n \in \mathbb{N}$, $\mathrm{I}_{n+2} = \exp(\mathrm{I}_n)$. Soit (a_n) et (b_n) les suites définies par $a_0 = a$, $b_0 = 0$ et pour tout $n \in \mathbb{N}$, $a_{n+1} = e^{a_n}$ et $b_{n+1} = e^{b_n}$. Elles sont strictement croissantes et vérifient $a_n < b_n < a_{n+1}$ pour tout $n \in \mathbb{N}$ (cela résulte de $a_0 < b_0 < a_1$). Pour tout $n \in \mathbb{N}$, on a

$$\mathrm{I}_{2n+1} = \exp^n(]a, 0]) =]a_n, b_n]$$
$$\mathrm{I}_{2n+2} = \exp^n(\mathrm{I}_2) = \exp^n(]0, e^a]) =]b_n, a_{n+1}].$$

On constate que $\mathbb{R}$ est la réunion disjointe des intervalles I_n. Montrons que f est déterminée par sa restriction à I_0. Notons f_n la restriction de la fonction f à I_n. Elle réalise une bijection de I_n sur I_{n+1}. Soit $x \in \mathrm{I}_{n+1}$ et $y = f_n^{-1}(x) \in \mathrm{I}_n$. On a

$$f_{n+1}(x) = (f \circ f)(y) = e^y = e^{f_n^{-1}(x)}.$$

Ainsi f_{n+1} s'exprime en fonction de f_n et donc f est déterminée par la donnée de f_0.

• *Synthèse.* Montrons que toutes les fonctions de cette forme conviennent. Soit a un réel strictement négatif, (a_n) et (b_n) les suites définies comme précédemment. On pose $\mathrm{I}_0 =]-\infty, a]$, $\mathrm{I}_1 =]a, 0]$ et, pour tout $n \in \mathbb{N}$, $\mathrm{I}_{n+2} = \exp(\mathrm{I}_n)$. On alors, pour tout $n \in \mathbb{N}$, $\mathrm{I}_{2n+1} =]a_n, b_n]$ et $\mathrm{I}_{2n+2} =]b_n, a_{n+1}]$ et $\mathbb{R}$ est la réunion disjointe des intervalles I_n. Soit f_0 une bijection continue croissante quelconque de I_0 sur I_1. On définit par récurrence sur n une bijection croissante f_n de I_n sur I_{n+1} en posant $f_{n+1} = \exp \circ f_n^{-1}$. Si f_n est une bijection de I_n sur I_{n+1}, f_n^{-1} est une bijection de I_{n+1} sur I_n et exp une bijection de I_n sur I_{n+2}, donc f_{n+1} est une bijection de I_{n+1} sur I_{n+2}. On appelle f la fonction de $\mathbb{R}$ dans $\mathbb{R}$ dont la restriction à chaque intervalle I_n est f_n. La fonction f est une bijection croissante de $\mathbb{R}$ sur $\bigcup_{n \in \mathbb{N}^*} \mathrm{I}_n =]a, +\infty[$. Par construction, on a pour tout $x \in \mathrm{I}_n$,

$$(f \circ f)(x) = f_{n+1}(f_n(x)) = e^x.$$

Reste à vérifier la continuité de f. Par construction elle est continue sur chaque intervalle I_n. Il reste à vérifier pour tout $n \in \mathbb{N}^*$, la continuité à

droite en la borne inférieure de l'intervalle I_n. Comme pour tout $n \in \mathbb{N}$, $f_{n+1} = \exp \circ f_n^{-1}$, il suffit de le vérifier pour I_1, c'est-dire de démontrer la continuité de f en a à droite. Pour $x \in I_1$, on a $f(x) = \exp(f_0^{-1}(x))$. Comme f_0 est une bijection de $]-\infty, a]$ sur $]a, 0]$, on a $\lim\limits_{a} f_0 = -\infty$ et donc $\lim\limits_{a+} f = \lim\limits_{-\infty} \exp = 0 = f(a)$.

Les solutions du problème sont donc toutes les fonctions ainsi construites, $a < 0$ et f_0 étant quelconques. ◁

4.20. Une équation fonctionnelle (6)

Soient a, b, c des réels strictement positifs deux à deux distincts. Trouver les fonctions $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^\infty$ telles que pour tout x,

$$f(ax) + f(bx) + f(cx) = 0.$$

(École normale supérieure)

▷ **Solution.**

On va montrer qu'il n'y a que la fonction nulle. On peut supposer sans perte de généralité que $0 < a < b < c$. Soit f une solution et $n \in \mathbb{N}$. En dérivant n fois l'égalité, on obtient, pour tout réel x,

$$a^n f^{(n)}(ax) + b^n f^{(n)}(bx) + c^n f^{(n)}(cx) = 0.$$

Cela peut s'écrire, en remplaçant x par $\dfrac{x}{c}$,

$$\forall x \in \mathbb{R}, \quad f^{(n)}(x) = -\frac{a^n}{c^n} f^{(n)}\left(\frac{ax}{c}\right) - \frac{b^n}{c^n} f^{(n)}\left(\frac{bx}{c}\right).$$

Choisissons n tel que $\dfrac{a^n}{c^n} + \dfrac{b^n}{c^n} < 1$. C'est possible car $\dfrac{a^n}{c^n} + \dfrac{b^n}{c^n}$ tend vers 0. Soit $\alpha > 0$ et M_α le maximum de $f^{(n)}$ sur l'intervalle $[-\alpha, \alpha]$. On a, pour tout $x \in [-\alpha, \alpha]$,

$$|f^{(n)}(x)| \leqslant \frac{a^n}{c^n}\left|f^{(n)}\left(\frac{ax}{c}\right)\right| + \frac{b^n}{c^n}\left|f^{(n)}\left(\frac{bx}{c}\right)\right| \leqslant \left(\frac{a^n}{c^n} + \frac{b^n}{c^n}\right) M_\alpha,$$

car $\dfrac{ax}{c}$ et $\dfrac{bx}{c}$ sont dans $[-\alpha, \alpha]$. On en déduit que $M_\alpha \leqslant \left(\dfrac{a^n}{c^n} + \dfrac{b^n}{c^n}\right) M_\alpha$ et donc $M_\alpha = 0$, puisque $\dfrac{a^n}{c^n} + \dfrac{b^n}{c^n} < 1$. Ainsi la fonction $f^{(n)}$ est nulle sur $[-\alpha, \alpha]$ et comme ceci est vrai pour tout $\alpha > 0$, c'est la fonction nulle. On en déduit que f est une fonction polynomiale de degré inférieur ou égale à $n-1$. Si $f = \sum\limits_{k=0}^{n-1} a_k x^k$, on obtient, pour tout réel x,

$$f(ax) + f(bx) + f(cx) = \sum_{k=0}^{n-1} a_k(a^k + b^k + c^k)x^k = 0.$$

Comme $a^k + b^k + c^k > 0$ pour tout k, on a $a_k = 0$ pour tout k et donc $f = 0$. $\lhd$

4.21. Valeurs propres d'un opérateur

1. Trouver toutes les fonctions continues $f : [0,1] \to \mathbb{R}$ telles que $f(x) = \sum_{n=1}^{+\infty} \frac{f(x^n)}{3^n}$ pour tout $x \in [0,1]$.

2. Même question avec la condition $f(x) = \sum_{n=1}^{+\infty} \frac{f(x^n)}{2^n}$ pour tout $x \in [0,1]$.

3. Pour $0 < c < 1$, déterminer toutes les fonctions continues $f : [0,1] \to \mathbb{R}$ telles que, pour tout $x \in [0,1]$, $f(x) = c \sum_{n=1}^{+\infty} \frac{f(x^n)}{2^n}$. Traiter également le cas $c = 2$.

(**École normale supérieure, école polytechnique**)

$\rhd$ **Solution.**

1. Soit f continue sur $[0,1]$ vérifiant $f(x) = \sum_{n=1}^{+\infty} \frac{f(x^n)}{3^n}$ pour tout $x \in [0,1]$. Comme $[0,1]$ est compact, f est bornée et on a, pour $0 \leqslant x \leqslant 1$,

$$|f(x)| \leqslant \sum_{n=1}^{+\infty} \frac{\|f\|_\infty}{3^n} = \frac{1}{3}\frac{1}{1-1/3}\|f\|_\infty = \frac{1}{2}\|f\|_\infty.$$

Il en résulte que $\|f\|_\infty \leqslant \frac{1}{2}\|f\|_\infty$ et donc $\|f\|_\infty = 0$. La fonction f est nulle.

2. La démarche précédente ne donne rien ici car la somme de la série géométrique $\sum_{n=1}^{+\infty} \frac{1}{2^n}$ est égale à 1. Il en résulte que toutes les fonctions constantes conviennent. On peut aussi noter que l'ensemble des solutions est un espace vectoriel. Soit f une solution. Quitte à lui retrancher la constante $f(0)$, on peut supposer que $f(0) = 0$.

Soit $a \in]0,1[$. Comme f est continue sur le compact $[0,a]$, il existe $b \in [0,a]$ tel que $f(b) = \max_{[0,a]} f$. Par hypothèse $f(b) = \sum_{n=1}^{+\infty} \frac{f(b^n)}{2^n}$. Comme pour tout $n \geqslant 1$, $0 \leqslant b^n \leqslant b \leqslant a$, on a $f(b^n) \leqslant f(b)$ par définition de b. On a donc

$$f(b) \leqslant \sum_{n=1}^{+\infty} \frac{f(b)}{2^n} = f(b)$$

ct l'inégalité serait stricte s'il existait un entier n tel que $f(b^n) < f(b)$. On a donc pour tout $n \in \mathbb{N}$, $f(b^n) = f(b)$. Comme $0 \leqslant b < 1$, la suite (b^n) converge vers 0 et $\lim_{n\to+\infty} f(b^n) = f(0) = 0$, puisque f est continue. Ainsi pour tout $x \in [0, a]$, $f(x) \leqslant 0$. Comme a a été choisi arbitrairement dans $[0, 1[$, $f(x) \leqslant 0$ pour $0 \leqslant x < 1$ et par continuité, $f \leqslant 0$ sur $[0, 1]$.

Si f est solution, $-f$ est aussi solution, donc on a aussi $-f \leqslant 0$ et finalement f est nulle.

Conclusion. Les solutions du problème sont les fonctions constantes.

3. En majorant comme dans la question 1, on a $\|f\|_\infty \leqslant c\|f\|_\infty$, et si $0 < c < 1$, cela impose $\|f\|_\infty = 0$ et f est nulle. Supposons maintenant $c = 2$. On a, après simplification, pour tout $x \in [0, 1]$

$$0 = \sum_{n=2}^{+\infty} \frac{f(x^n)}{2^{n-1}}, \text{ c'est-à-dire } f(x^2) = -\sum_{n=3}^{+\infty} \frac{f(x^n)}{2^{n-2}}.$$

On obtient en particulier

$$f(0) = -\sum_{n=3}^{+\infty} \frac{f(0)}{2^{n-2}} = -\frac{f(0)}{2} \quad \text{et} \quad f(1) = -\sum_{n=3}^{+\infty} \frac{f(1)}{2^{n-2}} = -\frac{f(1)}{2},$$

et donc $f(0) = f(1) = 0$. Notons pour $0 \leqslant x \leqslant 1$, $\mathrm{M}(x) = \max_{0\leqslant t\leqslant x} |f(t)|$. On a pour $0 \leqslant t \leqslant x$,

$$|f(t^2)| \leqslant \sum_{n=3}^{+\infty} \frac{\mathrm{M}(x^3)}{2^{n-2}} = \mathrm{M}(x^3)$$

car si $n \geqslant 3$, $t^n \leqslant x^n \leqslant x^3$. Il s'ensuit que $\mathrm{M}(x^2) \leqslant \mathrm{M}(x^3)$. Mais $\mathrm{M}(x^3)$ est inférieur à $\mathrm{M}(x^2)$, puisque $x^3 \leqslant x^2$. On en déduit que $\mathrm{M}(x^2) = \mathrm{M}(x^3)$. Il en résulte que pour $0 \leqslant x \leqslant 1$ et $n \geqslant 1$,

$$\mathrm{M}(x) = \mathrm{M}(x^{3/2}) = \mathrm{M}(x^{(3/2)^2}) = \cdots = \mathrm{M}(x^{(3/2)^n}).$$

Montrons que M est continue en 0. On a $\mathrm{M}(0) = 0 = f(0)$. Soit $\varepsilon > 0$. Il existe $\eta > 0$ tel que pour $0 \leqslant x \leqslant \eta$, $|f(x)| \leqslant \varepsilon$ puisque f est continue en 0. Il s'ensuit que pour $0 \leqslant x \leqslant \eta$, $|\mathrm{M}(x)| \leqslant \varepsilon$. D'où la continuité de M en 0.

Comme pour $0 \leqslant x < 1$, la suite $(x^{(3/2)^n})$ converge vers 0, on a $\mathrm{M}(x) = \mathrm{M}(x^{(3/2)^n}) \xrightarrow[n\to+\infty]{} \mathrm{M}(0) = 0$. Ainsi M est nulle sur $[0, 1[$, et f aussi.

Conclusion. La seule solution est la fonction nulle. ◁

Les deux énoncés qui suivent concernent la continuité uniforme, notion qui émerge dans les travaux de Dirichlet (1854) et Weierstrass (1861). C'est Heine qui, en 1870, en donne la définition que l'on connaît. Le célèbre théorème qui porte son nom affirme qu'une fonction continue sur un compact est automatiquement uniformément continue. Le premier exercice, très classique, montre qu'une application uniformément continue sur $\mathbb{R}$ est un $O(x)$ en $\pm\infty$.

4.22. Domination d'une fonction uniformément continue

Soit $f : \mathbb{R} \to \mathbb{R}$ uniformément continue sur $\mathbb{R}$. Montrer qu'il existe a et $b > 0$ tels que pour tout $x \in \mathbb{R}$, $|f(x)| \leqslant a|x| + b$. La réciproque est-elle vraie ?

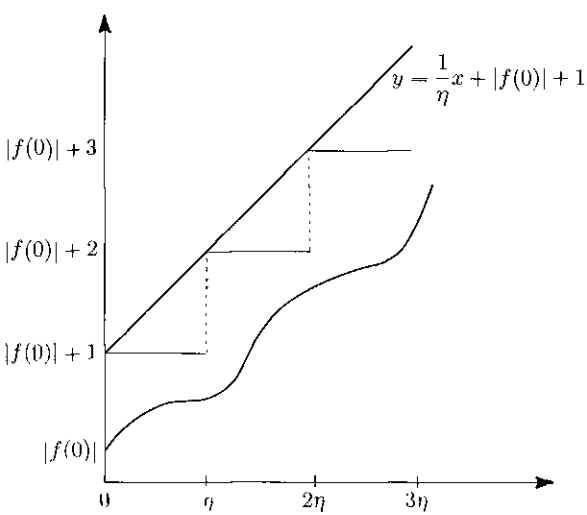
(École polytechnique)

▷ **Solution.**

Puisque f est uniformément continue sur $\mathbb{R}$, il existe $\eta > 0$ tel que si $|x - y| \leqslant \eta$ alors $|f(x) - f(y)| \leqslant 1$. On en déduit que pour $x \in [0, \eta]$ on a $|f(x)| \leqslant |f(0)| + 1$, puis que pour $x \in [\eta, 2\eta]$, $|f(x)| \leqslant |f(\eta)| + 1 \leqslant |f(0)| + 2$. Une récurrence immédiate montre que pour tout $n \in \mathbb{N}$ on a $|f(x)| \leqslant |f(0)| + n + 1$ sur l'intervalle $[n\eta, (n+1)\eta]$. Sur cet intervalle on peut majorer n par $\dfrac{x}{\eta}$. On a donc prouvé que pour tout $x \geqslant 0$,

$$|f(x)| \leqslant |f(0)| + 1 + \frac{x}{\eta}.$$

La figure suivante représente le graphe de $|f|$ et les majorations effectuées.



On peut faire un travail analogue sur $\mathbb{R}^-$. Sur l'intervalle $[-\eta, 0]$ on a aussi $|f(x)| \leqslant |f(0)| + 1$, et par récurrence on montre que pour tout $n \geqslant 1$, on a $|f(x)| \leqslant |f(0)| + n$ sur l'intervalle $[-n\eta, (-n+1)\eta]$. Comme pour $x \in [-n\eta, (-n+1)\eta]$, on a $n \leqslant \dfrac{|x|}{\eta} + 1$, il en résulte que pour tout $x \leqslant 0$,

$$|f(x)| \leqslant |f(0)| + 1 + \frac{|x|}{\eta}.$$

On a donc obtenu une majoration de la forme souhaitée.

Bien entendu, la réciproque est fausse. Par exemple la fonction bornée $x \longmapsto \sin(x^2)$ n'est pas uniformément continue : en effet, lorsqu'une fonction f est uniformément continue sur un intervalle I, pour toutes suites $(x_n)_{n\in\mathbb{N}}$ et $(y_n)_{n\in\mathbb{N}}$ de I telle que $\lim\limits_{n\to\infty} x_n - y_n = 0$, on a $\lim\limits_{n\to\infty} f(x_n) - f(y_n) = 0$. En prenant $x_n = \sqrt{2\pi n}$ et $y_n = \sqrt{2\pi n + \dfrac{\pi}{6}}$, on obtient que $x \longmapsto \sin(x^2)$ ne peut être uniformément continue. ◁

4.23. Lemme de Croft

Soit $f : \mathbb{R}_+ \to \mathbb{R}$ telle que pour tout $x > 0$, la suite $f(nx)$ tend vers 0.

1. Montrer que si f est uniformément continue, on a $\lim\limits_{x\to+\infty} f(x) = 0$.

2. Donner un exemple où f ne tend pas vers 0 en $+\infty$.

(**École polytechnique**)

▷ **Solution.**

1. L'hypothèse signifie que f tend vers 0 selon toute suite arithmétique de la forme $(nx)_{n\geqslant 0}$. Lorsque la fonction est uniformément continue, on peut facilement contrôler son comportement entre deux termes consécutifs de la suite. Plus précisément, soit $\varepsilon > 0$. La continuité uniforme de f permet de choisir $\eta > 0$ tel que $|f(x) - f(y)| \leqslant \varepsilon$ lorsque $|x - y| \leqslant \eta$. La suite $(f(n\eta))_{n\geqslant 0}$ tend vers 0 par hypothèse. Fixons N tel que $|f(n\eta)| \leqslant \varepsilon$ pour $n \geqslant \mathrm{N}$. Si x est un réel supérieur à $\mathrm{N}\eta$, il existe un entier $n \geqslant \mathrm{N}$ tel que $|x - n\eta| \leqslant \eta$. On a alors $|f(x) - f(n\eta)| \leqslant \varepsilon$ de sorte que $|f(x)| \leqslant 2\varepsilon$. Cela prouve que f tend vers 0 en $+\infty$.

2. L'exemple qui suit nous a été proposé par Daniel Kitachewsky, alors élève en classe de Mathématiques Supérieures. On pose $f(\pi^n) = 1$ pour tout $n \in \mathbb{N}$ et $f(x) = 0$ sinon. La fonction f ne tend pas vers 0 en $+\infty$, mais pour tout $x > 0$ la suite $f(nx)$ tend vers 0 car elle prend au plus une valeur non nulle : en effet, s'il y en avait deux distinctes, il existerait $0 < k < l$ et $n < m$ des entiers tels que $\pi^m = kx$ et $\pi^n = lx$.

Dans ces conditions, le rationnel $\frac{l}{k}$ serait égal à π^{n-m}, ce qui est exclu puisque π est transcendant. ◁

Le fait que π est transcendant constitue le théorème de Lindemann établi en 1882 : le nombre π n'est pas racine d'un polynôme non nul à coefficients entiers.

Le résultat de la première question reste vrai si la fonction f est seulement supposée continue. La démonstration utilise le théorème de Baire et est un peu plus délicate : on la trouvera dans le chapitre topologie du troisième tome d'analyse.

Les exercices suivants sont consacrés à la dérivabilité. Le développement des méthodes infinitésimales s'accéléra au XVII^{e} *siècle avec les problèmes de calcul de vitesse instantanée, de direction de tangentes, de minima et maxima, de longueur de trajectoire... Dans ce domaine, nous devons beaucoup à Pascal, à Fermat qui utilisait la notion d'infiniment petit, à Descartes qui critiquait déjà ce langage (la polémique ne s'éteindra vraiment qu'au début du* XIX^{e} *siècle). Mais c'est Newton (1642-1727) et Leibniz (1646-1716) qui, par leur travaux, contribuèrent à donner de solides fondements à cette nouvelle branche des mathématiques. Reprenant au début les infiniment petits, Newton s'en éloigna dans son exposé de la méthode dites des «fluxions» et enfin, dans* Quadratura curvarum (1704), *il donna une version très proche de l'actuelle définition de la dérivée, comme limite du rapport* $\frac{f(x+h)-f(x)}{h}$. *De manière indépendante, Leibniz créa une partie du formalisme moderne (avec les notations dx, $\frac{dx}{dt}$, $\int$), introduisit les dérivées d'ordre supérieur et donna bon nombre de règles de dérivation des quantités composées par sommes, produits...*

4.24. Étude d'un taux d'accroissement

Soit $f : \mathbb{R} \to \mathbb{R}$ continue en 0. Déterminer une condition nécessaire et suffisante pour que $\frac{f(2x)-f(x)}{x}$ admette une limite en 0.

(École polytechnique)

▷ **Solution.**

• La quantité $\frac{f(2x)-f(x)}{x}$ est le taux d'accroissement de f entre x et $2x$. Il est facile de l'estimer lorsque la fonction f est dérivable en 0. En effet, on peut alors écrire $f(x) = f(0) + f'(0)x + o(x)$. Il en résulte que

$f(2x) = f(0) + 2f'(0)x + o(x)$ et le rapport $\dfrac{f(2x) - f(x)}{x} = f'(0) + o(1)$ tend donc vers $f'(0)$.

• On va montrer réciproquement que, si le rapport $\dfrac{f(2x) - f(x)}{x}$ admet une limite a en 0, alors f est dérivable en 0 avec $f'(0) = a$. Quitte à considérer la fonction $g(x) = f(x) - ax$, on peut supposer que $a = 0$. Posons $\varepsilon(x) = \dfrac{f(2x) - f(x)}{x}$ prolongée par continuité en 0 par $\varepsilon(0) = 0$. On a pour tout entier $n \geqslant 1$,

$$f(x) - f\left(\frac{x}{2^n}\right) = \sum_{k=1}^{n} \left(f\left(\frac{x}{2^{k-1}}\right) - f\left(\frac{x}{2^k}\right)\right) = \sum_{k=1}^{n} \frac{x}{2^k} \varepsilon\left(\frac{x}{2^k}\right).$$

Comme $f\left(\frac{x}{2^n}\right)$ tend vers $f(0)$ par continuité de f en 0, on a en faisant tendre n vers l'infini

$$f(x) - f(0) = x \underbrace{\sum_{k=1}^{+\infty} \frac{\varepsilon(x/2^k)}{2^k}}_{\tau(x)}.$$

Prenons $\alpha > 0$ tel que ε soit bornée sur $[-\alpha, \alpha]$ (un tel intervalle existe puisque ε tend vers 0 en 0). La série définissant τ converge normalement sur $[\alpha, -\alpha]$, de sorte que τ est continue en 0. Ainsi, $\dfrac{f(x) - f(0)}{x}$ tend vers $\tau(0) = 0$ quand x tend vers 0, *i.e.* f est dérivable en 0 avec $f'(0) = 0$. ◁

La notion de dérivabilité prête à plusieurs généralisations. Pour définir la pseudo-dérivée dans l'exercice suivant, on utilise une expression symétrique par rapport au point considéré.

4.25. Pseudo-dérivée

Soit I un intervalle ouvert non vide de $\mathbb{R}$ et $f : \mathrm{I} \to \mathbb{R}$ continue sur I, et telle que

$$\tilde{f}(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h}$$

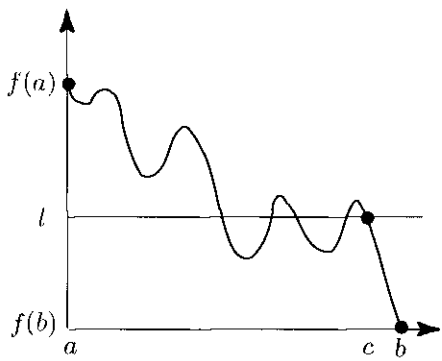
existe pour tout $x \in \mathrm{I}$ (nous dirons que f est *pseudo-dérivable* sur I). Montrer que si $\tilde{f} \geqslant 0$, alors f est croissante (on pourra d'abord supposer $\tilde{f} \geqslant \alpha > 0$).

(École polytechnique)

▷ **Solution.**

• Supposons $\tilde{f} \geqslant \alpha$ avec $\alpha > 0$. Raisonnons par l'absurde et supposons que f ne soit pas croissante. Dans ces conditions, il existe $a < b$ tel que $f(a) > f(b)$. On va essayer de trouver un point c dans l'intervalle $]a, b[$ pour lequel il existe des valeurs de $h > 0$ aussi petites que l'on souhaite telles que $f(c-h) \geqslant f(c+h)$. Pour une telle valeur h le taux d'accroissements $\dfrac{f(c+h)-f(x-h)}{2h}$ est négatif ou nul ce qui en passant à la limite fournira une contradiction avec le fait que $\tilde{f}(c) > 0$.

Choisissons $l \in]f(b), f(a)[$ et considérons $\mathrm{E} = \{x \in [a, b], f(x) > l\}$. Comme f est continue, E est un ouvert (relatif) de $[a, b]$ qui contient a. Il existe donc $\eta > 0$ tel que $[a, a+\eta] \subset \mathrm{E}$.



Posons $c = \sup \mathrm{E}$. On a $a < c < b$, la deuxième inégalité stricte étant justifiée par le fait que si c était égal à b, $f(b)$ serait par continuité supérieur ou égal à l. Par construction, on a $f(x) \leqslant l$ pour tout $x \in]c, b]$. On va maintenant montrer qu'à gauche de c il y a des points x aussi proches de c que l'on veut tels que $f(x) > l$ (autrement dit des points de E). Comme E est ouvert, $c \notin \mathrm{E}$ et comme $c = \sup \mathrm{E}$, c est un point d'accumulation de E : il existe donc une suite de réels strictement positifs $(h_n)_{n\in\mathbb{N}}$, tendant vers 0, telle que $c - h_n \in \mathrm{E}$ pour tout $n \in \mathbb{N}$. On obtient alors la contradiction annoncée plus haut. Pour tout n, on a

$$\frac{f(c+h_n)-f(c-h_n)}{2h_n} \leqslant \frac{l-l}{2h_n} = 0,$$

et donc $\tilde{f}(c) = \lim\limits_{n\to\infty} \dfrac{f(c+h_n)-f(c-h_n)}{2h_n} \leqslant 0$, ce qui contredit l'hypothèse. On conclut donc que si $\tilde{f} \geqslant \alpha > 0$, f est croissante.

• Supposons maintenant $\tilde{f} \geqslant 0$. Considérons pour tout $\alpha > 0$, la fonction $g_\alpha : x \in \mathrm{I} \longmapsto f(x) + \alpha x$. Il est clair que g_α est pseudo-dérivable et que pour tout $x \in \mathrm{I}$, $\tilde{g_\alpha}(x) = \tilde{f}(x) + \alpha \geqslant \alpha > 0$. D'après ce qui précède, g_α est croissante et donc pour $x \leqslant y$ dans I, on a $g_\alpha(x) \leqslant g_\alpha(y)$, ce qui

s'écrit $f(x)+\alpha x \leqslant f(x)+\alpha y$. Si x, y sont fixés, cette inégalité est vraie pour tout $\alpha > 0$. En faisant tendre α vers 0, on obtient $f(x) \leqslant f(y)$. Donc f est bien croissante. $\lhd$

Il est clair qu'une fonction dérivable en un point est aussi pseudo-dérivable en ce point. La réciproque est fausse, la fonction n'étant même pas nécessairement continue. Par exemple, la fonction f qui est nulle partout, sauf en 0 où elle vaut 1, n'est pas continue en 0, mais elle est pseudo-dérivable avec $\tilde{f}(0) = 0$. Même si on rajoute une hypothèse de continuité, une fonction peut être pseudo-dérivable sans être dérivable : c'est par exemple le cas de $x \longmapsto |x|$ en 0.

Dire que f est une fonction dérivable en un point x_0 signifie que $f(x) = f(x_0) + f'(x_0)(x - x_0) + o(x - x_0)$, c'est-à-dire qu'en remplaçant f par l'équation de sa tangente en x_0, on commet une erreur qui est un $o(x - x_0)$: c'est d'ailleurs la formulation de la dérivabilité que donne Weierstrass. Cette simple approximation est la clé de bien des exercices dont le suivant.

4.26. Limite d'une somme

Soit $f : \mathbb{R} \to \mathbb{R}$ dérivable en 0 et vérifiant $f(0) = 0$. Déterminer la limite de la suite de terme général

$$s_n = \sum_{k=0}^{n} f\left(\frac{k}{n^2}\right).$$

(École normale supérieure, École polytechnique)

$\rhd$ **Solution.**

L'idée est de remplacer f par son approximation affine en 0, qui conduit à une somme qui se calcule. On a $f(x) = f'(0)x + x\varepsilon(x)$ où ε est une fonction de limite nulle en 0. On a alors

$$s_n = f'(0) \sum_{k=0}^{n} \frac{k}{n^2} + \underbrace{\sum_{k=0}^{n} \frac{k}{n^2} \varepsilon\left(\frac{k}{n^2}\right)}_{= r_n}.$$

Comme $\sum_{k=0}^{n} k = \dfrac{n(n+1)}{2}$, le premier terme ci-dessus tend vers $\dfrac{f'(0)}{2}$. Montrons que r_n tend vers 0. Soit $\alpha > 0$. Il existe $\eta > 0$ tel que $|\varepsilon(x)| \leqslant \alpha$ sur $[0, \eta]$. Soit N tel que $\dfrac{1}{\mathrm{N}} \leqslant \eta$. Alors, pour $n \geqslant \mathrm{N}$, toutes les fractions

$\frac{k}{n^2}$, pour $k \in [\![0.n]\!]$ sont dans l'intervalle $[0,\eta]$. Ainsi, pour tout $n \geqslant N$, $|r_n| \leqslant \sum_{k=0}^{n} \frac{k}{n^2}\alpha = \frac{n+1}{2n}\alpha \leqslant \alpha$. D'où le résultat : $\lim s_n = \frac{f'(0)}{2}$. ◁

L'exercice qui suit possède des solutions très variées.

4.27. La fonction sinus n'est pas rationnelle

Montrer que la fonction sinus n'est pas la restriction à $]a,b[$ d'une fraction rationnelle.

(École normale supérieure)

▷ **Solution.**

Cet exercice est facile si on pense à utiliser la notion de degré d'une fraction rationnelle : le degré de $F = \frac{P}{Q}$ est par définition l'entier relatif $\deg P - \deg Q$ si $P \neq 0$ et $-\infty$ si $P = 0$ (il est immédiat que cette définition ne dépend pas de la représentation choisie pour F).

La fonction sinus est égale à sa dérivée quatrième. Raisonnons par l'absurde en supposant que la fraction rationnelle F coïncide avec la fonction sinus sur l'intervalle $]a,b[$. On a alors $F^{(4)}(x) = F(x)$ pour tout $x \in]a,b[$, et F est non nulle, puisque la fonction sinus ne saurait être nulle sur un intervalle ouvert non vide. La fraction rationnelle $F^{(4)} - F$ possède une infinité de zéros (puisque $]a,b[$ est infini). Or une fraction rationnelle non nulle n'a qu'un nombre fini de zéros. Donc $F^{(4)} - F$ est la fraction rationnelle nulle, autrement dit $F^{(4)} = F$.

Or si F est de degré n, F' est de degré $\leqslant n-1$ comme on le vérifie sans peine, et $F^{(4)}$ est de degré $\leqslant n-4$. L'égalité $F^{(4)} = F$ est donc impossible, ce qui fournit la contradiction attendue. ◁

Leibniz écrivait déjà à la fin du XVII^e siècle que le sinus n'était pas fonction rationnelle de l'angle et donc que la quadrature algébrique du cercle est sans espoir.

La formule des accroissements finis et l'inégalité qui en découle sont des outils puissants de l'Analyse. C'est Lagrange qui en 1797 énonce ce résultat fondamental : si f est une fonction à valeurs réelles continue sur $[a,b]$, dérivable sur $]a,b[$, il existe c dans $]a,b[$ tel que

$$f(b) - f(a) = f'(c)(b-a).$$

La démonstration utilise le théorème de Rolle établi en 1690 (il s'agissait de prouver qu'entre deux racines d'un polynôme P, on peut en trouver une de P'). La formule des accroissements finis permet de caractériser

la monotonie d'une fonction, sur tout intervalle où elle est dérivable, à l'aide du signe de la dérivée.

Dans l'exercice suivant on utilise la structure des ouverts de $\mathbb{R}$: un ouvert non vide U de $\mathbb{R}$ s'écrit de manière unique comme réunion d'une famille d'intervalles ouverts non vides deux à deux disjoints (qui sont les composantes connexes de l'ouvert) ; cette famille est au plus dénombrable (puisque dans chaque intervalle ouvert on peut trouver un rationnel). Le lecteur trouvera tous ces résultats dans l'exercice 1.27.

4.28. Un cas simple du lemme de Sard

Une partie A de $\mathbb{R}$ est dite négligeable (ou de mesure nulle) si pour tout $\varepsilon > 0$, il existe une suite (I_n) d'intervalles ouverts telle que $\mathrm{A} \subset \bigcup_{n\in\mathbb{N}} \mathrm{I}_n$ et $\sum_{n\in\mathbb{N}} \mu(\mathrm{I}_n) < \varepsilon$ ($\mu(\mathrm{I})$ désigne la longueur d'un intervalle I).

1. Montrer qu'une réunion dénombrable de parties négligeables est négligeable.

2. Soit $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^1$. On note C l'ensemble des zéros de f'. Montrer que $f(\mathrm{C})$ est négligeable.

(École polytechnique)

▷ **Solution.**

1. Soit $(\mathrm{A}_p)_{p\in\mathbb{N}}$ une famille dénombrable de parties négligeables, $\mathrm{A} = \bigcup_{p\in\mathbb{N}} \mathrm{A}_p$ et $\varepsilon > 0$. Pour tout p, on peut trouver une famille dénombrable $(\mathrm{I}_{n,p})_{n\in\mathbb{N}}$ d'intervalles ouverts, telle que $\sum_{n\in\mathbb{N}} \mu(\mathrm{I}_{n,p}) \leqslant \dfrac{\varepsilon}{2^p}$. La famille $(\mathrm{I}_{n,p})_{(n,p)\in\mathbb{N}^2}$ est alors une famille dénombrable d'intervalles ouverts. Le théorème d'associativité pour les familles sommables de réels positifs assure la sommabilité de la famille $(\mu(\mathrm{I}_{n,p}))_{(n,p)\in\mathbb{N}^2}$ et on a

$$\sum_{(n,p)\in\mathbb{N}^2} \mu(\mathrm{I}_{n,p}) = \sum_{p=0}^{\infty}\sum_{n=0}^{\infty} \mu(\mathrm{I}_{n,p}) \leqslant \sum_{p=0}^{\infty} \frac{\varepsilon}{2^p} = 2\varepsilon.$$

Comme A est incluse dans la réunion des $\mathrm{I}_{n,p}$, le résultat est démontré.

Le résultat de cette question figure aussi dans l'exercice 1.23.

2. D'après la question précédente, il suffit de prouver que l'ensemble $f(\mathrm{C} \cap]-\mathrm{N}, \mathrm{N}[)$ est négligeable pour tout $\mathrm{N} \in \mathbb{N}^*$. Soit $\varepsilon > 0$. Posons $\mathrm{A} = \{x \in]-\mathrm{N}, \mathrm{N}[,\ |f'(x)| < \varepsilon\}$. Par continuité de f', A est un ouvert. Sur toute composante connexe $]a, b[$ de cet ouvert, f est ε-lipschitzienne par la formule des accroissements finis. Donc $f(]a, b[)$ est un intervalle (par le théorème des valeurs intermédiaires) dont la

longueur est inférieure à $(b-a)\varepsilon$. L'ensemble des composantes connexes de A est au plus dénombrable puisque chacune d'entre elle contient un rationnel. Si on note $(\mathrm{U}_i)_{i\in\mathrm{I}}$ cette famille de composantes connexes, $f(\mathrm{C}\cap\,]-\mathrm{N},\mathrm{N}[)\subset f(\mathrm{A})\subset\bigcup_{i\in\mathrm{I}} f(\mathrm{U}_i)$ et

$$\sum_{i\in\mathrm{I}}\mu(f(\mathrm{U}_i))\leqslant\varepsilon\sum_{i\in\mathrm{I}}\mu(\mathrm{U}_i)\leqslant 2\mathrm{N}\varepsilon.$$

Justifions la dernière inégalité : les U_i étant deux à deux disjoints, pour toute partie $\mathrm{J}\subset\mathrm{I}$, on a $\sum_{i\in\mathrm{J}}\mu(\mathrm{U}_j)\leqslant\mu(]-\mathrm{N},\mathrm{N}[)=2\mathrm{N}$ car $\mathrm{A}\subset\,]-\mathrm{N},\mathrm{N}[$.
On en déduit que $f(\mathrm{C}\cap\,]-\mathrm{N},\mathrm{N}[)$ est bien négligeable. ◁

Le résultat reste vrai si la fonction f est seulement supposée dérivable, mais la preuve est un peu plus compliquée. Il se généralise[1] *également aux applications de classe $\mathcal{C}^1$ d'un ouvert de $\mathbb{R}^n$ dans $\mathbb{R}^n$.*

Les fonctions dérivées, qui ne sont en général pas continues, ont tout de même des propriétés remarquables. En particulier, elles vérifient le théorème des valeurs intermédiaires. C'est ce qu'affirme le théorème de Darboux qui fait l'objet de l'exercice suivant. On en déduit donc qu'une fonction qui vérifie le théorème des valeurs intermédiaires n'est pas nécessairement continue (il suffit de prendre la dérivée d'une fonction dérivable que n'est pas de classe $\mathcal{C}^1$, comme par exemple $x\mapsto x^2\sin 1/x$ prolongée par continuité en 0).

4.29. Théorème de Darboux

Soit I un intervalle et $f:\mathrm{I}\to\mathbb{R}$ une fonction dérivable. Montrer que $f'(\mathrm{I})$ est un intervalle.

(École polytechnique)

▷ **Solution.**
L'ensemble $f'(\mathrm{I})$ est un intervalle s'il contient $[\alpha,\beta]$ dès que α et β sont dans $f'(\mathrm{I})$. Soit $a<b$ deux points de I et λ un réel compris entre $f'(a)$ et $f'(b)$. On va montrer qu'il existe $x\in[a,b]$ tel que $f'(x)=\lambda$. En vertu de la formule des accroissements finis, il nous suffit de montrer qu'il existe deux points du graphe de f définissant un segment de pente λ. On va alors s'intéresser aux pentes des segments issus des deux points

1. Le lecteur pourra trouver le lemme de Sard pour les fonctions à plusieurs variables sous forme d'exercices dans GONNORD (S.) & TOSEL (N.), *Thèmes d'Analyse pour l'Agrégation, Calcul différentiel*, Ellipses 1998.

$(a, f(a))$ et $(b, f(b))$. On pose donc $\varphi(t) = \dfrac{f(t) - f(a)}{t - a}$ pour $t \in]a, b]$ et $\psi(t) = \dfrac{f(b) - f(t)}{b - t}$ pour $t \in [a, b[$. En fait, φ (resp. ψ) se prolonge en a (resp. en b) par $\varphi(a) = f'(a)$ (resp. $\psi(b) = f'(b)$) et sont donc continues sur $[a, b]$. Les images de l'intervalle $[a, b]$ par les applications φ et ψ sont donc deux intervalles qui contiennent respectivement $f'(a)$ et $f'(b)$ et qui se coupent puisqu'ils contiennent tous les deux le taux d'accroissement $\dfrac{f(b) - f(a)}{b - a} = \varphi(b) = \psi(a)$. Par conséquent leur réunion est un intervalle qui contient l'intervalle délimité par $f'(a)$ et $f'(b)$. En particulier, l'un des deux au moins contient λ. La formule des accroissements finis permet de conclure.

Voici une seconde solution. Posons $\mathrm{T} = \{(x, y) \in \mathrm{I}^2 \mid y < x\}$ et soit $\psi : \mathrm{T} \to \mathbb{R}$ définie par

$$\psi(x, y) = \frac{f(y) - f(x)}{y - x}.$$

On a

$$\psi(\mathrm{T}) \subset f'(\mathrm{I}) \subset \overline{\psi(\mathrm{T})}.$$

La première inclusion provient de la formule des accroissements finis, et la seconde de la définition de la dérivée comme limite de taux d'accroissements. Or T est connexe (et même convexe) et ψ est continue. Donc $\psi(\mathrm{T})$ est un intervalle de $\mathbb{R}$ et son adhérence $\overline{\psi(\mathrm{T})}$ est l'intervalle fermé qui a les mêmes extrémités. Dans ces conditions $f'(\mathrm{I})$ ne peut être qu'un intervalle. ◁

Il semble que ce théorème fut énoncé par Serret en 1868 et prouvé par Bonnet. Darboux l'utilise en 1875 pour prouver que si f' est une fonction intégrable au sens de Riemann sur $[a, b]$, alors $\int_a^b f' = f(b) - f(a)$.

Si $f : \mathbb{R} \to \mathbb{R}$ est dérivable, la formule des accroissements finis affirme que tout taux d'accroissement $\dfrac{f(y) - f(x)}{y - x}$ s'écrit $f'(c)$ pour un point c de $]x, y[$. L'exercice suivant montre que les polynômes de degré $\leqslant 2$ sont les seules fonctions pour lesquelles le point c peut toujours être pris au milieu du segment $[x, y]$.

4.30. Sur la formule des accroissements finis

Soit a et b deux réels. Trouver les fonctions $f : \mathbb{R} \to \mathbb{R}$ dérivables telles que, si $x \neq y$,

$$f'(ax + by) = \frac{f(y) - f(x)}{y - x}.$$

(École polytechnique)

▷ **Solution.**

Soit f une fonction dérivable vérifiant la propriété de l'énoncé. On va commencer par étudier la régularité de f et notamment montrer que f est nécessairement de classe $\mathcal{C}^\infty$. Pour cela, il va suffire d'exprimer f' en fonction de f. On change de variable en posant $X = ax + by$ et on élimine x.

Dans le cas particulier où $a = 0$, on obtient, en fixant y quelconque, $f(x) = f'(by)x + f(y) - f'(by)y$ pour tout $x \in \mathbb{R}$. Cela montre que f est affine. Réciproquement, on peut observer que les fonctions affines conviennent et cela quelles que soient les valeurs de a et de b. Dans la suite on suppose $a \neq 0$. On a alors, pour tout $X \in \mathbb{R}$ et tout y réel tel que $(a+b)y - X \neq 0$,

$$f'(X) = \frac{f(y) - f\left(\dfrac{X - by}{a}\right)}{y - \dfrac{X - by}{a}} = a\frac{f(y) - f\left(\dfrac{X - by}{a}\right)}{(a+b)y - X}.$$

• Si $a + b \neq 0$, cette relation prouve que f' est dérivable sur $\mathbb{R}$. Par suite f est deux fois dérivable, donc par la relation ci-dessus f' aussi. Une récurrence immédiate montre que f est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$.

• Dans le cas où $a + b = 0$, on obtient pour tout $y \in \mathbb{R}$, $f'(X) = a\dfrac{f(y + \frac{X}{a}) - f(y)}{X}$, ce qui donne, en faisant tendre X vers 0, $\lim\limits_{X \to 0} f'(X) = f'(y)$. Cela implique que la fonction f' est constante, et donc que f est affine.

On continue l'étude dans le cas $a + b \neq 0$. Comme f est de classe $\mathcal{C}^\infty$ on peut dériver par rapport à x, puis par rapport à y, l'égalité $(y - x)f'(ax + by) = f(y) - f(x)$. On obtient

$$\begin{cases} -f'(ax + by) + a(y - x)f''(ax + by) &= -f'(x) \\ f'(ax + by) + b(y - x)f''(ax + by) &= f'(y), \end{cases}$$

En multipliant la première relation par $-b$, la seconde par a et en faisant la somme on trouve $(a + b)f'(ax + by) = bf'(x) + af'(y)$. En dérivant

de nouveau cette relation par rapport à x et y, on a alors, pour tout $(x, y) \in \mathbb{R}^2$,

$$\begin{cases} (a+b)af''(ax+by) = bf''(x) \\ (a+b)bf''(ax+by) = af''(y), \end{cases}$$

Il en résulte que $f''(y) = \dfrac{b^2}{a^2} f''(x)$ pour tout $(x, y) \in \mathbb{R}^2$. Si $\dfrac{b^2}{a^2} = 1$, on en déduit que f'' est constante, *i.e.* que f est une fonction polynôme de degré $\leqslant 2$. Si $\dfrac{b^2}{a^2} \neq 1$, on obtient, en prenant $y = x$ dans la relation précédente, $f'' = 0$; donc f est affine.

Il reste à effectuer la synthèse de ce travail. On a déjà dit plus haut que les fonctions affines conviennent. Soit $f : x \longmapsto \alpha x^2 + \beta x + \gamma$ une fonction polynôme de degré $\leqslant 2$. On a

$$f'(ax+by) = 2\alpha(ax+by) \text{ et } \frac{f(y)-f(x)}{y-x} = \alpha(x+y) + \beta.$$

On en déduit que, si $(a, b) \neq (\frac{1}{2}, \frac{1}{2})$, il faut $\alpha = 0$ et que sinon α, β, γ peuvent être quelconques.

Conclusion. Si $(a, b) \neq (\frac{1}{2}, \frac{1}{2})$, les seules fonctions f qui conviennent sont les fonctions affines. Si $(a, b) = (\frac{1}{2}, \frac{1}{2})$, on trouve les polynômes de degré inférieur ou égal à 2. ◁

La dérivation D *est une application linéaire de* $\mathcal{C}^1([0,1], \mathbb{R})$ *dans* $\mathcal{C}^0([0,1], \mathbb{R})$ *telle que, pour toute fonction* f *qui admet en un point* x_0 *un maximum local, on a* $\mathrm{D}(f)(x_0) = 0$. *L'exercice qui suit recherche tous les opérateurs ayant cette propriété sur l'espace* $\mathcal{C}^2([0,1], \mathbb{R})$.

4.31. Maximum local

1. Soit $\mathrm{T} : \mathcal{C}^2([0,1], \mathbb{R}) \to \mathcal{C}^0([0,1], \mathbb{R})$ une application linéaire vérifiant : pour tout $f \in \mathcal{C}^2([0,1], \mathbb{R})$ et pour tout $x_0 \in]0,1[$ tel que f possède un maximum local en x_0, on a $\mathrm{T}(f)(x_0) = 0$. Montrer qu'il existe $\varphi \in \mathcal{C}^0([0,1], \mathbb{R})$ telle que pour tout $f \in \mathcal{C}^2([0,1], \mathbb{R})$, $\mathrm{T}(f) = \varphi f'$.

2. On remplace dans la question 1 l'espace $\mathcal{C}^2([0,1], \mathbb{R})$ par $\mathcal{C}^0([0,1], \mathbb{R})$. Montrer que $\mathrm{T} = 0$.

(École normale supérieure)

▷ **Solution.**

1. Trouver l'application φ qui va convenir est chose facile : il suffit de prendre f telle que $f' = 1$, par exemple $f = \mathrm{Id}_{[0,1]}$. On pose donc $\varphi = \mathrm{T}(\mathrm{Id}_{[0,1]})$.

Remarquons que si f est une fonction constante, on a par hypothèse $\mathrm{T}(f)(x_0) = 0$ pour tout $x_0 \in]0,1[$ et donc $\mathrm{T}(f) = 0$, par continuité. Soit $f \in \mathcal{C}^2([0,1],\mathbb{R})$ et $x_0 \in]0,1[$. Considérons la fonction $g \in \mathcal{C}^2([0,1],\mathbb{R})$ définie par

$$g(x) = f(x) - f(x_0) - (x - x_0)f'(x_0).$$

D'après la formule de Taylor-Young, on a, au voisinage de x_0,

$$g(x) = \frac{1}{2}(x - x_0)^2 f''(x_0) + o(x - x_0)^2.$$

Si $f''(x_0) < 0$, g s'annule en x_0 et est négative au voisinage de x_0. Elle possède en x_0 un maximum relatif et on a donc $\mathrm{T}(g)(x_0) = 0$. Or la linéarité de T ainsi que la nullité de T pour les fonctions constantes conduisent à $\mathrm{T}(g) = \mathrm{T}(f) - f'(x_0)\varphi$. L'égalité $\mathrm{T}(g)(x_0) = 0$ s'écrit donc $\mathrm{T}(f)(x_0) = f'(x_0)\varphi(x_0)$. En considérant la fonction $-f$, on obtient le même résultat si $f''(x_0) < 0$.

Regardons maintenant le cas où $f''(x_0) = 0$. On distingue deux cas. Si f'' est nulle sur un voisinage I de x_0, la fonction g est nulle sur I donc possède encore un maximum relatif en x_0 et le même raisonnement s'applique. Dans le cas où x_0 n'est pas intérieur à l'ensemble des zéros de f, on peut trouver une suite (y_n) de points de $]0,1[$ qui converge vers x_0 avec $f''(y_n) \neq 0$ pour tout n. D'après ce qui précède, on a $\mathrm{T}(f)(y_n) = f'(y_n)\varphi(y_n)$ pour tout n. La continuité de $\mathrm{T}(f)$, f' et φ entraîne que $\mathrm{T}(f)(x_0) = f'(x_0)\varphi(x_0)$.

On a donc, pour tout $x_0 \in]0,1[$, $\mathrm{T}(f)(x_0) = f'(x_0)\varphi(x_0)$. Toutes les fonctions qui interviennent étant continues sur $[0,1]$, on en déduit que $\mathrm{T}(f) = f'\varphi$. On a donc déterminé $\varphi \in \mathcal{C}^0([0,1],\mathbb{R})$ tel que T soit l'application $f \longmapsto \varphi f'$. On peut remarquer, réciproquement, que toute application de ce type a la propriété voulue.

2. On démontre, comme précédemment, que T est nulle pour les fonctions constantes.

Soit $x_0 \in]0,1[$. Si $f \in \mathcal{C}^0([0,1],\mathbb{R})$ s'annule en x_0, la fonction $-f^2$ possède un maximum relatif en x_0. On a donc $\mathrm{T}(-f^2)(x_0) = 0$ et $\mathrm{T}(f^2)(x_0) = 0$. Si f est une fonction positive qui s'annule en x_0, on obtient $\mathrm{T}(f)(x_0) = \mathrm{T}((\sqrt{f})^2)(x_0) = 0$, d'après ce qui précède, puisque $\sqrt{f}$ est continue. Enfin, si f est un fonction quelconque s'annulant en x_0, on écrit $f = \dfrac{f + |f|}{2} - \dfrac{|f| - f}{2}$, différence de deux fonctions continues et positives, s'annulant en x_0. Par linéarité, on obtient $\mathrm{T}(f)(x_0) = 0$.

Si maintenant f est quelconque, on pose $g = f - f(x_0)$. L'application T étant nulle pour les fonctions constantes, on a $\mathrm{T}(f)(x_0) = \mathrm{T}(g)(x_0) = 0$, puisque $g(x_0) = 0$.

On a donc, pour tout $x \in]0,1[$, $\mathrm{T}(f)(x) = 0$. Par continuité, on en déduit que $\mathrm{T}(f) = 0$. Nous avons démontré que $\mathrm{T} = 0$. ◁

L'exercice suivant montre qu'une fonction quelconque (sans aucune hypothèse de régularité) définie sur $\mathbb{R}$ ne peut pas avoir beaucoup de maxima locaux distincts.

4.32. Dénombrabilité des maxima locaux

Soit $f : \mathbb{R} \to \mathbb{R}$ et A l'ensemble des réels a tels que f admet un maximum local en a. Montrer que $f(\mathrm{A})$ est au plus dénombrable.
(École polytechnique)

▷ **Solution.**

Soit $\alpha \in f(\mathrm{A})$. Comme α est un maximum local, il existe $x_\alpha \in \mathbb{R}$ tel que $f(x_\alpha) = \alpha$ et $a_\alpha < x_\alpha < b_\alpha$ tels que pour tout $x \in [a_\alpha, b_\alpha]$, $f(x) \leqslant \alpha = f(x_\alpha)$. (Notons que l'existence de la fonction $\alpha \mapsto x_\alpha$ utilise l'axiome du choix, que l'on peut contourner dans le cas où f est continue.) Comme $\mathbb{Q}$ est dense dans $\mathbb{R}$, on peut choisir a_α et b_α dans $\mathbb{Q}$. On a $\alpha = \sup_{x \in [a_\alpha, b_\alpha]} f(x)$ de sorte que l'application

$$\varphi : \begin{array}{ccc} f(\mathrm{A}) & \longrightarrow & \mathbb{Q}^2 \\ \alpha & \longmapsto & (a_\alpha, b_\alpha) \end{array}$$

est injective. Or $\mathbb{Q}^2$ est dénombrable, donc $f(\mathrm{A})$ est au plus dénombrable. ◁

On en déduit facilement qu'une application continue $f : \mathbb{R} \to \mathbb{R}$ qui a un extremum local en tout point est constante. En effet, si B *est l'ensemble des points où f a un minimum, $f(\mathrm{B})$ est aussi au plus dénombrable. Donc $f(\mathbb{R}) = f(\mathrm{A}) \cup f(\mathrm{B})$ est au plus dénombrable. Mais si f est continue, son image est un intervalle par le théorème des valeurs intermédiaires, donc un singleton ou un ensemble indénombrable.*

L'énoncé suivant a été tiré du volume «Fonctions d'une variable réelle» du célèbre traité de Nicolas Bourbaki.

4.33. Majoration de $f^2 + f'^2$

Soit $f : \mathbb{R} \to \mathbb{R}$ deux fois dérivable. On suppose que $f^2(x) \leqslant a$ et $f'^2(x) + f''^2(x) \leqslant b$ pour tout réel x. Montrer que, pour tout réel x,

$$f^2(x) + f'^2(x) \leqslant \max(a, b).$$

(École polytechnique)

▷ **Solution.**

Posons $g(x) = f^2(x) + f'^2(x)$ et supposons, par l'absurde, qu'il existe $x_0 \in \mathbb{R}$ tel que $g(x_0) > \max(a, b)$. L'idée est de montrer que g admet alors un maximum local strictement supérieur à $\max(a, b)$, maximum en lequel il y aura une contradiction.

Supposons que la fonction g reste supérieure à $g(x_0)$ sur tout l'intervalle $]-\infty, x_0]$. On a alors, pour $x \leqslant x_0$, $f'^2(x) \geqslant g(x_0) - a > 0$. Donc f' ne s'annule pas sur $]-\infty, x_0]$ et y garde un signe constant. Si elle est positive, $f'(x) \geqslant \sqrt{g(x_0) - a} > 0$ pour tout $x \leqslant x_0$, et si elle est négative, $f'(x) \leqslant -\sqrt{g(x_0) - a} < 0$ pour tout $x \leqslant x_0$. Les deux cas sont impossibles puisque f est bornée. Il existe donc $x_1 < x_0$ tel que $g(x_1) < g(x_0)$. Un raisonnement analogue permet d'établir qu'il existe $x_2 > x_0$ tel que $g(x_2) < g(x_0)$.

La fonction g est continue sur le compact $[x_1, x_2]$ donc atteint son maximum en un point $\alpha \in]x_1, x_2[$. En ce point $g'(\alpha) = 0$, c'est-à-dire $[f(\alpha) + f''(\alpha)]f'(\alpha) = 0$.

- Si $f'(\alpha) = 0$, on a $g(\alpha) = f^2(\alpha) \leqslant a$, ce qui est impossible car $g(\alpha) \geqslant g(x_0) > a$.
- Si $f(\alpha) + f''(\alpha) = 0$, on a $g(\alpha) = f'^2(\alpha) + f''^2(\alpha) \leqslant b$: c'est aussi exclu.

L'hypothèse initiale est absurde et on a donc l'inégalité demandée. ◁

Voici deux énoncés qui concernent des questions d'interpolation.

4.34. Interpolation par splines cubiques

Soit $[a, b]$ un segment de $\mathbb{R}$ et $\sigma = \{x_0 = a < x_1 < \cdots < x_n = b\}$ une subdivision de $[a, b]$. On considère l'ensemble $\mathcal{S}$ des applications de classe $\mathcal{C}^2$ de $[a, b]$ dans $\mathbb{R}$ dont la restriction à chaque segment $[x_i, x_{i+1}]$ est un polynôme de degré au plus 3. Un élément de $\mathcal{S}$ est appelé une *fonction spline cubique*.

1.a. Déterminer la nature de $\mathcal{S}$ et sa dimension.

b. Pour $y = (y_0, \dots, y_n) \in \mathbb{R}^{n+1}$, on pose

$$\mathcal{T}(y) = \{\varphi \in \mathcal{S}, \forall i \in [\![0, n]\!],\ \varphi(x_i) = y_i\}.$$

Déterminer la nature de $\mathcal{T}(y)$ et sa dimension.

c. Pour $(\lambda, \mu) \in \mathbb{R}^2$, démontrer l'existence et l'unicité d'une fonction $\varphi \in \mathcal{T}(y)$ telle que $\varphi'(a) = \lambda$ et $\varphi'(b) = \mu$ (considérer, pour $\varphi \in \mathcal{T}(0)$ telle que $\varphi'(a) = \varphi'(b) = 0$, l'intégrale $\int_a^b (\varphi''(x))^2\,\mathrm{d}x$).

2. Soit f une application $\mathcal{C}^2$ de $[a, b]$ dans $\mathbb{R}$ et φ l'élément de $\mathcal{S}$ tel que $\varphi(x_i) = f(x_i)$ pour tout $i \in [\![0, n]\!]$, $\varphi'(a) = f'(a)$ et $\varphi'(b) = f'(b)$. Calculer $\int_a^b (\varphi'' - f'')^2$, en fonction d'intégrales portant sur f, φ et leurs dérivées. Interpréter.

(École polytechnique)

▷ **Solution.**

1.a. $\mathcal{S}$ est clairement un sous-espace vectoriel de $\mathcal{C}^2([a, b], \mathbb{R})$. Si φ appartient à $\mathcal{S}$, alors, pour tout $i \in [\![0, n-1]\!]$, la restriction φ_i de φ à $[x_i, x_{i+1}]$ est un polynôme de degré inférieur ou égal à 3. Il existe donc $(\alpha_i, \beta_i, \gamma_i, \delta_i) \in \mathbb{R}^4$ tel que, pour tout $x \in [x_i, x_{i+1}]$,

$$\varphi_i(x) = \alpha_i(x - x_i)^3 + \beta_i(x - x_i)^2 + \gamma_i(x - x_i) + \delta_i.$$

La fonction φ est de classe $\mathcal{C}^2$ sur $[a, b]$ si, pour tout $i \in [\![0, n-2]\!]$, φ_i, φ_i' et φ_i'' coïncident avec φ_{i+1}, φ_{i+1}' et φ_{i+1}'' en x_{i+1}. Cette condition équivaut à

$$\begin{cases} \varphi_i(x_{i+1}) = \delta_{i+1} \\ \varphi_i'(x_{i+1}) = \gamma_{i+1} \\ \varphi_i''(x_{i+1}) = \beta_{i+1}, \end{cases}$$

c'est-à-dire à

$$\begin{cases} \alpha_i h_i^3 + \beta_i h_i^2 + \gamma_i h_i + \delta_i = \delta_{i+1} \\ 3\alpha_i h_i^2 + 2\beta_i h_i + \gamma_i = \gamma_{i+1} \\ 6\alpha_i h_i + 2\beta_i = 2\beta_{i+1}, \end{cases}$$

où $h_i = x_{i+1} - x_i$.

Pour définir φ on a donc $4n$ degrés de liberté (pour choisir les n polynômes de degré $\leqslant 3$) et $3(n-1)$ contraintes (pour assurer un raccord de classe $\mathcal{C}^2$ en $n-1$ points). Ces contraintes se traduisent par des formes linéaires. Si ces formes sont indépendantes, ce qui intuitivement semble être le cas, la dimension de $\mathcal{S}$ sera égale à $4n - 3(n-1) = n + 3$. Utilisons le langage des systèmes linéaires. En posant $\mathrm{X} = (\alpha_0, \dots, \alpha_{n-1}, \beta_0, \dots, \beta_{n-1}, \gamma_0, \dots, \gamma_{n-1}, \delta_0, \dots, \delta_{n-1}) \in \mathbb{R}^{4n}$, les $3(n-1)$

équations ci-dessus se traduisent par un système linéaire homogène $AX = 0$ où A est une matrice de taille $(3(n-1), 4n)$. L'espace $\mathcal{S}$ est clairement isomorphe au noyau de A et sa dimension est égale à $4n - \text{rg A}$.

On voit ci-dessus que les inconnues β_{i+1} et γ_{i+1} et δ_{i+1} sont parfaitement déterminées si on connaît $\alpha_i, \beta_i, \gamma_i$ et δ_i. On peut alors se donner arbitrairement $\alpha_0, \beta_0, \gamma_0$ et δ_0 (cela revient à choisir le polynôme sur le premier intervalle). Les valeurs de β_1, γ_1 et δ_1 sont alors fixées mais on peut choisir librement α_1 (cela signifie que le raccord $\mathcal{C}^2$ établi, il n'y a plus qu'un degré de liberté sur le second intervalle). Cela détermine alors uniquement β_2, γ_2 et δ_2. On choisit α_2 de manière quelconque et on poursuit de proche en proche. Bref, on a ainsi montré qu'on peut choisir $\alpha_0, \beta_0, \gamma_0, \delta_0, \alpha_1, \ldots, \alpha_n$ comme inconnues secondaires et que le système est alors un système de Cramer en les inconnues restantes. Il en résulte donc bien que la dimension de $\mathcal{S}$ vaut $n+3$.

b. On garde les mêmes notations et on identifie toujours $\mathcal{S}$ avec un sous-espace de $\mathbb{R}^{4n}$. Le vecteur X doit vérifier, pour que φ appartienne à $\mathcal{T}(y)$, les conditions supplémentaires

$$\begin{cases} \text{pour tout } i \in [\![0, n-1]\!], \delta_i = y_i \text{ et} \\ \alpha_{n-1} h_{n-1}^3 + \beta_{n-1} h_{n-1}^2 + \gamma_{n-1} h_{n-1} + \delta_{n-1} = y_n. \end{cases}$$

On rajoute donc $n+1$ équations (avec second membre) au système précédent, pour obtenir un système de $3(n-1)+n+1 = 4n-2$ équations. On peut déjà dire que si $\mathcal{T}(y)$ n'est pas vide, il s'agit d'un sous-espace affine de S dont l'espace vectoriel associé est $\mathcal{T}(0)$.

Cette fois, $\delta_0, \ldots, \delta_{n-1}$ sont déterminés. On choisit arbitrairement α_{n-1} et β_{n-1}. La deuxième équation ci-dessus détermine γ_{n-1}. On voit alors de proche en proche que tous les autres coefficients sont déterminés. Le système possède donc des solutions : $\mathcal{T}(y)$ n'est pas vide. Puisque deux inconnues peuvent être choisies arbitrairement, on en déduit que $\mathcal{T}(y)$ est un espace affine de dimension 2.

Conclusion. Pour tout $y \in \mathbb{R}^{n+1}$, $\mathcal{T}(y)$ est un espace affine de dimension 2.

c. • Considérons, pour commencer, comme nous y invite l'énoncé, une application $\varphi \in \mathcal{T}(0)$ telle que $\varphi'(a) = \varphi'(b) = 0$ et calculons $\mathrm{I} = \displaystyle\int_a^b (\varphi''(x))^2 \, \mathrm{d}x$. Avec les notations précédentes, on a

$$\mathrm{I} = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} (\varphi_i''(x))^2 \, \mathrm{d}x.$$

En intégrant par parties, on obtient, pour tout $i \in [\![0, n-1]\!]$,

$$\int_{x_i}^{x_{i+1}} (\varphi_i''(x))^2 \, \mathrm{d}x = [\varphi_i'(x)\varphi_i''(x)]_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} \varphi_i'(x)\varphi_i'''(x) \mathrm{d}x.$$

La dernière intégrale vaut

$$6\alpha_i \int_{x_i}^{x_{i+1}} \varphi_i'(x)\mathrm{d}x = 6\alpha_i(\varphi_i(x_{i+1}) - \varphi_i(x_i)) = 0.$$

Il reste donc

$$\mathrm{I} = \sum_{i=0}^{n-1} [\varphi'(x)\varphi''(x)]_{x_i}^{x_{i+1}} = \varphi'(b)\varphi''(b) - \varphi'(a)\varphi''(b) = 0.$$

Puisque I est nulle, la fonction φ'' est nulle. La fonction φ' est constante et s'annule en a : c'est la fonction nulle. De même, φ qui est constante et s'annule en a est la fonction nulle.

• X doit vérifier par rapport à la question précédente deux conditions supplémentaires. On a donc affaire à un système de $4n$ équations à $4n$ inconnues.

Considérons le système homogène associé. C'est le système que doivent vérifier les coefficients d'une fonction φ qui appartient à $\mathcal{T}(0)$ et vérifie de plus $\varphi'(a) = \varphi'(b) = 0$. D'après ce qui précède, une telle fonction est la fonction nulle. Autrement dit, le système homogène n'a pas d'autre solution que la solution nulle. Le système est donc de Cramer. Il a donc, pour tout $y \in \mathbb{R}^{n+1}$ et tout $(\lambda, \mu) \in \mathbb{R}^2$ une solution unique.

Conclusion. Pour tout $y \in \mathbb{R}^{n+1}$, il existe, pour tout $(\lambda, \mu) \in \mathbb{R}^2$ une unique fonction $\varphi \in \mathcal{T}(y)$ telle que $\varphi'(a) = \lambda$ et $\varphi'(b) = \mu$.

2. L'existence et l'unicité de φ résulte de la question précédente, avec $y_i = f(x_i)$ pour tout $i \in [\![0, n]\!]$, $\lambda = f'(a)$ et $\mu = f'(b)$.

• Montrons que $\displaystyle\int_a^b (\varphi'' - f'')^2 = \int_a^b (f'')^2 - \int_a^b (\varphi'')^2$.

Le même raisonnement que dans la question 1.c montre que

$$\begin{aligned}
\int_a^b (f'' - \varphi'')\varphi'' &= \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} (f'' - \varphi'')\varphi'' \\
&= \sum_{i=0}^{n-1} \left([(f' - \varphi')\varphi'']_{x_i}^{x_{i+1}} - \int_{x_i}^{x_{i+1}} (f' - \varphi')\varphi''' \right) \\
&= [(f' - \varphi')\varphi'']_a^b - \sum_{i=0}^{n-1} 6\alpha_i \int_{x_i}^{x_{i+1}} (f' - \varphi') = 0,
\end{aligned}$$

car $f' - \varphi'$ s'annule en a et b et $f - \varphi$ s'annule en x_i pour tout $i \in [\![0, n]\!]$.

On en déduit que $\displaystyle\int_a^b f''\varphi'' = \int_a^b (\varphi'')^2$, puis que

$$\begin{aligned}
\int_a^b (f'' - \varphi'')^2 &= \int_a^b (f'')^2 - 2\int_a^b f''\varphi'' + \int_a^b (\varphi'')^2 \\
&= \int_a^b (f'')^2 - \int_a^b (\varphi'')^2.
\end{aligned}$$

• Interprétons ce résultat. On a

$$\int_a^b (f'')^2 - \int_a^b (\varphi'')^2 = \int_a^b (f'' - \varphi'')^2 \geqslant 0.$$

L'égalité $\int_a^b (f'')^2 = \int_a^b (\varphi'')^2$ est obtenue quand $\int_a^b (f'' - \varphi'')^2 = 0$. Alors $f'' - \varphi''$ est nulle ; $f' - \varphi'$ est constante et donc nulle, car elle s'annule en a ; $f - \varphi$ est constante, donc nulle, car elle s'annule en a. Il n'y a donc égalité entre $\int_a^b (f'')^2$ et $\int_a^b (\varphi'')^2$ que si $f = \varphi$.

Soit $y \in \mathbb{R}^{n+1}$ et $(\lambda, \mu) \in \mathbb{R}^2$ fixés et $\mathcal{U}$ l'ensemble des applications de classe $\mathcal{C}^2$ de $[a, b]$ dans $\mathbb{R}$, vérifiant $f(x_i) = y_i$, pour tout $i \in [\![0, n]\!]$, $f'(a) = \lambda$ et $f'(b) = \mu$. L'ensemble $\mathcal{U}$ contient une unique fonction spline cubique φ. Ce qui précède démontre que, pour tout $f \in \mathcal{U}$, on a

$$\int_a^b (f'')^2 \geqslant \int_a^b (\varphi'')^2,$$

avec égalité seulement si $f = \varphi$. L'application φ réalise le minimum sur $\mathcal{U}$ de $\int_a^b (f'')^2$. On peut dire que c'est la fonction de $\mathcal{U}$ qui est la moins «bosselée». ◁

4.35. Interpolation d'Hermite

1. Soit $f : [0, 1] \to \mathbb{R}$ une fonction de classe $\mathcal{C}^4$, telle que $f(0) = f'(0) = f(1) = f'(1) = 0$. Montrer que pour tout $x \in [0, 1]$, il existe $\xi \in]0, 1[$ tel que

$$f(x) = \frac{f^{(4)}(\xi)}{24} x^2 (1 - x)^2.$$

2. Soit $a = x_0 < x_1 < \cdots < x_n = b$ une subdivision du segment $[a, b]$ et $f : [a, b] \longmapsto \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. Montrer l'existence de $\Phi : [a, b] \to \mathbb{R}$ de classe $\mathcal{C}^1$, polynomiale de degré au plus trois sur chacun des segments $[x_k, x_{k+1}]$ et telle que, pour tout $k \in [\![0, n]\!]$, on ait $\Phi(x_k) = f(x_k)$ et $\Phi'(x_k) = f'(x_k)$.

3. Dans les conditions de la question 2, on suppose en outre que f est de classe $\mathcal{C}^4$. Montrer que

$$\|f - \Phi\|_\infty \leqslant \frac{\|f^{(4)}\|_\infty}{384} \sup_{1 \leqslant i \leqslant n-1} (x_{i+1} - x_i)^2.$$

(École polytechnique)

▷ **Solution.**

1. Si $x = 0$ ou 1, la propriété à démontrer est évidente : on prend ξ quelconque dans $]0,1[$. On suppose dans la suite que $x \in]0,1[$ et on s'inspire de la preuve classique du théorème de Taylor-Lagrange. On considère la fonction φ définie sur $[0,1]$ par $\varphi(t) = f(t) - \dfrac{A}{24}t^2(1-t)^2$, A étant choisi pour que $\varphi(x) = 0$. Comme f, la fonction φ est de classe $\mathcal{C}^4$ et vérifie $\varphi(0) = \varphi(1) = \varphi(x) = \varphi'(0) = \varphi'(1) = 0$. En appliquant le théorème de Rolle à φ sur les intervalles $[0,x]$ et $[x,1]$, on montre que φ' s'annule en $c \in]0,x[$ et $d \in]x,1[$: φ' s'annule donc quatre fois au moins sur $[0,1]$. En appliquant de nouveau le théorème de Rolle aux fonctions φ', φ'' et $\varphi^{(3)}$, on montre successivement que φ'' s'annule au moins trois fois, $\varphi^{(3)}$ au moins deux fois et $\varphi^{(4)}$ au moins une fois sur $]0,1[$: soit $\xi \in]0,1[$ tel que $\varphi^{(4)}(\xi) = 0$. Pour tout $t \in [0,1]$, $\varphi^{(4)}(t) = f^{(4)}(t) - A$ et donc $A = f^{(4)}(\xi)$. Sachant que $\varphi(x) = 0$, on obtient $f(x) = \dfrac{A}{24}x^2(1-x)^2 = \dfrac{f^{(4)}(\xi)}{24}x^2(1-x)^2$.

2. Montrons que, pour tout $k \in [\![0, n-1]\!]$, il existe $\Phi_k \in \mathbb{R}_3[X]$ tel que $\Phi_k(x_k) = f(x_k)$, $\Phi_k(x_{k+1}) = f(x_{k+1})$, $\Phi_k'(x_k) = f'(x_k)$ et $\Phi'(x_{k+1}) = f'(x_{k+1})$. Soit $u_k : \mathbb{R}_3[X] \to \mathbb{R}^4$ l'application linéaire qui à tout polynôme $P \in \mathbb{R}_3[X]$ associe $u_k(P) = (P(x_k), P(x_{x+1}), P'(x_k), P'(x_{x+1}))$. Cette application u_k est injective car si $P \in \operatorname{Ker} u_k$, P admet x_k et x_{k+1} comme racines avec une multiplicité $\geqslant 2$. Comme $\deg P \leqslant 3$, cela impose $P = 0$. Comme $\mathbb{R}_3[X]$ et $\mathbb{R}^4$ sont de même dimension, égale à 4, l'application u_k est un isomorphisme. Cela garantit l'existence, et même l'unicité, de Φ_k.

Soit alors $\Phi : [a,b] \longrightarrow \mathbb{R}$ l'application dont la restriction à chaque segment $[x_k, x_{k+1}]$ est Φ_k. Elle est de classe $\mathcal{C}^1$ sur chaque intervalle $]x_k, x_{k+1}[$. De plus, par construction, pour tout $k \in [\![0, n-1]\!]$, on a $\Phi(x_k) = \lim\limits_{x \to x_k} \Phi(x) = f(x_k)$, donc Φ est continue sur $[a,b]$; enfin, toujours par construction, Φ est dérivable à gauche et à droite en x_k avec $\Phi_d'(x_k) = \Phi_g'(x_k) = f'(x_k)$. Donc Φ est dérivable en chaque point x_k et on a $\lim\limits_{x \to x_k} \Phi'(x) = f'(x_k) = \Phi'(x_k)$. Donc Φ est de classe $\mathcal{C}^1$ sur $[a,b]$.

3. Considérons la restriction f_k de $f - \Phi$ à $[x_k, x_{k+1}]$; f_k est de classe $\mathcal{C}^4$ sur $[x_k, x_{k+1}]$ et vérifie $f_k(x_k) = f_k(x_{k+1}) = f_k'(x_k) = f_k'(x_{k+1}) = 0$. On va se ramener au segment $[0,1]$ pour appliquer la question 1. La fonction g_k définie sur $[0,1]$ par $g_k(t) = f_k((1-t)x_k + tx_{k+1})$ vérifie les conditions de la question 1. Pour tout $x \in [x_k, x_{k+1}]$, il existe $t \in [0,1]$ tel que $f_k(x) = g_k(t)$ et d'après 1, il existe $\xi \in]0,1[$ tel que $g_k(t) = \dfrac{g_k^{(4)}(\xi)}{24}t^2(1-t)^2$. Sachant que $g_k^{(4)}(t) = (x_{k+1} - x_k)^4 f^{(4)}((1-t)x_k + tx_{k+1})$, on en déduit que

$$|f_k(x)| \leqslant \frac{(x_{k+1} - x_k)^4}{24}\|f^{(4)}\|_\infty t^2(1-t)^2,$$

puis que $|f_k(x)| \leqslant \dfrac{(x_{k+1}-x_k)^4}{384}\|f^{(4)}\|_\infty$, car $\sup\limits_{t\in[0,1]} t^2(1-t)^2 = \dfrac{1}{4^2} = \dfrac{1}{16}$.

Conclusion. $\boxed{\|f-\Phi\|_\infty \leqslant \dfrac{\|f^{(4)}\|_\infty}{384} \sup\limits_{1\leqslant i\leqslant n-1}(x_{i+1}-x_i)^2}$. ◁

Les exercices qui suivent font appel aux différentes formules de Taylor. Les développements «tayloriens» apparaissent dans les travaux de Taylor sur l'interpolation polynomiale (1715) et ceux de Mac Laurin (1742). C'est Lagrange qui établit en 1797 la célèbre formule dite de Taylor-Lagrange : soit $f : [a,b] \to \mathbb{R}$ de classe $\mathcal{C}^n$ sur $[a,b]$ et $n+1$ fois dérivable sur $]a.b[$. Il existe alors $c \in]a,b[$ tel que

$$f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(b-a)^k + \frac{f^{(n+1)}(c)}{(n+1)!}(b-a)^{n+1}.$$

Lagrange pense alors que toute fonction indéfiniment dérivable peut se développer en série de Taylor. C'est Cauchy qui donnera le premier un contre-exemple. Un exemple classique de fonction f de classe $\mathcal{C}^\infty$ dont la série de Taylor ne converge pas vers f est fourni par $f(x) = e^{-1/x^2}$, prolongée par continuité en 0.

Lorsque f est de classe $\mathcal{C}^{n+1}$ sur $[a,b]$, le reste peut s'écrire sous forme exacte à l'aide d'une intégrale :

$$f(b) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!}(b-a)^k + \int_a^b \frac{(b-x)^n}{n!} f^{(n+1)}(x)\,dx.$$

Il est à noter que cette formule avec reste intégral est valable pour une fonction à valeurs vectorielles (et se démontre par récurrence par une simple intégration par parties).

4.36. Dérivées dominées par un polynôme

Soit $P \in \mathbb{R}[X]$ un polynôme de degré impair et $f \in \mathcal{C}^\infty(\mathbb{R},\mathbb{R})$ telle que $|f^{(n)}(x)| \leqslant |P(x)|$ pour tout $n \in \mathbb{N}$ et tout $x \in \mathbb{R}$. Que dire de f ?

(École normale supérieure)

▷ **Solution.**

Comme P est un polynôme de degré impair, on pense tout de suite à utiliser le fait qu'il admet au moins une racine réelle α. Par hypothèse, toutes les dérivées de f en α sont nulles. Pour $x \in \mathbb{R}$, on va estimer $f(x)$ en utilisant la formule de Taylor-Lagrange en α à l'ordre n. Il existe c_n

entre α et x tel que

$$f(x) = \frac{f^{n+1}(c_n)(x-\alpha)^{n+1}}{(n+1)!}.$$

On a donc $|f(x)| \leqslant \dfrac{|\mathrm{P}(c_n)||x-\alpha|^{n+1}}{(n+1)!}$, et le majorant tend vers 0 lorsque n tend vers l'infini car P est borné sur le segment $[\alpha, x]$ (ou $[x, \alpha]$). Ainsi, $f(x) = 0$ et comme cela vaut pour tout x, f est nulle. $\lhd$

Le résultat devient faux avec un polynôme de degré pair (prendre par exemple $f(x) = \sin x$ et $\mathrm{P}(x) = x^2 + 1$).

4.37. Une généralisation du théorème de Rolle

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^p$.

1. On suppose qu'il existe $n \in [\![0, p-1]\!]$ tel que $f(x) = o(x^n)$ lorsque $|x|$ tend vers $+\infty$. Montrer que $f^{(p)}$ s'annule en un point.

2. Comment modifier la preuve précédente si on suppose simplement que f est p fois dérivable sur $\mathbb{R}$?

3. Que se passe-t-il si $n = p$?

(École polytechnique)

$\rhd$ **Solution.**

1. Le cas $p = 1$ et $n = 0$ généralise le théorème de Rolle : si f tend vers 0 en $\pm\infty$, il existe un réel c tel que $f'(c) \neq 0$. En effet, dans le cas contraire, f' garderait un signe constant sur $\mathbb{R}$ (car elle est continue) et f serait strictement monotone. Elle ne pourrait donc pas tendre vers 0 à la fois en $-\infty$ et en $+\infty$.

Passons maintenant au cas général. Il suffit bien entendu de traiter le cas où $n = p-1$. Supposons par l'absurde que la fonction $f^{(p)}$ ne s'annule pas sur $\mathbb{R}$. Comme elle est continue, elle garde un signe constant sur $\mathbb{R}$ et, quitte à prendre $-f$, on peut supposer qu'elle est strictement positive. Soit a un réel. La formule de Taylor-Lagrange permet d'affirmer que pour tout réel x, il existe un réel α_x tel que

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \cdots + \frac{f^{(p-1)}(a)}{(p-1)!}(x-a)^{p-1} + \frac{f^{(p)}(\alpha_x)}{p!}(x-a)^p.$$

On divise cela par x^{p-1} et on fait tendre x vers $+\infty$ et vers $-\infty$. On obtient

$$\frac{f^{(p-1)}(a)}{(p-1)!} = -\lim_{x\to+\infty} \frac{x f^{(p)}(\alpha_x)}{p!} = -\lim_{x\to-\infty} \frac{x f^{(p)}(\alpha_x)}{p!}.$$

Mais comme $f^{(p)}(\alpha_x) > 0$, la première limite est positive et la seconde négative. Il en résulte que $f^{(p-1)}(a) = 0$. Comme cela vaut pour tout réel a, $f^{(p-1)}$ est identiquement nulle et cela contredit la stricte positivité de $f^{(p)}$.

On pouvait également argumenter de la sorte : si $f^{(p)} > 0$, $f^{(p-1)}$ est strictement croissante et il existe donc $a > 0$ et $\alpha \in \mathbb{R}$ tels que $f^{(p-1)}(x) \geqslant a$ pour tout $x \geqslant \alpha$ ou bien $f^{(p-1)}(x) \leqslant -a$ pour tout $x \leqslant \alpha$. Supposons que nous sommes dans le premier cas, $f^{(p-1)}(x) \geqslant a$ pour $x \in [\alpha, +\infty[$. Par intégrations successives, on obtient $f(x) \geqslant \mathrm{P}(x)$ avec $\mathrm{P}(x) \underset{x\to+\infty}{\sim} \dfrac{ax^{p-1}}{(p-1)!}$, d'où la contradiction recherchée.

2. Dans la preuve précédente, on a toujours le droit d'écrire la formule de Taylor-Lagrange, mais on ne dispose plus de l'argument de continuité pour dire que $f^{(p)}$ garde un signe constant sur $\mathbb{R}$. En réalité, cela reste vrai car une fonction dérivée vérifie le théorème des valeurs intermédiaires. C'est un théorème de Darboux que l'on trouvera dans l'exercice 4.29.

Il est possible de contourner l'utilisation du théorème de Darboux : comme $f^{(p)}$ ne s'annule pas, $f^{(p-1)}$ est injective d'après le théorème de Rolle et par conséquent strictement monotone (on se reportera aux propos précédant l'exercice 4.12). La deuxième solution de la première question peut alors être reprise.

3. En revanche si $p = n$ le résultat n'est plus vrai. Pour $p = 1$ il suffit de prendre une fonction bornée à dérivée strictement positive, par exemple $x \longmapsto \arctan x$ ou $x \longmapsto \mathrm{th}\, x$. On peut alors en déduire des contre-exemples pour p quelconque. En effet, soit f de classe $\mathcal{C}^1$ telle que $f(x) = o(x)$ et $f' > 0$ sur $\mathbb{R}$ et g telle que $g^{(p-1)} = f$. On a donc $g^{(p)} = f'$ ne s'annule pas sur $\mathbb{R}$. Pourtant, comme $g^{(p-1)}(x) = o(x)$ on a successivement $g^{(p-2)} = o(x^2)$, $g^{(p-3)}(x) = o(x^3), \ldots, g(x) = o(x^p)$ en $+\infty$ et en $-\infty$ (par le théorème d'intégration des relations de négligeabilité). ◁

4.38. Minoration de la dérivée seconde

Soit $f \in \mathcal{C}^1([0,1], \mathbb{R})$ une fonction deux fois dérivable sur $]0,1[$ vérifiant $f(0) = f'(0) = f'(1) = 0$ et $f(1) = 1$. Montrer qu'il existe $x \in]0,1[$ tel que $|f''(x)| \geqslant 4$.

(École normale supérieure)

▷ **Solution.**

On dispose d'une information sur f en deux points distincts 0 et 1. Pour utiliser simultanément ces deux informations, on va estimer $f\left(\frac{1}{2}\right)$ par la formule de Taylor-Lagrange à l'ordre 2 appliquée en 0 puis en 1.

Il existe donc $c \in \left]0, \frac{1}{2}\right[$ et $d \in \left]\frac{1}{2}, 1\right[$ tels que

$$f\left(\frac{1}{2}\right) = f(0) + \frac{1}{2}f'(0) + \frac{1}{8}f''(c) = \frac{1}{8}f''(c)$$

$$\text{et} \quad f\left(\frac{1}{2}\right) = f(1) - \frac{1}{2}f'(1) + \frac{1}{8}f''(d) = 1 + \frac{1}{8}f''(d).$$

Ainsi on a $f''(c) - f''(d) = 8$, de sorte que $|f''(c)| + |f''(d)| \geqslant 8$ et soit $|f''(c)| \geqslant 4$ soit $|f''(d)| \geqslant 4$. ◁

Application pratique : si un candidat à l'X court le 100 m en 10 secondes, il y a un moment où son accélération est supérieure à 40 ms^{-2}.

Une autre solution consiste à raisonner par l'absurde en supposant $|f''| < 4$ et à intégrer cette information en tenant compte des conditions initiales en 0 et 1. Cette approche permet même d'améliorer le résultat car l'hypothèse $|f''| \leqslant 4$ suffit pour obtenir la contradiction. Par conséquent, il existe toujours un point $x \in]0,1[$ tel que $|f''(x)| > 4$.

4.39. Zéros des dérivées successives

Soit $f : \mathbb{R}_+ \to \mathbb{R}$ de classe $\mathcal{C}^\infty$ avec $f(0) > 0$ et $f'(0) > 0$. On suppose que la limite de f en $+\infty$ est nulle.

1. Montrer qu'il existe $x_1 > 0$ tel que $f'(x_1) = 0$.

2. Montrer qu'il existe une suite strictement croissante $(x_n)_{n\geqslant 1}$ telle que pour tout $n \geqslant 1$, $f^{(n)}(x_n) = 0$.

(École polytechnique)

▷ **Solution.**

1. Si f' ne s'annule pas, elle garde un signe constant (par continuité) et comme $f'(0) > 0$, elle est strictement positive. Mais alors f est strictement croissante sur $\mathbb{R}_+$ et, comme $f(0) > 0$, elle ne peut pas tendre vers 0 en $+\infty$. Il existe donc $x_1 > 0$ tel que $f'(x_1) = 0$.

2. On a naturellement l'idée de construire la suite par récurrence sur n. Supposons $x_1 < x_2 < \cdots < x_n$ construits avec $f^{(k)}(x_k) = 0$ pour tout $k \in [\![1, n]\!]$. Supposons par l'absurde que $f^{(n+1)}$ ne s'annule pas sur $]x_n, +\infty[$. Elle y garde alors un signe constant, par exemple strictement positif. Donc $f^{(n)}$ est strictement croissante sur $[x_n, +\infty[$ et donc strictement positive sur $]x_n, +\infty[$. Appliquons alors la formule de Taylor-Lagrange à l'ordre n en un point $a > x_n$, pour estimer $f(x)$. Pour tout $x > a$, il existe $c \in]a, x[$ tel que

$$f(x) = f(a) + (x-a)f'(a) + \cdots + \frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1} + \frac{f^{(n)}(c)}{n!}(x-a)^n.$$

On en déduit que

$$f(x) \geqslant f(a)+(x-a)f'(a)+\cdots+\frac{f^{(n-1)}(a)}{(n-1)!}(x-a)^{n-1}+\frac{f^{(n)}(a)}{n!}(x-a)^n.$$

Or le polynôme qui minore $f(x)$ tend vers $+\infty$ lorsque x tend vers $+\infty$. D'où la contradiction avec l'hypothèse que f a une limite nulle. $\triangleleft$

4.40. Théorème de Glaeser pour une variable (1963)

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^2$, positive sur $\mathbb{R}$.

1. Trouver une condition nécessaire et suffisante pour que $\sqrt{f}$ soit dérivable sur $\mathbb{R}$.

2. On suppose que $f(0) = f'(0) = f''(0) = 0$. Soit $\alpha > 0$ et $\mathrm{M}(\alpha) = \sup_{t\in[-2\alpha,2\alpha]} |f''(t)|$. En utilisant une formule de Taylor, montrer que pour $x \in [-\alpha, \alpha]$ on a $f'^2(x) \leqslant 2f(x)\mathrm{M}(\alpha)$.

3. En déduire une condition nécessaire et suffisante pour que $\sqrt{f}$ soit de classe $\mathcal{C}^1$.

(École normale supérieure)

$\triangleright$ **Solution.**

1. Si $x_0 \in \mathbb{R}$ est tel que $f(x_0) > 0$, alors le théorème de dérivation des fonctions composées garantit que $\sqrt{f}$ est dérivable en x_0, de dérivée $\dfrac{f'(x_0)}{2\sqrt{f(x_0)}}$. En revanche si x_0 est un zéro de f on peut avoir des problèmes du fait de la non-dérivabilité de la fonction racine carrée en 0. Par exemple, si $f(x) = x^2$, on a $\sqrt{f(x)} = |x|$ qui n'est pas dérivable en 0. Elle est toutefois dérivable à droite et à gauche. On va voir que ce fait est général.

Soit x_0 un zéro de f. Comme f admet en x_0 un minimum local, on a $f'(x_0) = 0$. On va étudier le taux d'accroissement de $\sqrt{f}$ entre x_0 et x en utilisant un développement limité de f. On a $f(x) = \dfrac{1}{2}(x-x_0)^2 f''(x_0) + o(x-x_0)^2$ par la formule de Taylor-Young à l'ordre 2. Il en résulte que

$$\frac{\sqrt{f(x)}-\sqrt{f(x_0)}}{x-x_0} = \frac{|x-x_0|}{x-x_0}\sqrt{\frac{f''(x_0)}{2}+o(1)}$$

(comme $f \geqslant 0$ on a nécessairement $f''(x_0) \geqslant 0$). Il en résulte que $\sqrt{f}$ est dérivable à droite et à gauche en x_0 avec une dérivée à droite qui vaut $\sqrt{\dfrac{f''(x_0)}{2}}$ et une dérivée à gauche qui vaut $-\sqrt{\dfrac{f''(x_0)}{2}}$. Ainsi, $\sqrt{f}$ est dérivable en x_0 si et seulement si $f''(x_0) = 0$. La dérivée en x_0 de $\sqrt{f}$ est alors nulle.

Conclusion. $\sqrt{f}$ est dérivable sur $\mathbb{R}$ si et seulement si en tout zéro x_0 de f on a $f''(x_0) = 0$.

2. Soit $x \in [-\alpha, \alpha]$. Pour $|h| \leqslant \alpha$, on a, d'après la formule de Taylor-Lagrange,

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2} f''(c) \geqslant 0,$$

avec c entre x et $x+h$. *A fortiori*, on a

$$\mathrm{P}(h) = \mathrm{M}(\alpha)\frac{h^2}{2} + hf'(x) + f(x) \geqslant 0,$$

car $|c| \leqslant 2\alpha$. Le polynôme P est positif sur l'intervalle $[-\alpha, \alpha]$. Or le minimum de ce polynôme est atteint en $\dfrac{-f'(x)}{\mathrm{M}(\alpha)}$ et comme il existe d entre 0 et x tel que $f'(x) = f'(0) + xf''(d) = xf''(d)$, on a

$$\left|\frac{f'(x)}{\mathrm{M}(\alpha)}\right| = \left|\frac{xf''(d)}{\mathrm{M}(\alpha)}\right| \leqslant \alpha.$$

Le minimum de P est atteint dans le segment $[-\alpha, \alpha]$; il est donc positif. Donc P est positif sur $\mathbb{R}$ et son discriminant est négatif : $f'(x)^2 \leqslant 2f(x)\mathrm{M}(\alpha)$. C'est l'inégalité demandée.

3. On va montrer que la condition obtenue dans la question 1 suffit pour que $\sqrt{f}$ soit de classe $\mathcal{C}^1$. Il suffit de prouver que la dérivée de $\sqrt{f}$ est continue en tout zéro x_0 de f. À translation près on peut se placer en $x_0 = 0$. On est alors ramené aux hypothèses de la question précédente : $f(0) = f'(0) = f''(0) = 0$. En reprenant les notations ci-dessus, on a pour tout point x de $[-\alpha, \alpha]$ tel que $f(x) \neq 0$,

$$|(\sqrt{f})'(x)| = \frac{|f'(x)|}{2\sqrt{f(x)}} \leqslant \frac{\sqrt{\mathrm{M}(\alpha)}}{\sqrt{2}}.$$

L'inégalité $|(\sqrt{f})'(x)| \leqslant \dfrac{\sqrt{\mathrm{M}(\alpha)}}{\sqrt{2}}$ demeure pour un point x tel que $f(x) = 0$ puisqu'en un tel point la dérivée de $\sqrt{f}$ est nulle. Comme f'' est continue, la quantité $\mathrm{M}(\alpha)$ tend vers 0 lorsque α tend vers 0. Cela prouve la continuité de $(\sqrt{f})'$ en 0.

Conclusion. $\sqrt{f}$ est de classe $\mathcal{C}^1$ si et seulement si en tout zéro x_0 de f on a $f''(x_0) = 0$. ◁

Ce résultat, prouvé par Glaeser, se généralise aux fonctions positives de classe $\mathcal{C}^2$ sur un ouvert de $\mathbb{R}^n$. Le lecteur pourra en rédiger la démonstration en adaptant la solution de l'exercice.

4.41. Étude du maximum de $f \longmapsto f(0) - 2f(\frac{1}{2}) + f(1)$

Soit $\mathcal{E}$ l'ensemble des fonctions $f : [0,1] \to \mathbb{R}$ continues, de classe $\mathcal{C}^2$ sur $]0,1[$ telles que, pour tout $x \in]0,1[$, $|f''(x)| \leqslant 1$. Pour $f \in \mathcal{E}$, on note $\mathrm{A}(f) = f(0) - 2f(\frac{1}{2}) + f(1)$.

1. Montrer que A est bornée sur $\mathcal{E}$ et déterminer $\sup_{f\in\mathcal{E}} \mathrm{A}(f) = \mathrm{M}$.

2. Résoudre $\mathrm{A}(f) = \mathrm{M}$ dans $\mathcal{E}$.

(École polytechnique)

▷ **Solution.**

1. Si on pose $g(t) = f\left(t + \frac{1}{2}\right) - f(t)$ pour t dans $\left[0, \frac{1}{2}\right]$, on a $\mathrm{A}(f) = g\left(\frac{1}{2}\right) - g(0)$. En utilisant la formule des accroissements finis, il existe donc $c \in \left]0, \frac{1}{2}\right[$ tel que $\mathrm{A}(f) = \frac{1}{2}g'(c)$. On a de plus $g'(c) = f'\left(c+\frac{1}{2}\right) - f'(c)$. En utilisant une nouvelle fois les accroissements finis, on obtient l'existence de $c' \in \left]0, \frac{1}{2}\right[$ tel que $\mathrm{A}(f) = \frac{1}{4}f''(c+c')$. On en déduit donc que $|\mathrm{A}(f)| \leqslant \frac{1}{4}$ pour toute fonction f de $\mathcal{E}$.

La valeur $\frac{1}{4}$ est atteinte. En effet, si on prend pour f une fonction telle que $f'' = 1$, c'est-à-dire de la forme $t \longmapsto \frac{1}{2}t^2 + at + b$, on obtient exactement $\mathrm{A}(f) = \frac{1}{4}$.

Conclusion. On a $\sup_{f\in\mathcal{E}} \mathrm{A}(f) = \max_{f\in\mathcal{E}} \mathrm{A}(f) = \frac{1}{4}$.

2. On va prouver que les seules fonctions $f \in \mathcal{E}$ telle que $\mathrm{A}(f) = \frac{1}{4}$ sont les fonctions polynômes $t \longmapsto \frac{1}{2}t^2 + at + b$. Soit $f \in \mathcal{E}$ telle que $\mathrm{A}(f) = \frac{1}{4}$. On pose toujours $g(t) = f\left(t + \frac{1}{2}\right) - f(t)$ pour $t \in \left[0, \frac{1}{2}\right]$. La fonction g est de classe $\mathcal{C}^2$ sur l'intervalle $\left]0, \frac{1}{2}\right[$. En fait, comme $|f''| \leqslant 1$, la fonction f' admet des limites finies en 0 et en 1 (f'' est sommable sur $]0,1[$ donc $\int_x^{1/2} f'' = f'(1/2) - f'(x)$ admet une limite quand x tend vers 0^+ et de même pour $\int_{1/2}^x f'' = f'(x) - f'(1/2)$ quand x tend vers 1). Le théorème sur la limite de la dérivée[2] nous assure de la dérivabilité de f en 0 et 1 et même du caractère $\mathcal{C}^1$ de la fonction f

2. Rappelons que si $f : [a,b] \to \mathbb{R}$ est une fonction continue sur $[a,b]$ et dérivable sur $]a,b[$ et telle que f' admette une limite finie ℓ en b, alors f est dérivable en b et $f'(b) = \ell$: c'est une conséquence immédiate du théorème des accroissements finis.

sur $[0,1]$. Ainsi, g est donc de classe $\mathcal{C}^1$ sur $\left[0,\frac{1}{2}\right]$. On va alors utiliser des expressions intégrales. On a, pour tout $t \in \left]0,\frac{1}{2}\right[$,

$$g'(t) = f'\left(t+\frac{1}{2}\right) - f'(t) = \int_t^{t+\frac{1}{2}} f''(u)\mathrm{d}u \leqslant \frac{1}{2}.$$

On sait par ailleurs que $\mathrm{A}(f) = g\left(\frac{1}{2}\right) - g(0) = \int_0^{1/2} g'(t)\mathrm{d}t = \frac{1}{4}$. La dernière égalité impose que $g'(t) = \frac{1}{2}$ pour tout $t \in \left[0,\frac{1}{2}\right]$ et cela donne que $f''(u) = 1$ pour tout $u \in]0,1[$. Il en résulte l'existence de a et b réels tels que, si $t \in]0,1[$, $f(t) = \frac{1}{2}t^2 + at + b$. Cette formule reste vraie en 0 et 1 puisque f est continue.

Conclusion. Les solutions de $\mathrm{A}(f) = \frac{1}{4}$ sont donc les fonctions du type $t \in [0,1] \longmapsto \frac{t^2}{2} + at + b$ avec $(a,b) \in \mathbb{R}^2$ quelconque. ◁

4.42. Puissances entières

Soit $\mathrm{E} = \mathcal{C}^\infty(\mathbb{R}_+^*, \mathbb{R})$. Pour $f \in \mathrm{E}$ on note $\Delta f \in \mathrm{E}$ l'application définie pour tout $x > 0$ par $\Delta f(x) = f(x+1) - f(x)$.

1. Soit $f \in \mathrm{E}$. Montrer que pour tout $x > 0$ et tout $n \in \mathbb{N}^*$, il existe $\theta \in]0,1[$, tel que $\Delta^n f(x) = f^{(n)}(x + n\theta)$.

2. En déduire l'ensemble des réels c tels que n^c soit entier pour tout $n \geqslant 1$.

(École normale supérieure)

▷ **Solution.**

1. Soit $x > 0$. On procède par récurrence sur n, le cas $n = 1$ étant une application directe de la formule des accroissements finis. Supposons le résultat vrai au rang n. On applique l'hypothèse de récurrence à la fonction $g = \Delta f$. Soit $x > 0$. Il existe $\theta \in]0,1[$ tel que $(\Delta^n g)(x) = g^{(n)}(x + n\theta)$. Cela s'écrit $\Delta^{n+1} f(x) = f^{(n)}(x + n\theta + 1) - f^{(n)}(x + n\theta)$. On applique alors l'égalité des accroissements finis

$$f^{(n)}(x + n\theta + 1) - f^{(n)}(x + n\theta) = f^{(n+1)}(x + n\theta + \theta'),$$

où $\theta' \in]0,1[$, et on conclut en remarquant que $0 < n\theta + \theta' < n + 1$ donc qu'il existe $\theta'' \in]0,1[$ tel que $n\theta + \theta' = (n+1)\theta''$.

Cela ne se passe pas aussi bien si on écrit $\Delta^{n+1} f = \Delta(\Delta^n f)$ (ci-dessus on utilise la relation $\Delta^{n+1} f = \Delta^n(\Delta f)$). En effet, pour $x > 0$ on trouve $\theta \in]0,1[$ tel que $\Delta^n f(x) = f^{(n)}(x + n\theta)$. Mais par malheur ce θ dépend de x. Donc $\Delta(\Delta^n f)(x) = f^{(n)}(x + 1 + n\theta') - f^{(n)}(x + n\theta)$. Et ici, c'est ennuyeux car l'écart entre les deux points n'est pas de 1.

2. Notons Ω l'ensemble des réels c tels que, pour tout $n \geqslant 1$, $n^c \in \mathbb{N}$. Il est clair que $\mathbb{N} \subset \Omega$. On va prouver par l'absurde qu'il y a égalité en supposant qu'il existe $\alpha \in \Omega$, $\alpha \notin \mathbb{N}$. La fonction $f(x) = x^\alpha$ est dans E et ne prend que des valeurs entières sur $\mathbb{N}^*$. Il en est donc de même de $\Delta^p f$ pour tout $p \in \mathbb{N}$. Appliquons le résultat de la question précédente à f avec $n = \mathrm{E}(\alpha) + 1$. Comme $f^{(n)}(x) = \alpha(\alpha-1)\dots(\alpha-n+1)x^{\alpha-n}$, on a pour tout $x \in \mathbb{N}^*$,

$$\Delta^n f(x) = \alpha(\alpha-1)\dots(\alpha-n+1)(x+n\theta)^{\alpha-n},$$

avec $\theta \in]0,1[$.

Pour tout entier x cette quantité est non nulle (car $\alpha \notin \mathbb{N}$) et entière d'après la remarque qui précède. De plus elle tend vers 0 lorsque x tend vers l'infini (puisque $\alpha - n < 0$). C'est absurde, car une suite d'entiers non nuls ne peut converger vers 0. ◁

L'énoncé suivant est un thème très classique des sujets de concours.

4.43. Inégalités de Kolmogorov

1. Soit $f \in \mathcal{C}^2(\mathbb{R},\mathbb{C})$. On suppose que f et f'' sont bornées sur $\mathbb{R}$ et on pose $\mathrm{M}_0 = \sup_{x\in\mathbb{R}} |f(x)|$, $\mathrm{M}_2 = \sup_{x\in\mathbb{R}} |f''(x)|$. Montrer que

$$\mathrm{M}_1 = \sup_{x\in\mathbb{R}} |f'(x)| \leqslant \sqrt{2\mathrm{M}_0\mathrm{M}_2}.$$

2. Soit $f \in \mathcal{C}^n(\mathbb{R},\mathbb{C})$. Pour tout $k \in [\![0,n]\!]$, on pose $\mathrm{M}_k = \sup_{x\in\mathbb{R}} |f^{(k)}(x)|$. On suppose que M_0 et M_n sont finis. Montrer que pour tout $k \in [\![0,n]\!]$, M_k est fini et

$$\mathrm{M}_k \leqslant 2^{\frac{k(n-k)}{2}} \mathrm{M}_0^{1-\frac{k}{n}} \mathrm{M}_n^{\frac{k}{n}}.$$

(ENS Ulm)

▷ **Solution.**

1. Une idée naturelle consiste à utiliser l'inégalité de Taylor-Lagrange. Soit x un réel et $h > 0$. On peut écrire

$$|f(x+h) - f(x) - hf'(x)| \leqslant \frac{h^2\mathrm{M}_2}{2}.$$

Grâce à l'inégalité triangulaire, on en déduit que $h|f'(x)| \leqslant 2\mathrm{M}_0 + \dfrac{h^2\mathrm{M}_2}{2}$ ou encore $|f'(x)| \leqslant \dfrac{2\mathrm{M}_0}{h} + \dfrac{h\mathrm{M}_2}{2}$. Cela étant valable pour tout $h > 0$, on va choisir h de manière à minimiser le majorant. On peut étudier la fonction $h \longmapsto \dfrac{2\mathrm{M}_0}{h} + \dfrac{h\mathrm{M}_2}{2}$ ou plus simplement invoquer l'inégalité entre la moyenne géométrique et la moyenne arithmétique : le produit de $\dfrac{2\mathrm{M}_0}{h}$ et $\dfrac{h\mathrm{M}_2}{2}$ est constant et vaut $\mathrm{M}_0\mathrm{M}_2$. La somme sera minimale lorsque les deux termes sont égaux et valent donc chacun $\sqrt{\mathrm{M}_0\mathrm{M}_2}$. On en déduit que

$$|f'(x)| \leqslant 2\sqrt{\mathrm{M}_0\mathrm{M}_2},$$

ce qui est plus faible que l'inégalité demandée par l'énoncé. Pour aboutir à la constante de l'énoncé, on raffine le travail précédent en écrivant aussi l'inégalité de Taylor-Lagrange entre x et $x - h$:

$$|f(x-h) - f(x) + hf'(x)| \leqslant \frac{h^2\mathrm{M}_2}{2}.$$

En la combinant avec l'inégalité utilisée précédemment, on obtient, toujours grâce à l'inégalité triangulaire,

$$|f(x+h) - f(x-h) - 2hf'(x)| \leqslant h^2\mathrm{M}_2$$

puis

$$2h|f'(x)| \leqslant h^2\mathrm{M}_2 + |f(x+h) - f(x-h)| \leqslant h^2\mathrm{M}_2 + 2\mathrm{M}_0.$$

Ainsi, pour tout $h > 0$, $|f'(x)| \leqslant \dfrac{h\mathrm{M}_2}{2} + \dfrac{\mathrm{M}_0}{h}$. En prenant la valeur de h qui minimise le majorant (à savoir $h = \sqrt{\dfrac{\mathrm{M}_2}{2\mathrm{M}_0}}$), on obtient $|f'(x)| \leqslant \sqrt{2\mathrm{M}_0\mathrm{M}_2}$. Comme cela vaut pour tout réel x, f' est bornée et $\mathrm{M}_1 \leqslant \sqrt{2\mathrm{M}_0\mathrm{M}_2}$. Cela reste bien entendu vrai si $\mathrm{M}_0 = 0$, puisqu'alors f est nulle.

2. Montrons pour commencer que les M_k sont finis. Soit x un réel et $h > 0$. L'inégalité de Taylor-Lagrange entre x et $x + h$ conduit à

$$\left| f(x+h) - f(x) - hf'(x) - \cdots - \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x) \right| \leqslant \frac{h^n \mathrm{M}_n}{n!}$$

soit, par inégalité triangulaire, à

$$\left| hf'(x) + \cdots + \frac{h^{n-1}}{(n-1)!} f^{(n-1)}(x) \right| \leqslant 2\mathrm{M}_0 + \frac{h^n \mathrm{M}_n}{n!}$$

Écrivons cette inégalité pour $n - 1$ valeurs distinctes de h, $0 < h_1 < h_2 < \cdots < h_{n-1}$. Si on note X le vecteur colonne $(f'(x), \ldots, f^{(n-1)}(x))$,

les inégalités précédentes signifient que $\|AX\|_\infty \leqslant K = 2M_0 + \frac{h_{n-1}^n M_n}{n!}$, où A est la matrice

$$A = \begin{pmatrix} h_1 & \frac{h_1^2}{2!} & \cdots & \frac{h_1^{n-1}}{(n-1)!} \\ \vdots & \vdots & & \vdots \\ h_{n-1} & \frac{h_{n-1}^2}{2!} & \cdots & \frac{h_{n-1}^{n-1}}{(n-1)!} \end{pmatrix}.$$

Or A est clairement inversible (en sortant les factorielles et en mettant en facteur les h_k, on se ramène à une matrice de Vandermonde). On a donc $\|X\|_\infty \leqslant \|A^{-1}\|K$. Comme cette inégalité vaut pour tout réel x, les quantités M_k sont bien finies.

On va maintenant prouver les inégalités proposées par récurrence sur n. On observe que pour $k = 0$ et $k = n$ l'inégalité est triviale.

• Pour $n = 2$ et $k = 1$, on obtient $M_1 \leqslant \sqrt{2M_0M_2}$, ce qui a été prouvé dans la première question.

• Supposons que le résultat est vrai jusqu'au rang n et considérons une fonction f de classe $\mathcal{C}^{n+1}$. Fixons k entre 1 et n. En appliquant le cas $n = 2$ à la fonction $f^{(k-1)}$ (ce qui est possible d'après ce qu'on vient de voir) on a l'inégalité

$$M_k^2 \leqslant 2M_{k-1}M_{k+1}.$$

On va maintenant majorer M_{k-1} et M_{k+1} en utilisant l'hypothèse de récurrence. Pour le premier, on utilise l'hypothèse de récurrence au rang k. Il vient

$$M_{k-1} \leqslant 2^{\frac{k-1}{2}} M_0^{\frac{1}{k}} M_k^{\frac{k-1}{k}}.$$

Pour le second terme, on utilise l'hypothèse de récurrence au rang $n+1-k$ pour la fonction $f^{(k)}$. Il vient

$$M_{k+1} \leqslant 2^{\frac{n-k}{2}} M_k^{\frac{n-k}{n+1-k}} M_{n+1}^{\frac{1}{n+1-k}}.$$

En utilisant ces deux majorations, on aboutit à

$$M_k^2 \leqslant 2^{\frac{n+1}{2}} M_0^{\frac{1}{k}} M_k^{\frac{k-1}{k}+\frac{n-k}{n+1-k}} M_{n+1}^{\frac{1}{n+1-k}}.$$

En élevant cela à la puissance $\frac{k(n+1-k)}{n+1}$, on obtient le résultat souhaité

$$M_k \leqslant 2^{\frac{k(n+1-k)}{2}} M_0^{1-\frac{k}{n+1}} M_{n+1}^{\frac{k}{n+1}},$$

ce qui termine la récurrence. ◁

Les exercices suivants concernent les fonctions de classe $\mathcal{C}^\infty$.

4.44. Théorème de division

Soit $f : \mathbb{R} \to \mathbb{R}$ de classe $\mathcal{C}^\infty$ telle que $f(0) = 0$.

1. Montrer que l'application $g : x \longmapsto \dfrac{f(x)}{x}$ se prolonge en une application de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$.

2. On suppose de plus que $f''(0) \neq 0$ et $f(x) > 0$ pour $x \neq 0$. Montrer qu'il existe $g \in \mathcal{C}^\infty(\mathbb{R}, \mathbb{R})$ telle que $g^2 = f$.

(École polytechnique)

▷ **Solution.**

1. On va donner deux preuves différentes de ce résultat.

• La fonction g est continue sur $\mathbb{R}^*$ et tend vers $f'(0)$ en 0. On pose donc $g(0) = f'(0)$ de sorte que g est continue sur $\mathbb{R}$. La fonction g est bien entendu de classe C^∞ sur $\mathbb{R}^*$. On va montrer qu'elle est aussi dérivable à tout ordre en 0.

Pour $x \neq 0$, on a

$$g'(x) = \frac{xf'(x) - f(x)}{x^2} = \frac{xf'(0) + x^2 f''(0) - xf'(0) - \frac{1}{2}x^2 f''(0) + o(x^2)}{x^2},$$

ce qui tend vers $\frac{1}{2}f''(0)$. Donc, par le théorème sur la limite de la dérivée, g est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ avec $g'(0) = \frac{1}{2}f''(0)$. Montrons alors par récurrence sur $n \geqslant 1$ que g est de classe $\mathcal{C}^n$ sur $\mathbb{R}$ avec $g^{(n)}(0) = \dfrac{f^{(n+1)}(0)}{n+1}$. Le cas $n = 1$ vient d'être traité. Supposons le résultat établi au rang $n-1$. $n \geqslant 2$. Pour simplifier on va se ramener au cas où les $n+1$ premières dérivées en 0 de f sont nulles. Considérons pour cela la fonction

$$\tilde{f}(x) = f(x) - xf'(0) - \frac{x^2}{2}f''(0) - \cdots - \frac{x^{n+1}}{(n+1)!}f^{(n+1)}(0).$$

Cette fonction est de classe $\mathcal{C}^\infty$ et ses $n+1$ premières dérivées sont nulles en 0. Par hypothèse de récurrence, $\tilde{g}(x) = \dfrac{\tilde{f}(x)}{x}$ est de classe $\mathcal{C}^{n-1}$ avec $\tilde{g}(0) = \tilde{g}'(0) = \cdots = \tilde{g}^{(n-1)}(0) = 0$. En dérivant la relation $\tilde{g}(x) = \dfrac{\tilde{f}(x)}{x}$ à l'aide de la formule de Leibniz, il vient, pour tout x non nul :

$$\tilde{g}^{(n)}(x) = \frac{\tilde{f}^{(n)}(x)}{x} + \sum_{k=0}^{n-1} \mathrm{C}_n^k (-1)^{n-k}(n-k)!\frac{\tilde{f}^{(k)}(x)}{x^{n-k+1}}.$$

On effectue un $\mathrm{DL}_{n-k+1}(0)$ de $\tilde{f}^{(k)}(x)$. Comme les dérivées de $\tilde{f}$ en 0 jusqu'à l'ordre $n+1$ sont nulles, on a $\tilde{f}^{(k)}(x) = o(x^{n-k+1})$. Il en résulte

que $\tilde{g}^{(n)}(x) = o(1)$ tend vers 0 en 0 et le théorème sur la limite de la dérivée permet de conclure que $\tilde{g}$ est de classe $\mathcal{C}^n$ sur $\mathbb{R}$ avec $\tilde{g}^{(n)}(0) = 0$. La relation

$$g(x) = \frac{f(x)}{x} = \tilde{g}(x) + f'(0) + \frac{x}{2}f''(0) + \cdots + \frac{x^n}{(n+1)!}f^{(n+1)}(0)$$

montre que g est aussi de classe $\mathcal{C}^n$ sur $\mathbb{R}$ et que $g^{(n)}(0) = \dfrac{f^{(n+1)}(0)}{n+1}$. Cela termine la récurrence.

Conclusion. La fonction g se prolonge en une fonction de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ avec pour tout $n \geqslant 0$, $g^{(n)}(0) = \dfrac{f^{(n+1)}(0)}{n+1}$.

Le résultat que l'on vient de démontrer est évident si la fonction f est développable en série entière en 0.

• Voici une seconde solution, nettement plus rapide, qui utilise le théorème de dérivation sous le signe intégral. La fonction f étant de classe $\mathcal{C}^1$, on peut écrire pour x réel

$$f(x) = f(0) + \int_0^x f'(t)\mathrm{d}t = x\int_0^1 f'(ux)\mathrm{d}u$$

en faisant le changement de variables $t = ux$. Ainsi, pour x non nul,

$$g(x) = \int_0^1 f'(ux)\mathrm{d}u.$$

Si on pose $g(0) = f'(0)$, on voit que g est la fonction $x \longmapsto \int_0^1 f'(ux)\mathrm{d}u$ qui est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ puisque f' l'est.

2. Notons qu'au voisinage de 0, on a $f(x) = xf'(0) + \dfrac{x^2}{2}f''(0) + o(x^2)$. Pour que f garde un signe constant, il faut $f'(0) = 0$. Posons alors $h(x) = \dfrac{f(x)}{x^2}$ pour x non nul. La fonction h se prolonge par continuité en 0 par $h(0) = \dfrac{f''(0)}{2}$. Le théorème de division de la question précédente, appliqué deux fois, montre que h est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$. Par ailleurs, h est strictement positive sur $\mathbb{R}$, donc la fonction $\sqrt{h}$ est également de classe $\mathcal{C}^\infty$. Il nous suffit alors de poser $g(x) = x\sqrt{h(x)}$ pour obtenir une fonction répondant au problème posé. ◁

On sait qu'une fonction développable en série entière sur $\mathbb{R}$ est parfaitement déterminée par la suite de ses dérivées en 0. *Il n'en est pas de même pour une fonction de classe $\mathcal{C}^\infty$.*

4.45. Le théorème de réalisation de Borel

Soit E l'ensemble des fonctions réelles de classe $\mathcal{C}^\infty$ à support compact.

1. Soit $p \in \mathbb{N}$. On pose $\mathrm{N}_p(f) = \max\limits_{0 \leqslant q \leqslant p} \|f^{(q)}\|_\infty$. Montrer que N_p est une norme sur E.

2. Soit φ une fonction $\mathcal{C}^\infty$, nulle en dehors de $[-1,1]$, et telle que $\varphi(0) = 1$ et $\varphi^{(p)}(0) = 0$ si $p \geqslant 1$. Comment construire une telle fonction ?

3. Montrer que, pour tout $n \in \mathbb{N}$, $\lim\limits_{\varepsilon \to 0^+} \mathrm{N}_n\left(\varphi\left(\frac{x}{\varepsilon}\right) x^{n+1}\right) = 0$.

4. On se donne une suite (a_n) de réels. Construire une suite décroissante de réels strictement positifs (ε_n), telle que

$$\mathrm{N}_n\left(\varphi\left(\frac{x}{\varepsilon_n}\right) x^{n+1}\right) \leqslant \min\left(1, \frac{1}{|a_{n+1}|}\right).$$

On pose $g(x) = a_0 + \sum\limits_{n=1}^{+\infty} \varphi\left(\frac{x}{\varepsilon_{n-1}}\right) \frac{a_n x^n}{n!}$. Montrer que g est de classe $\mathcal{C}^\infty$, et vérifie $g^{(n)}(0) = a_n$ pour tout $n \in \mathbb{N}$.

(École polytechnique)

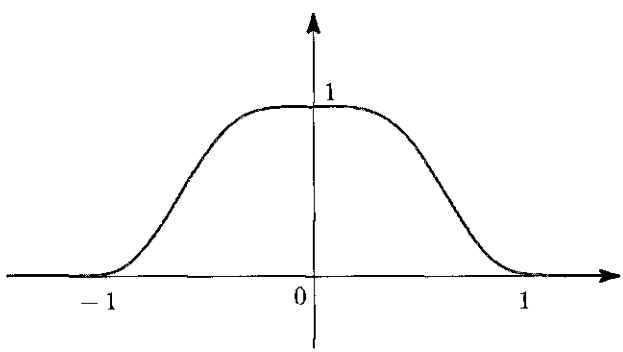
▷ **Solution.**

1. Remarquons que si $f \in \mathrm{E}$, f et toutes ses dérivées sont continues et à support compact, donc bornées, ce qui assure l'existence de $\mathrm{N}_p(f)$. La vérification que N_p est une norme est aisée. En utilisant la linéarité de la dérivation et l'inégalité triangulaire, on montre que si $(f,g) \in \mathrm{E}^2$ et $\lambda \in \mathbb{R}$, alors

$$\mathrm{N}_p(f+g) \leqslant \mathrm{N}_p(f) + \mathrm{N}_p(g) \text{ et } \mathrm{N}_p(\lambda f) = |\lambda| \mathrm{N}_p(f).$$

De plus $\mathrm{N}_p(f) = 0$ implique $\|f\|_\infty = 0$ et donc $f = 0$.

2. On va construire une fonction de classe $\mathcal{C}^\infty$ dont le graphe ressemble à ce qui suit :



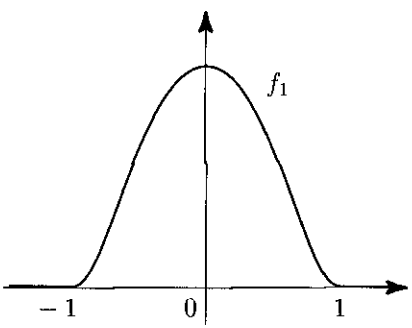
Pour simplifier on va la choisir paire. La partie difficile de la construction est bien entendu d'avoir les dérivées d'ordre $\geqslant 1$ nulle en 0, et les toutes les dérivées à gauche en 1 également nulles afin d'avoir un raccord de classe $\mathcal{C}^\infty$.

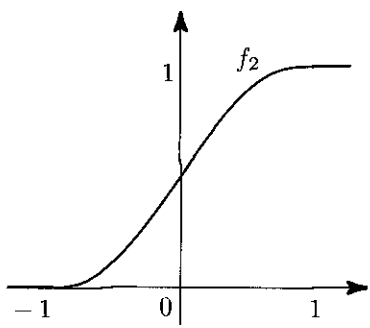
On va partir d'une fonction nulle sur $\mathbb{R}_-$ et strictement positive sur $\mathbb{R}_+^*$ qui est de classe $\mathcal{C}^\infty$. On peut vérifier que la fonction f définie par

$$\begin{cases} f(x) = 0 \text{ si } x \leqslant 0 \\ f(x) = e^{-1/x} \text{ si } x > 0 \end{cases}$$

convient. En effet, f est $\mathcal{C}^\infty$ sur $]-\infty, 0[$ et $]0, +\infty[$ et il suffit d'étudier les dérivées en 0. Pour cela, on montre par récurrence que, pour tout $n \in \mathbb{N}$, il existe un polynôme $\mathrm{P}_n \in \mathbb{R}[\mathrm{X}]$ tel que $f^{(n)}(x) = \dfrac{\mathrm{P}_n(x)}{x^{2n}} e^{-1/x}$, pour tout $x > 0$. Cela établi, on montre par récurrence sur n que f est de classe $\mathcal{C}^n$ sur $\mathbb{R}$, pour tout n. On a $\lim\limits_{x \to 0^+} e^{-1/x} = 0$, donc f est continue. Supposons que f est de classe $\mathcal{C}^n$. Alors, par croissance comparée, $\lim\limits_{0^+} f^{(n+1)} = \lim\limits_{0^-} f^{(n+1)} = 0$. Du théorème sur la limite de la dérivée, on déduit que $f^{(n)}$ est dérivable en 0 et que $f^{(n+1)}$ est continue en 0. La fonction f est donc $\mathcal{C}^{n+1}$ sur $\mathbb{R}$ et cela termine la récurrence.

On considère alors la fonction f_1 définie par $f_1(x) = f(1+x)f(1-x)$. Elle est de classe $\mathcal{C}^\infty$, paire, nulle en dehors du segment $[-1, 1]$ et, par la formule de Leibniz, toutes ses dérivées en -1 et 1 sont nulles. Elle ne répond pas encore à la question posée car ses dérivées en 0 ne conviennent pas. Considérons alors f_2 définie par





$$f_2(x) = \frac{\displaystyle\int_{-1}^{x} f_1(t)\mathrm{d}t}{\displaystyle\int_{-1}^{1} f_1(t)\mathrm{d}t}.$$

La fonction f_2 est toujours de classe $\mathcal{C}^\infty$ et à support dans $[-1, 1]$. On a $f_2^{(n)}(-1) = 0$ pour tout n, $f_2(1) = 1$ et $f_2^{(n)}(1) = 0$ pour tout $n \geqslant 1$. Le

graphe de f_2 sur $[-1,1]$ est exactement ce qu'on souhaite obtenir pour φ sur $[-1,0]$.

On termine donc en posant $\varphi(x) = \varphi(-x) = f_2(2x+1)$ pour tout $x \leqslant 0$. La fonction ainsi construite répond à la question posée.

On dit que φ est une fonction plateau.

3. Soit $\varepsilon > 0$, $n \in \mathbb{N}$ et $\varphi_{\varepsilon,n} : x \longmapsto \varphi\left(\frac{x}{\varepsilon}\right) x^{n+1}$. La fonction $\varphi_{\varepsilon,n}$ est $\mathcal{C}^\infty$ et nulle en dehors de $[-\varepsilon,\varepsilon]$. Soit $p \in [\![0,n]\!]$. La formule de Leibniz donne, pour tout $x \in [-\varepsilon,\varepsilon]$,

$$\varphi_{\varepsilon,n}^{(p)}(x) = \sum_{k=0}^{p} \mathrm{C}_p^k (n+1)n\ldots(n+2-k)x^{n+1-k}\varphi^{(p-k)}\left(\frac{x}{\varepsilon}\right)\varepsilon^{-(p-k)}.$$

On en déduit que $\|\varphi_{\varepsilon,n}^{(p)}\|_\infty \leqslant \mathrm{N}_n(\varphi)\varepsilon^{n+1-p}\sum_{k=0}^{p}\mathrm{C}_p^k\frac{(n+1)!}{(n+1-k)!}$. Ceci montre que, pour $0 \leqslant p \leqslant n$, $\lim\limits_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \|\varphi_{\varepsilon,n}^{(p)}\|_\infty = 0$ et donc que

$$\lim_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \mathrm{N}_n(\varphi_{\varepsilon,n}) = 0.$$

4. • On construit la suite (ε_n) par récurrence. On a $\lim\limits_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \mathrm{N}_0(\varphi_{\varepsilon,0}) = 0$. Il existe donc $\varepsilon_0 > 0$ tel que $\mathrm{N}_0(\varphi_{\varepsilon_0,0}) \leqslant \min\left(1, \frac{1}{|a_1|}\right)$. Les réels $\varepsilon_0,\ldots,\varepsilon_{n-1}$ étant choisis, de $\lim\limits_{\substack{\varepsilon\to 0\\ \varepsilon>0}} \mathrm{N}_n(\varphi_{\varepsilon,n}) = 0$ on déduit l'existence de $\varepsilon_n > 0$ tel que $\varepsilon_n < \varepsilon_{n-1}$ et $\mathrm{N}_n(\varphi_{\varepsilon_n,n}) \leqslant \min\left(1, \frac{1}{|a_{n+1}|}\right)$.

• Posons, pour tout réel x, $g_0(x) = a_0$ et $g_n(x) = \varphi_{\varepsilon_{n-1},n-1}(x)\frac{a_n}{n!}$ si $n \geqslant 1$. On a, par le choix de la suite (ε_n), $\mathrm{N}_{n-1}(g_n) \leqslant \frac{1}{n!}$. On en déduit que, pour tout $p \in \mathbb{N}$, on a, si $n > p$,

$$\|g_n^{(p)}\|_\infty \leqslant \frac{1}{n!}.$$

Il en résulte que la série $\sum g_n^{(p)}$ (la sommation portant sur n) est normalement convergente, pour tout p. Le théorème de dérivation des séries de fonctions permet de conclure que g est $\mathcal{C}^\infty$ et que, pour tout $p \in \mathbb{N}$, $g^{(p)} = \sum\limits_{n=0}^{+\infty} g_n^{(p)}$.

• Calculons $g^{(p)}(0)$, pour tout p, par la formule de Leibniz. Si $p < n$, alors on peut mettre en facteur x^{n-p} dans l'expression de $g_n^{(p)}(x)$; on a donc $g_n^{(p)}(0) = 0$. Si $p > n$, $g_n^{(p)}(0)$ est une combinaison linéaire de $\varphi^{(k)}(0)$ avec $k \geqslant p-n \geqslant 1$. On a encore $g_n^{(p)}(0) = 0$. Enfin $g_n^{(n)}(0) = \varphi(0)a_n = a_n$.

On en déduit que, pour tout $p \in \mathbb{N}$, $g^{(p)}(0) = \sum_{n=0}^{+\infty} g_n^{(p)}(0) = a_p$.

Conclusion. Nous avons démontré le théorème de Borel : pour toute suite de réels (a_n), il existe une fonction g de classe $\mathcal{C}^\infty$ telle que, pour tout $n \in \mathbb{N}$, $g^{(n)}(0) = a_n$. $\triangleleft$

L'exercice suivant s'intéresse à l'ensemble des zéros d'une fonction de classe $\mathcal{C}^\infty$.

4.46. Théorème de Whitney

Montrer que tout fermé de $\mathbb{R}$ est l'ensemble des zéros d'une fonction $\mathcal{C}^\infty$ de $\mathbb{R}$ dans $\mathbb{R}$.

(École polytechnique)

$\triangleright$ **Solution.**

Soit F un fermé non vide de $\mathbb{R}$. On suppose que $F \neq \mathbb{R}$, sinon la fonction nulle convient. D'après l'exercice 1.27, l'ouvert non vide $\mathbb{R} \setminus F$ est une réunion au plus dénombrable d'intervalles ouverts non vides disjoints.

Pour tout intervalle ouvert non vide I de $\mathbb{R}$, on construit une fonction $f_I : \mathbb{R} \longmapsto [0,1]$ de classe $\mathcal{C}^\infty$ positive ou nulle telle que, pour tout réel x, $f_I(x) = 0$ soit équivalent à $x \notin I$. On sait (voir la solution de l'exercice 4.45) que la fonction $f : x \longmapsto \begin{cases} 0 & \text{si } x \leqslant 0 \\ e^{-\frac{1}{x}} & \text{si } x > 0 \end{cases}$ est de classe $\mathcal{C}^\infty$. Elle est à valeurs dans $[0,1]$.

Si $I =]a,b[$, avec a et b réels la fonction $f_I : x \longmapsto f(-(x-a)(x-b))$ convient : elle est $\mathcal{C}^\infty$ comme composée de fonctions $\mathcal{C}^\infty$ et $f_I(x) = 0$ équivaut à $(x-a)(x-b) \geqslant 0$, *i.e.* à $x \notin]a,b[$. De même si $I =]a,+\infty[$ avec a réel, on prend $f_I : x \longmapsto f(x-a)$ et si $I =]-\infty,b[$ avec b réel, on prend $f_I : x \longmapsto f(b-x)$.

Si $\mathbb{R} \setminus F$ est une réunion d'un nombre fini d'intervalles ouverts on écrit $\mathbb{R} \setminus F = \bigcup_{i=1}^{n} I_k$. La fonction $\varphi = \sum_{k=1}^{n} f_{I_k}$ est $\mathcal{C}^\infty$. Comme chacune des fonctions est positive, on a pour tout réel x,

$$\varphi(x) = 0 \Longleftrightarrow \forall k \in [\![1,n]\!],\ f_{I_k}(x) = 0 \Longleftrightarrow \forall k \in [\![1,n]\!],\ x \notin I_k \Longleftrightarrow x \in F.$$

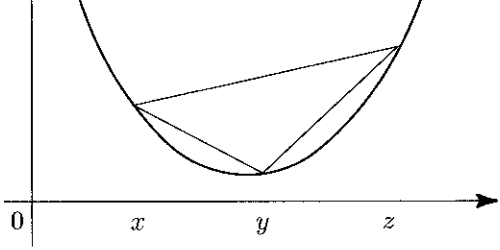
Si $\mathbb{R} \setminus F$ est une réunion d'un nombre dénombrable d'intervalles ouverts, $I = \bigcup_{n \in \mathbb{N}^*} I_n$, c'est un peu plus compliqué. On pose $\varphi = \sum_{n=1}^{+\infty} a_n f_{I_n}$, où les a_n seront strictement positifs et choisis de telle façon que φ soit

définie sur $\mathbb{R}$ et $\mathcal{C}^\infty$. Chaque fonction f_{I_n} est de classe $\mathcal{C}^\infty$ et pour tout $k \in \mathbb{N}$, $f_{\mathrm{I}_n}^{(k)}$ est bornée. En effet si I_n est borné, $f_{\mathrm{I}_n}^{(k)}$ est à support compact et sinon elle tend vers 0 en $\pm\infty$ (cela résulte de la croissance comparée d'un polynôme et d'une exponentielle). Pour tout $n \in \mathbb{N}^*$, posons $a_n = \dfrac{1}{n^2 \max(\|f_{\mathrm{I}_n}\|_\infty, \|f'_{\mathrm{I}_n}\|_\infty, \ldots, \|f_{\mathrm{I}_n}^{(n)}\|_\infty)}$. Pour tout $k \in \mathbb{N}$, on a, pour $n \geqslant k$, $a_n \leqslant \dfrac{1}{n^2 \|f_{\mathrm{I}_n}^{(k)}\|_\infty}$ et donc $a_n \|f_{\mathrm{I}_n}^{(k)}\|_\infty \leqslant \dfrac{1}{n^2}$. La série $\sum a_n f_{\mathrm{I}_n}^{(k)}$ est donc normalement convergente pour tout $k \in \mathbb{N}$. En appliquant de manière réitérée le théorème de dérivation des séries de fonctions, on prouve que φ est de classe $\mathcal{C}^\infty$. On montre comme dans le cas fini que l'ensemble des zéros de φ est F. $\lhd$

La série d'exercices qui suit porte sur la convexité. Rappelons qu'une fonction f d'un intervalle I *dans* $\mathbb{R}$ *est dite convexe lorsque son graphe est en dessous de toutes ses cordes, c'est-à-dire si*

$$\forall (x, y) \in \mathrm{I}^2, \quad \forall t \in [0, 1], \quad f((1-t)x + ty) \leqslant (1-t)f(x) + tf(y).$$

Si l'inégalité est stricte chaque fois que $x \neq y$ et $0 < t < 1$, la fonction est dite strictement convexe.



On déduit facilement de cette définition le théorème des pentes croissantes : f est convexe sur I *si et seulement si pour tout triplet (x, y, z) de points de* I *tel que $x < y < z$ on a*

$$\frac{f(y) - f(x)}{y - x} \leqslant \frac{f(z) - f(x)}{z - x} \leqslant \frac{f(z) - f(y)}{z - y}.$$

À partir de cela on montre facilement qu'une fonction convexe $f : \mathrm{I} \to \mathbb{R}$ admet en tout point intérieur à I *une dérivée à droite et une dérivée à gauche : en effet, si a est un point intérieur, la fonction qui à x associe la pente de la sécante entre a et x est croissante et bornée sur un voisinage*

pointé de x ; cette fonction admet donc des limites finies à droite et à gauche en x.

En particulier f est continue sur l'intérieur de I.

Lorsque f est dérivable sur I, *elle est convexe si et seulement si f' est croissante. Dans le cas où f est deux fois dérivable, elle est donc convexe si et seulement si $f'' \geqslant 0$.*

4.47. Convexité

Soit I un intervalle de $\mathbb{R}_+^*$ et $f : \mathrm{I} \to \mathbb{R}$. Montrer que la fonction $x \longmapsto xf(x)$ est convexe si et seulement si $x \longmapsto f\left(\frac{1}{x}\right)$ est convexe.

(École polytechnique)

▷ **Solution.**

Posons pour $x > 0$, $g(x) = xf(x)$ et $h(x) = f\left(\frac{1}{x}\right)$.

• On peut déjà vérifier le résultat lorsque f est deux fois dérivable. On a $g''(x) = 2f'(x) + xf''(x)$ et $h''(x) = \frac{1}{x^3}\left(2f'\left(\frac{1}{x}\right) + \frac{1}{x}f''\left(\frac{1}{x}\right)\right) = \frac{1}{x^3}g''\left(\frac{1}{x}\right)$ pour $x > 0$. Il apparaît alors que $g'' \geqslant 0$ équivaut à $h'' \geqslant 0$, ce qui permet de conclure pour la convexité. Traitons maintenant le cas général en revenant à la définition de la convexité.

• Supposons g convexe et montrons que h est convexe. Considérons des réels strictement positifs $x < z < y$. Il s'agit d'établir que

$$f\left(\frac{1}{z}\right) \leqslant \frac{y-z}{y-x}f\left(\frac{1}{x}\right) + \frac{z-x}{y-x}f\left(\frac{1}{y}\right).$$

Comme $\frac{1}{y} \leqslant \frac{1}{z} \leqslant \frac{1}{x}$ et g est convexe, on peut écrire

$$\frac{1}{z}f\left(\frac{1}{z}\right) \leqslant \frac{\frac{1}{z}-\frac{1}{y}}{\frac{1}{x}-\frac{1}{y}}\frac{1}{x}f\left(\frac{1}{x}\right) + \frac{\frac{1}{x}-\frac{1}{z}}{\frac{1}{x}-\frac{1}{y}}\frac{1}{y}f\left(\frac{1}{y}\right),$$

ce qui donne, après simplification, l'inégalité voulue.

• Supposons h convexe. On peut écrire $h(x) = x\frac{1}{x}f\left(\frac{1}{x}\right) = x\mathrm{F}(x)$, où $\mathrm{F} : x \in \mathbb{R}_+^* \longmapsto \frac{1}{x}f\left(\frac{1}{x}\right)$. D'après ce que nous venons de faire, la fonction $x \longmapsto \mathrm{F}\left(\frac{1}{x}\right)$ est convexe. Or, si $x > 0$, $\mathrm{F}\left(\frac{1}{x}\right) = xf(x) = g(x)$. La fonction g est bien convexe. ◁

Si f est une fonction convexe sur un intervalle I, *on montre facilement par récurrence sur n que si $x_1, \dots, x_n$ sont des points de* I, *et $t_1, \dots, t_n$ des réels positifs de somme* 1, *on a*

$$f(t_1x_1 + \cdots + t_nx_n) \leqslant t_1f(x_1) + \cdots + t_nf(x_n).$$

Cela offre un outil puissant pour établir des inégalités. Le lecteur pourra par exemple retrouver l'inégalité arithmético-géométrique en utilisant la convexité de la fonction $x \longmapsto -\ln x$. Les exercices ci-après proposent d'autres exemples.

4.48. Inégalité de convexité (1)

Soit $a, b \geqslant 0$ tels que $a + b = 1$. Montrer que pour tous réels positifs x et y, on a $1 + x^a y^b \leqslant (1 + x)^a (1 + y)^b$.

(École polytechnique)

▷ **Solution.**

C'est trivial si $x = 0$ ou $y = 0$. Supposons $x > 0$ et $y > 0$. On pose $\mathrm{X} = \ln x$, $\mathrm{Y} = \ln y$. Il s'agit de prouver que

$$\ln\left(1 + e^{a\mathrm{X}+b\mathrm{Y}}\right) \leqslant a\ln(1 + e^{\mathrm{X}}) + b\ln(1 + e^{\mathrm{Y}}).$$

On reconnaît là une inégalité de convexité. Soit $f : u \in \mathbb{R} \longmapsto \ln(1+e^u)$. La fonction f est de classe $\mathcal{C}^\infty$ et un petit calcul montre que

$$f''(u) = \frac{e^u}{(1 + e^u)^2} \geqslant 0.$$

Donc f est convexe et $f(a\mathrm{X}+b\mathrm{Y}) \leqslant af(\mathrm{X})+bf(\mathrm{Y})$, ce qui est l'inégalité espérée. ◁

4.49. Inégalité de convexité (2)

Soit $\mathrm{A} \in \mathcal{M}_n(\mathbb{R})$ une matrice *bistochastique* (c'est-à-dire telle que $a_{ij} \geqslant 0$ pour tout couple $(i, j) \in [\![1, n]\!]^2$ et que la somme des coefficients de chaque ligne et de chaque colonne de A soit égale à 1). Soit $\mathrm{X} = (x_1, \dots, x_n)$ un vecteur de $\mathbb{R}^n$ à coordonnées positives. On pose $\mathrm{Y} = \mathrm{AX}$. Montrer que $\prod_{i=1}^{n} y_i \geqslant \prod_{i=1}^{n} x_i$.

(École normale supérieure)

▷ **Solution.**

Si l'un des x_i est nul l'inégalité est évidente. On supposera donc les x_i strictement positifs. On a pour tout $i \in [\![1,n]\!]$, $y_i = \sum_{j=1}^{n} a_{ij}x_j$. Il s'agit d'une combinaison convexe des x_i. Par concavité de la fonction logarithme sur $\mathbb{R}_+^*$ on a

$$\forall i \in [\![1,n]\!], \quad \ln y_i = \ln\left(\sum_{j=1}^{n} a_{ij}x_j\right) \geqslant \sum_{j=1}^{n} a_{ij}\ln x_j.$$

En additionnant ces inégalités il vient

$$\sum_{i=1}^{n}\ln y_i \geqslant \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}\ln x_j = \sum_{j=1}^{n}\underbrace{\left(\sum_{i=1}^{n} a_{ij}\right)}_{=1}\ln x_j = \sum_{j=1}^{n}\ln x_j.$$

Le résultat est donc établi en passant à l'exponentielle. ◁

4.50. Condition suffisante de convexité

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction continue telle que, pour x, y réels,

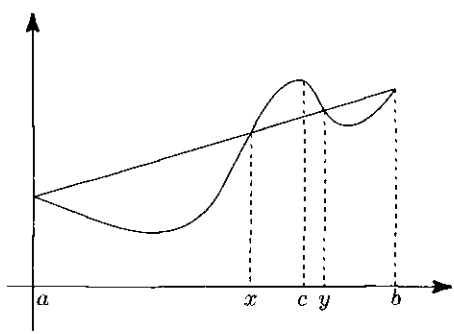
$$f\left(\frac{x+y}{2}\right) \leqslant \frac{f(x)+f(y)}{2}.$$

Montrer que f est convexe.

(École polytechnique)

▷ **Solution.**

On raisonne par l'absurde en supposant que f n'est pas convexe. On peut donc trouver $a < b$ et $c \in]a,b[$ tel que le point $(c, f(c))$ soit au-dessus de la corde entre $(a, f(a))$ et $(b, f(b))$.



Quitte à retrancher à f la fonction affine $x \longmapsto f(a) + \dfrac{f(b)-f(a)}{b-a}(x-a)$, on peut supposer $f(a) = f(b) = 0$. (Le dessin est fait dans le cas général.) On a alors $f(c) > 0$. L'idée du raisonnement apparaît alors clairement : on va obtenir une contradiction avec l'hypothèse en prenant les deux zéros de f de part et d'autre de c qui sont les plus proches de c. Précisons cela. L'ensemble $\{u \in [a,c],\ f(u) = 0\}$ est non vide (car il contient a), majoré par définition et fermé car f est continue. Il admet donc un plus grand élément x. On a $x < c$ car $f(c) > 0$ et f est strictement positive sur l'intervalle $]x,c]$ par le théorème des valeurs intermédiaires. De la même manière on justifie l'existence de $y > c$ tel que $f(y) = 0$ et $f > 0$ sur $[c,y[$. On a alors $f\left(\dfrac{x+y}{2}\right) > 0$ alors que $\dfrac{f(x)+f(y)}{2} = 0$ ce qui est la contradiction annoncée. $\triangleleft$

La solution ci-dessus montre, plus généralement, qu'une fonction continue f vérifiant

$$\forall (x,y) \in \mathbb{R}^2,\quad \exists t \in]0,1[,\quad f((1-t)x+ty) \leqslant (1-t)f(x) + tf(y),$$

est convexe (l'exercice est le cas particulier où on peut toujours prendre $t = 1/2$).

Le résultat de l'exercice ne reste pas vrai sans l'hypothèse de continuité. Il suffit de prendre pour f un morphisme discontinu du groupe $(\mathbb{R},+)$ dans lui-même. Le lecteur pourra en revanche montrer, mais c'est plus difficile, qu'avec l'hypothèse f majorée sur un intervalle ouvert I donné, on peut encore conclure que f est convexe.

4.51. Minimum d'une fonction convexe

Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction convexe.

1. Montrer que si f admet un minimum local, il s'agit d'un minimum global. Que peut-on dire de l'ensemble des points où ce minimum est atteint ?

2. On suppose que f est dérivable et qu'il existe un réel a tel que $f'(a) = 0$. Montrer que f admet en a un minimum global.

3. On suppose maintenant que f est deux fois dérivable et qu'il existe $\alpha \in \mathbb{R}$ tel que $f'' \geqslant \alpha > 0$. Montrer que f possède un minimum unique. A-t-on encore ce résultat avec l'hypothèse $f'' > 0$?

(École normale supérieure)

$\triangleright$ **Solution.**

1. Soit $a \in \mathbb{R}$ un point où f admet un minimum local. On peut donc trouver $\eta > 0$ tel que $f(a) \leqslant f(x)$ pour tout $x \in [a-\eta, a+\eta]$. Prenons $x > a + \eta$. D'après l'inégalité des pentes croissantes, on a

$$0 \leqslant \frac{f(a+\eta)-f(a)}{\eta} \leqslant \frac{f(x)-f(a)}{x-a}.$$

Il en résulte en particulier que $f(a) \leqslant f(x)$. On montre de même que pour tout $x < a-\eta$, $f(a) \leqslant f(x)$ de sorte que f admet en a un minimum global.

Supposons que b soit un autre point où ce minimum est atteint : $f(a) = f(b)$. Par convexité de f on a $f(x) \leqslant f(a)$ pour tout $x \in [a,b]$. Mais comme f atteint son minimum en a, on a nécessairement égalité. Il en résulte que f est constante sur tout le segment $[a,b]$. Cela prouve que l'ensemble des points où f atteint son minimum est une partie convexe de $\mathbb{R}$, donc un intervalle.

2. Comme f est convexe, la fonction f' est croissante. Elle est donc négative sur $]-\infty,a[$ et positive sur $]a,+\infty[$. La fonction f est donc décroissante sur $]-\infty,a[$ et croissante sur $]a,+\infty[$. Elle possède donc un minimum global en a.

3. La fonction $x \longmapsto f'(x)-\alpha x$ est de dérivée positive, donc croissante sur $\mathbb{R}$. Ainsi, pour $x \geqslant 0$, $f'(x)-\alpha x \geqslant f(0)$ et donc $f'(x) \geqslant \alpha x + f'(0)$. On en déduit que $\lim\limits_{+\infty} f' = +\infty$. On a de même, $f'(x) \leqslant \alpha x + f'(0)$ si $x \leqslant 0$. On en déduit que $\lim\limits_{-\infty} f' = -\infty$. Comme f' est strictement croissante sur $\mathbb{R}$, elle s'annule en un unique point a. D'après la question précédente f admet en a un minimum global. Il est nécessairement unique, car en un minimum local de f la dérivée f' doit s'annuler.

L'existence d'un minimum n'est plus assurée avec la seule hypothèse $f'' > 0$ comme le montre l'exemple de la fonction exponentielle. ◁

Bien entendu, pour une fonction dérivable quelconque $f : I \to \mathbb{R}$ (I ouvert), la condition $f'(a) = 0$ n'est pas suffisante pour que f admette un extremum (même local) en a. Toutefois, si f est de classe C^2 avec $f'(a) = 0$ et $f''(a) > 0$, alors f admet un minimum local en a (elle est convexe sur un intervalle ouvert contenant a).

4.52. Intervalle de monotonie d'une fonction strictement convexe

Soit f une fonction strictement convexe de $[a,b]$ dans $\mathbb{R}$ (avec $a < b$). Montrer qu'il existe un sous-intervalle J non vide de $[a,b]$, non réduit à un point, tel que la restriction de f à J soit strictement monotone.

(École polytechnique)

▷ **Solution.**

Considérons $a' < b'$ dans $]a,b[$. Comme f est convexe sur $[a,b]$, f est continue sur $[a',b']$ et il existe $\alpha \in [a',b']$ tel que $f(\alpha) = \min\limits_{x\in[a',b']} f(x)$. Par

stricte convexité, pour $x \neq \alpha$ dans $[a', b']$, $f(x) > f(\alpha)$. D'autre part, f n'admet pas de minimum local sur $[a', b']$ autre que α. En effet, par convexité, tout minimum local est automatiquement global.

Supposons $\alpha \neq b'$. Alors f est injective sur $[\alpha, b']$: en effet, si on suppose que $f(x) = f(y)$ avec $x < y$ dans $[\alpha, b']$, f admet par compacité un minimum sur $[x, y]$ en un point z. Or, par stricte convexité, z est dans l'intervalle ouvert $]x, y[$ et f admet donc en z un minimum local, ce qui est exclu. Comme une fonction continue injective sur un intervalle est strictement monotone (cf. les propos précédant l'exercice 4.12), f est bien strictement monotone sur $[\alpha, b']$.

Si $\alpha = b'$, on montre de même que f est strictement monotone sur $[a', \alpha] = [a', b']$. ◁

4.53. Étude asymptotique d'une fonction convexe

Soit $f : \mathbb{R}_+^* \to \mathbb{R}$ une application convexe.

1. Montrer que $\dfrac{f(x)}{x}$ admet une limite l dans $\mathbb{R} \cup \{+\infty\}$ quand x tend vers $+\infty$.

2. Montrer que si $l \leqslant 0$, alors f est décroissante.

3. Montrer que si l est fini, alors $f(x) - lx$ admet une limite dans $\overline{\mathbb{R}}$ quand x tend vers $+\infty$.

(École polytechnique)

▷ **Solution.**

1. Par le théorème des pentes croissantes, la fonction $\psi(x) = \dfrac{f(x) - f(1)}{x - 1}$ est croissante, donc admet une limite l dans $\mathbb{R} \cup \{+\infty\}$ quand x tend vers $+\infty$. Comme $\dfrac{f(x)}{x} = \dfrac{x-1}{x}\psi(x) + \dfrac{f(1)}{x}$, $\dfrac{f(x)}{x}$ tend également vers l en $+\infty$.

2. Si f n'est pas décroissante, il existe $a < b$ tels que $f(a) < f(b)$. Toujours en vertu du théorème des pentes croissantes, on aurait $\dfrac{f(x) - f(a)}{x - a} \geqslant \dfrac{f(b) - f(a)}{b - a} > 0$ pour tout $x \geqslant b$, ce qui est impossible car le terme de gauche de l'inégalité tend vers $l \leqslant 0$ en $+\infty$.

3. La fonction $g(x) = f(x) - lx$ est également convexe et $\dfrac{f(x)}{x}$ tend vers 0 en $+\infty$. D'après la question précédente, elle est décroissante sur $\mathbb{R}_+^*$. Elle admet donc une limite dans $\mathbb{R} \cup \{-\infty\}$ lorsque x tend vers $+\infty$. ◁

Une combinaison linéaire à coefficients positifs (en particulier une combinaison convexe) de fonctions convexes est encore convexe. L'exer-

cice suivant donne une condition nécessaire et suffisante pour que l'enveloppe convexe de n fonctions convexes contienne une fonction positive.

4.54. Combinaison convexe positive

Soit $f_1, \dots, f_n$ n fonctions convexes et continues sur $[0,1]$ telles que $\sup(f_1, \dots, f_n) \geqslant 0$. Montrer qu'il existe n réels positifs $\alpha_1, \dots, \alpha_n$ de somme 1 tels que $\alpha_1 f_1 + \cdots + \alpha_n f_n \geqslant 0$.

(École normale supérieure)

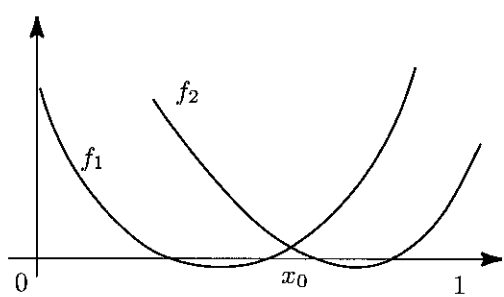
▷ **Solution.**

On va commencer par regarder le cas $n = 2$ avant d'envisager une récurrence sur n. Notons h la fonction $\sup(f_1, f_2)$. Si f_1 ou f_2 est positive, le résultat est trivial puisqu'il suffit de prendre $\alpha_1 = 1$ et $\alpha_2 = 0$ ou l'inverse. On suppose donc dans la suite que $\min f_1 < 0$ et $\min f_2 < 0$. La fonction h est continue et convexe sur $[0,1]$ et atteint son minimum en un point $x_0 \in [0,1]$. Par hypothèse $h(x_0) \geqslant 0$. On a alors les faits suivants.

• On a nécessairement $f_1(x_0) = f_2(x_0) = h(x_0)$. En effet, si par exemple $f_1(x_0) < f_2(x_0) = h(x_0)$, alors par continuité, h coïncide avec f_2 sur un petit voisinage de x_0. Si $x_0 \in]0,1[$, f_2 admet un minimum local en x_0, mais comme elle est convexe, il s'agit d'un minimum global (voir exercice 4.51) et cela contredit l'hypothèse $\min f_2 < 0$. On obtient la même contradiction si $x_0 = 0$ ou $x_0 = 1$.

• La fonction h n'atteint son minimum qu'en x_0. En effet, si $h(x_0') = h(x_0)$, avec par exemple $x_0 < x_0'$, alors h est constante sur le segment $[x_0, x_0']$. Mais d'après le point précédent, f_1 et f_2 seraient égales sur ce segment et par convexité positives sur $[0,1]$, ce qu'on a exclu.

On se retrouve donc avec une situation du type suivant :



On suppose dans la suite $x_0 \in]0,1[$, les cas où x_0 est au bord se traitant de la même manière.

• Montrons que les nombres dérivés à gauche en x_0 de f_1 et f_2 sont de signes opposés (au sens large). En effet, si $(f_1)'_g(x_0)$ et $(f_2)'_g(x_0)$ sont strictement positifs, f_1 et f_2 prennent toutes les deux des valeurs strictement inférieures à $h(x_0)$ à gauche de x_0, contredisant le fait que $h(x_0) = \min h$. Si $(f_1)'_g(x_0)$ et $(f_2)'_g(x_0)$ sont strictement négatifs, f_1 et f_2 prennent des valeurs supérieures à $h(x_0)$ à gauche de x_0. Considérons ce qui se passe à droite de x_0. Les deux nombres $(f_1)'_g(x_0)$ et $(f_2)'_g(x_0)$ ne peuvent pas être tous les deux strictement négatifs. Si on a par exemple $(f_1)'_d(x_0) \geqslant 0$, alors f_1 prend des valeurs supérieures à $h(x_0)$ à droite de x_0. La fonction f_1 a un minimum $h(x_0)$ positif, contrairement à notre hypothèse. On peut donc trouver deux réels positifs α_1 et α_2 de somme 1 tels que $g = \alpha_1 f_1 + \alpha_2 f_2$ ait une dérivée à gauche en x_0 nulle. Le théorème des pentes croissantes montre alors que $\dfrac{g(x) - g(x_0)}{x - x_0}$ est du signe de $x - x_0$, c'est-à-dire que g atteint son minimum en x_0. Comme $g(x_0) = h(x_0) \geqslant 0$ (car $f_1(x_0) = f_2(x_0)$), g est positive sur $[0, 1]$.

• On peut maintenant traiter le cas général par récurrence sur $n \geqslant 2$. Supposons le résultat prouvé avec $n - 1$ fonctions où $n \geqslant 3$ et prenons n fonctions $f_1, \ldots, f_n$ vérifiant les hypothèses. On a

$$h = \sup(f_1, \ldots, f_n) = \sup(f_1, \ldots, f_{n-2}, \sup(f_{n-1}, f_n)) \geqslant 0.$$

Comme $\sup(f_{n-1}, f_n)$ est convexe et continue, on peut appliquer l'hypothèse de récurrence. Il existe donc $\alpha_1, \ldots, \alpha_{n-1}$ réels positifs de somme 1 tels que

$$\alpha_1 f_1 + \alpha_2 f_2 + \cdots + \alpha_{n-1} \sup(f_{n-1}, f_n) \geqslant 0.$$

On pose alors $f = \alpha_1 f_1 + \cdots + \alpha_{n-2} f_{n-2} + \alpha_{n-1} f_{n-1}$ et $g = \alpha_1 f_1 + \cdots + \alpha_{n-2} f_{n-2} + \alpha_{n-1} f_n$. Ce sont deux fonctions convexes continues qui vérifient $\sup(f, g) \geqslant 0$. D'après le cas $n = 2$, on peut trouver $\alpha \in [0, 1]$ tel que $\alpha f + (1 - \alpha) g \geqslant 0$, ce qui donne

$$\alpha_1 f_1 + \cdots + \alpha_{n-2} f_{n-2} + \alpha\alpha_{n-1} f_{n-1} + (1 - \alpha)\alpha_{n-1} f_n \geqslant 0.$$

Cela termine la récurrence.

On notera que la réciproque du résultat de l'exercice est évidente. ◁

L'exercice suivant étudie les fonctions à variation bornée.

4.55. Fonctions à variation bornée

Soit une application $f : [a,b] \to \mathbb{R}$. À toute subdivision $\sigma = (a = x_0 < x_1 < \cdots < x_n = b)$ de $[a,b]$, on associe la *variation totale* de f sur σ : $\mathrm{V}(f,\sigma) = \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)|$. On pose $\mathrm{V}_a^b(f) = \sup_\sigma \mathrm{V}(f,\sigma) \in [0,+\infty]$, que l'on appelle la *variation* de f sur $[a,b]$. Si $\mathrm{V}_a^b(f)$ est fini, on dit que f est à *variation bornée* sur $[a,b]$.

1. Montrer que si f est $\mathcal{C}^1$ sur $[a,b]$, f est à variation bornée sur $[a,b]$, et calculer sa variation. Examiner le cas où f est lipschitzienne ou monotone.

2. Comparer $\mathrm{V}_a^b(f+g)$ à $\mathrm{V}_a^b(f) + \mathrm{V}_a^b(g)$.

3. Pour $c \in]a,b[$, comparer la variation de f sur $[a,b]$ à la somme de ses variations sur $[a,c]$ et $[c,b]$.

4. Montrer que f est à variation bornée sur $[a,b]$ si et seulement si f est somme de deux fonctions monotones sur $[a,b]$.

(École polytechnique)

▷ **Solution.**

1. Pour toute subdivision $\sigma = (a = x_0 < x_1 < \cdots < x_n = b)$ de $[a,b]$ et $k \in [\![0, n-1]\!]$, on a

$$|f(x_{k+1}) - f(x_k)| = \left| \int_{x_k}^{x_{k+1}} f'(t)\mathrm{d}t \right| \leqslant \int_{x_k}^{x_{k+1}} |f'(t)|\mathrm{d}t,$$

et donc $\mathrm{V}(f,\sigma) \leqslant \int_a^b |f'(t)|\mathrm{d}t$. La fonction f est donc à variation bornée sur $[a,b]$ et $\mathrm{V}_a^b(f) \leqslant \int_a^b |f'(t)|\mathrm{d}t$. Montrons, qu'en fait $\mathrm{V}_a^b(f)$ est égal à $\int_a^b |f'(t)|\mathrm{d}t$. Si σ est une subdivision de $[a,b]$ alors, pour tout $k \in [\![0, n-1]\!]$, il existe, d'après la formule des accroissements finis, $\xi_k \in [x_k, x_{k+1}]$ tel que $f(x_{k+1}) - f(x_k) = (x_{k+1} - x_k) f(\xi_k)$. On obtient $\mathrm{V}(f,\sigma) = \sum_{k=0}^{n-1} (x_{k+1} - x_k)|f'(\xi_k)|$. On reconnaît une somme de Riemann. Lorsque le pas de la subdivision tend vers 0, $\mathrm{V}(f,\sigma)$ tend vers $\int_a^b |f'(t)|\mathrm{d}t$, puisque $|f'|$ est continue. Pour tout $\varepsilon > 0$, il existe une subdivision σ telle que $\mathrm{V}(f,\sigma) \geqslant \int_a^b |f'(t)|\mathrm{d}t - \varepsilon$. On peut conclure que $\mathrm{V}_a^b(f) = \int_a^b |f'(t)|\mathrm{d}t$.

Si f est α-lipschitzienne, on a, avec les notations précédentes, pour

toute subdivision σ de $[a,b]$, $V(f,\sigma) \leqslant \sum_{k=0}^{n-1} \alpha(x_{k+1} - x_k) = \alpha(b-a)$. Ainsi f est à variation bornée sur $[a,b]$ et $V_a^b(f) \leqslant \alpha(b-a)$.

Si f est croissante sur $[a,b]$, on obtient, pour toute subdivision σ de $[a,b]$, $V(f,\sigma) = \sum_{k=0}^{n-1} (f(x_{k+1}) - f(x_k)) = f(b) - f(a)$. Ainsi f est à variation bornée sur $[a,b]$ et $V_a^b(f) = f(b) - f(a)$. On démontre, de même qu'une fonction décroissante est à variation bornée et que $V_a^b(f) = f(a) - f(b)$.

2. Montrons que si f et g sont deux fonctions définies sur $[a,b]$, on a $V_a^b(f+g) \leqslant V_a^b(f) + V_a^b(g)$. L'inégalité est évidente si $V_a^b(f)$ ou $V_a^b(g)$ est infini. Supposons donc f et g à variation bornée. Alors, pour toute subdivision σ de $[a,b]$,

$$\begin{aligned} V(f+g,\sigma) &= \sum_{k=0}^{n-1} |(f+g)(x_{k+1}) - (f+g)(x_k)| \\ &\leqslant \sum_{k=0}^{n-1} |f(x_{k+1}) - f(x_k)| + \sum_{k=0}^{n-1} |g(x_{k+1} - g(x_k)| \\ &\leqslant V(f,\sigma) + V(g,\sigma) \leqslant V_a^b(f) + V_a^b(g). \end{aligned}$$

On en déduit que $f+g$ est à variation bornée et que $V_a^b(f+g) \leqslant V_a^b(f) + V_a^b(g)$.

3. Soit σ une subdivision de $[a,b]$, $\sigma' = (a = x_0 < x_1 < \cdots < x_n = b)$ la subdivision de $[a,b]$ obtenue en rajoutant à σ le point c s'il n'y est pas. Notons p l'entier tel que $x_p = c$; $\sigma_1 = (a = x_0 < x_1 < \cdots < x_p = c)$ est une subdivision de $[a,c]$ et $\sigma_2 = (x_p < \cdots < x_n = b)$ une subdivision de $[c,b]$. On a $V(f,\sigma) \leqslant V(f,\sigma')$: on rajoute au plus un point dans la subdivision et l'inégalité résulte de l'inégalité triangulaire. D'autre part, $V(f,\sigma') = V(f,\sigma_1) + V(f,\sigma_2)$. On en déduit

$$V(f,\sigma) \leqslant V(f,\sigma_1) + V(f,\sigma_2) \leqslant V_a^c(f) + V_c^b(f).$$

On a donc $V_a^b(f) \leqslant V_a^c(f) + V_c^b(f)$. Montrons qu'il y a égalité. Supposons que f est à variation bornée sur $[a,c]$ et sur $[b,c]$. Soit $\varepsilon > 0$. On peut trouver des subdivisions σ_1 et σ_2 de $[a,c]$ et $[c,b]$ respectivement telle que : $V(f,\sigma_1) \geqslant V_a^c(f) - \frac{\varepsilon}{2}$ et $V(f,\sigma_2) \geqslant V_c^b(f) - \frac{\varepsilon}{2}$. Considérons la subdivision σ de $[a,b]$ obtenue en faisant la réunion de σ_1 et σ_2. On a alors $V(f,\sigma) = V(f,\sigma_1) + V(f,\sigma_2) \geqslant V_a^c(f) + V_c^b(f) - \varepsilon$. On obtient finalement $V_a^b(f) = V_a^c(f) + V_c^b(f)$.

Si $V_a^c(f)$ ou $V_c^b(f)$ est infini, alors $V_a^b(f)$ est infini, car si σ_1 est une subdivision quelconque de $[a,c]$, σ_2 une subdivision quelconque de $[b,c]$

et σ la réunion des deux, alors $V(f,\sigma) = V(f,\sigma_1) + V(f,\sigma_2)$, quantité qui n'est pas bornée.

4. D'après ce qui précède une fonction monotone est à variation bornée et la somme de deux fonctions à variation bornée est une fonction à variation bornée : la somme de deux fonctions monotones est donc une fonction à variation bornée.

Réciproquement, supposons que f est à variation bornée sur $[a,b]$. Alors pour tout $x \in [a,b]$, f est à variation bornée sur $[a,x]$. Soit g la fonction définie sur $[a,b]$ par $g(x) = V_a^x(f)$. Si $a \leqslant x \leqslant x' \leqslant b$, alors $g(x') = g(x) + V_x^{x'}(f) \geqslant g(x)$; g est donc croissante. D'autre part, avec les mêmes notations, $V_x^{x'}(f) \geqslant |f(x') - f(x)|$ (on prend la subdivision $\sigma = (x < x')$). On en déduit $g(x') \geqslant g(x) + f(x') - f(x)$, c'est-à-dire $f(x') - g(x') \leqslant f(x) - g(x)$. La fonction $h = f - g$ est décroissante. La fonction $f = g + h$ est somme d'une fonction croissante et d'une fonction décroissante. ◁

La notion de fonction à variation bornée intervient notamment dans l'étude des arcs rectifiables. Si au lieu de prendre f à valeurs réelles, on prend f à valeurs dans $\mathbb{R}^2$ ou $\mathbb{R}^3$, et si on remplace la valeur absolue par la norme euclidienne, la quantité $V_a^b(f)$ n'est autre que la longueur de l'arc défini par f. Cet arc est dit rectifiable si $V_a^b(f)$ est fini.

Le dernier thème de ce chapitre est celui des estimations asymptotiques. On peut y faire appel à différentes notions déjà mentionnées.

4.56. Un développement asymptotique

Pour $\lambda \in]0,1]$, on pose $f(\lambda) = \sup_{x \in [0, \frac{1}{\lambda}]} \left(e^{-x} - (1-\lambda x)^{\frac{1}{\lambda}}\right)$.

Déterminer un développement asymptotique à deux termes de la fonction f en 0.

(École polytechnique)

▷ **Solution.**

La fonction f est définie sur $]0,1]$, car $g_\lambda : x \longmapsto e^{-x} - (1-\lambda x)^{\frac{1}{\lambda}}$ est continue sur $\left[0, \frac{1}{\lambda}\right]$. Étudions les variations de g_λ. On a, pour $x \in \left[0, \frac{1}{\lambda}\right]$, $g'_\lambda(x) = -e^{-x} + (1-\lambda x)^{\frac{1}{\lambda}-1}$. La fonction g'_λ a même signe que h_λ, où $h_\lambda(x) = x + \left(\frac{1}{\lambda} - 1\right)\ln(1-\lambda x)$. Pour $x \in \left[0, \frac{1}{\lambda}\right[$, on a $h'_\lambda(x) = \frac{\lambda(1-x)}{1-\lambda x}$. La fonction h'_λ s'annule en 1 et h_λ croît sur $[0,1]$ et décroît sur $[1, \frac{1}{\lambda}[$. Comme on a de plus, $h_\lambda(0) = 0$ et $\lim_{x \to \frac{1}{\lambda}} h_\lambda(x) = -\infty$, on conclut que h_λ

est positive sur $[0,1]$ et s'annule une fois sur $\left[1, \frac{1}{\lambda}\right[$, en α ; la fonction g'_λ s'annule en α, est positive sur $[0,\alpha[$ et négative sur $]\alpha, \frac{1}{\lambda}[$. On a donc $f(\lambda) = g_\lambda(\alpha)$. Sachant que $g'_\lambda(\alpha) = -e^{-\alpha} + (1-\lambda\alpha)^{\frac{1}{\lambda}-1} = 0$, on obtient

$$f(\lambda) = e^{-\alpha} - (1-\lambda\alpha)^{\frac{1}{\lambda}} = (1-\lambda\alpha)^{\frac{1}{\lambda}-1} - (1-\lambda\alpha)^{\frac{1}{\lambda}} = (1-\lambda\alpha)^{\frac{1}{\lambda}-1}(\lambda\alpha).$$

En posant $u = \lambda\alpha$, on peut écrire $f(\lambda) = u(1-u)^{\frac{1}{\lambda}-1} = ue^{\left(\frac{1}{\lambda}-1\right)\ln(1-u)}$. Pour commencer, on détermine un développement asymptotique de u.

L'équation $h_\lambda(\alpha) = 0$ équivaut à $\lambda\alpha + (1-\lambda)\ln(1-\lambda\alpha) = 0$, soit $u + (1-\lambda)\ln(1-u) = 0$. Le réel u vérifie donc les conditions $u \in]\lambda, 1[$ et $\dfrac{\ln(1-u)}{u} = \dfrac{1}{\lambda-1}$. On étudie les variations de $\varphi : x \longmapsto \dfrac{\ln(1-x)}{x}$ sur $]0,1[$. On trouve $\varphi'(x) = \dfrac{\psi(x)}{x^2}$, où $\psi(x) = -\dfrac{x}{1-x} - \ln(1-x)$, puis $\psi'(x) = \dfrac{-x}{(1-x)^2} < 0$. Sachant que $\lim\limits_0 \psi = 0$, on en déduit que ψ est négative. La fonction φ est donc strictement décroissante. Elle réalise une bijection de $]0,1[$ sur $]-\infty,-1[$. On a donc $u = \varphi^{-1}\left(\dfrac{1}{\lambda-1}\right)$. On en déduit que $\lim\limits_{\lambda\to 0} u = \lim\limits_{-1} \varphi^{-1} = 0$. On écrit le développement limité de $\varphi(u)$ en 0 à l'ordre 2 :

$$\varphi(u) = \frac{\ln(1-u)}{u} = -1 - \frac{u}{2} - \frac{u^2}{3} + o(u^2)$$

et donc,

$$\varphi(u) + 1 = -\frac{u}{2} - \frac{u^2}{3} + o(u^2) = \frac{\lambda}{\lambda-1}.$$

On en déduit que $-\dfrac{u}{2} \underset{\lambda\to 0}{\sim} -\lambda$ et donc $u \underset{\lambda\to 0}{\sim} 2\lambda$. Ceci signifie que α tend vers 2 quand λ tend vers 0. Pour obtenir le terme suivant dans le développement de u, on écrit $u = 2\lambda(1+v)$, avec $v = o(1)$ et on reporte dans l'égalité précédente, qui devient

$$-\lambda(1+v) - \frac{4\lambda^2}{3} + o(\lambda^2) = \frac{\lambda}{\lambda-1},$$
$$v = \frac{-1}{\lambda-1} - 1 - \frac{4\lambda}{3} + o(\lambda) = -\frac{1}{3}\lambda + o(\lambda).$$

Finalement, on a le développement de u

$$u = 2\lambda\left(1 - \frac{1}{3}\lambda + o(\lambda)\right) = 2\lambda - \frac{2}{3}\lambda^2 + o(\lambda^2).$$

Compte tenu de ce qui précède, on peut simplifier l'expression de $f(\lambda)$. On a, en effet $\left(\dfrac{1}{\lambda} - 1\right)\ln(1-u) = \left(\dfrac{1}{\lambda} - 1\right)\dfrac{u}{\lambda-1} = -\dfrac{u}{\lambda}$. On en déduit

que

$$f(\lambda) = e^{-\frac{u}{\lambda}}u = e^{-2+\frac{2}{3}\lambda+o(\lambda)}(2\lambda - \frac{2}{3}\lambda^2 + o(\lambda^2)$$
$$= \lambda e^{-2}(1 + \frac{2}{3}\lambda + o(\lambda))(2 - \frac{2}{3}\lambda + o(\lambda)),$$

et finalement,

$$\boxed{f(\lambda) = 2e^{-2}\lambda + \frac{2}{3}e^{-2}\lambda^2 + o(\lambda^2).} \quad \triangleleft$$

4.57. Estimation du maximum d'une fonction polynôme

Soit $f_n : x \in [0,n] \mapsto \prod_{i=0}^{n}(x-i) \in \mathbb{R}$. Trouver un équivalent simple de $\|f_n\|_\infty$ lorsque n tend vers $+\infty$.

(**École normale supérieure**)

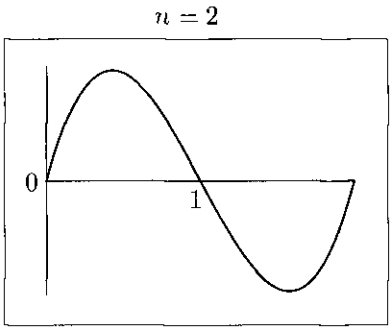
▷ **Solution.**

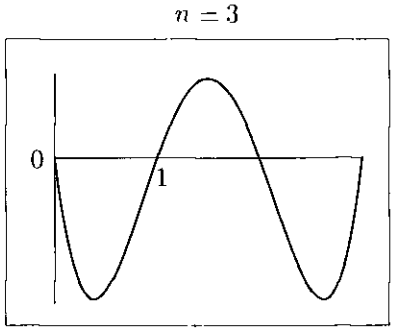
La fonction $|f_n|$ est continue sur $[0,n]$. Elle s'annule pour des valeurs entières. Sur chaque intervalle $[i,i+1]$, elle atteint son maximum sur l'intervalle $]i,i+1[$. De plus, f_n est dérivable sur $[0,n]$ et, pour $x \in [0,n]$, x non entier, on a

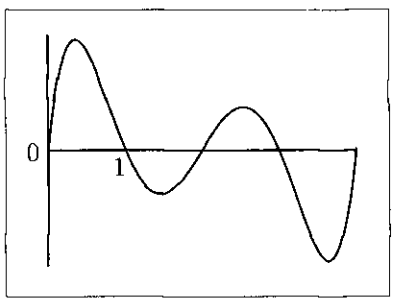
$$\frac{f_n'(x)}{f_n(x)} = \sum_{i=0}^{n}\frac{1}{x-i} = g_n(x).$$

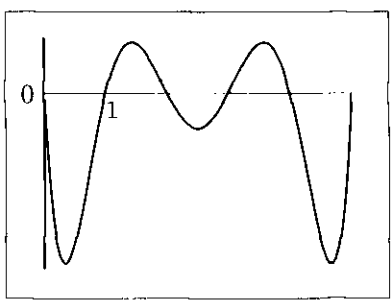
La fonction g_n est strictement décroissante sur chaque intervalle $]i,i+1[$ $(0 \leqslant i \leqslant n-1)$ et de plus, $\lim_{i^+} g_n = +\infty$ et $\lim_{(i+1)^-} g_n = -\infty$. Elle s'annule et change de signe une fois sur $]i,i+1[$. La fonction f_n gardant un signe constant sur cet intervalle, il en est de même pour f_n'. On en déduit que $|f_n|$ atteint son maximum sur $[i,i+1]$ en l'unique valeur qui annule g_n. Il nous faut maintenant déterminer sur quel intervalle $[i,i+1]$, le maximum de $|f_n|$ est atteint. Puisqu'on a, pour tout $x \in [0,n]$, $|f_n(n-x)| = |f_n(x)|$, on peut supposer $i < \frac{n}{2}$. Le tracé des courbes représentatives de quelques fonctions f_n indique clairement que le maximum de $|f_n|$ est atteint sur $[0,1]$ (et aussi sur $[n-1,n]$).

Pour le montrer, commençons par majorer $|f_n|$ sur $[i,i+1]$, pour $1 \leqslant i < \frac{n}{2}$.









On a, pour tout $x \in [i, i+1]$,

$$\begin{aligned}|f_n(x)| &= x(x-1)\dots(x-i+1)(x-i)(i+1-x)(i+2-x)\dots \\ &\quad \dots(n-1-x)(n-x) \\ &\leqslant (i+1)i\dots 2(x-i)(i+1-x)2\dots(n-i) \\ &\leqslant \frac{1}{4}(i+1)!(n-i)!,\end{aligned}$$

puisque $0 \leqslant x(1-x) \leqslant \frac{1}{4}$ si x est dans $[0,1]$. Posons $u_i = \frac{1}{4}(i+1)!(n-i)!$, pour $1 \leq i < \frac{n}{2}$. Montrons que la suite (u_i) est décroissante. Pour $1 \leqslant i < i+1 < \frac{n}{2}$, on a $\frac{u_{i+1}}{u_i} = \frac{i+2}{n-i} < 1$. On en déduit que $u_i \leqslant u_1 = \frac{(n-1)!}{2}$. On obtient finalement, pour tout $x \in [1, n-1]$,

$$|f_n(x)| \leqslant \frac{(n-1)!}{2}.$$

Voyons maintenant ce qui se passe sur $[0,1]$. La fonction $|f_n|$ y atteint son maximum en a_n, l'unique zéro de g_n dans $[0,1]$. Cherchons un équivalent de a_n puis de $|f_n(a_n)|$. On a $\sum_{i=0}^{n} \frac{1}{a_n - i} = 0$, ce qu'on écrit

$\frac{1}{a_n}+\frac{1}{a_n-1}=\sum_{i=2}^{n}\frac{1}{i-a_n}$. On en déduit, par décroissance de la fonction $t\longmapsto\frac{1}{t}$, que

$$\frac{1}{a_n}+\frac{1}{a_n-1}\leqslant\sum_{i=1}^{n-1}\frac{1}{i}\leqslant 1+\int_1^{n-1}\frac{1}{t}\mathrm{d}t\leqslant 1+\ln(n-1)\ \text{ et}$$
$$\frac{1}{a_n}+\frac{1}{a_n-1}\geqslant\sum_{i=2}^{n}\frac{1}{i}\geqslant\int_2^{n+1}\frac{1}{t}\mathrm{d}t\geqslant\ln(n+1)-\ln 2.$$

On obtient l'encadrement

$$\ln(n+1)-\ln 2\leqslant\frac{1}{a_n}+\frac{1}{a_n-1}\leqslant 1+\ln(n-1).$$

On en déduit que $\frac{1}{a_n}+\frac{1}{a_n-1}\underset{n\to\infty}{\sim}\ln n$ et donc $\lim\limits_{n\to+\infty}\frac{1}{a_n}+\frac{1}{a_n-1}=+\infty$. Puisque a_n-1 est négatif, on a *a fortiori* $\lim\limits_{n\to+\infty}\frac{1}{a_n}=+\infty$, c'est-à-dire $\lim\limits_{n\to+\infty}a_n=0$. On conclut finalement que

$$a_n\underset{n\to\infty}{\sim}\frac{1}{\ln n}.$$

Il faut maintenant déterminer un équivalent de $|f_n(a_n)|$. On écrit

$$|f_n(a_n)|=a_n\prod_{i=1}^{n}(i-a_n)=a_n n!\prod_{i=1}^{n}(1-\frac{a_n}{i}).$$

On sait que $a_n\underset{n\to\infty}{\sim}\frac{1}{\ln n}$. Il faut s'occuper du produit. On passe au logarithme et on considère $s_n=\sum\limits_{i=1}^{n}\ln\left(1-\frac{a_n}{i}\right)$. Chaque terme $\frac{a_n}{i}$ est inférieur à a_n et donc tend vers 0. Au voisinage de 0 on a $\ln(1-x)\underset{x\to 0}{\sim}-x$. Si on se donne $\varepsilon>0$, il existe $\eta>0$ tel que $|\ln(1-x)+x|\leqslant\varepsilon|x|$, pour $|x|\leqslant\eta$. On en déduit que, pour n assez grand (tel que $|a_n|\leqslant\eta$), on aura, pour tout $i\in[\![1,n]\!]$,

$$\left|\ln\left(1-\frac{a_n}{i}\right)+\frac{a_n}{i}\right|\leqslant\varepsilon\left|\frac{a_n}{i}\right|\ \text{ et donc }\ \left|s_n+\sum_{i=1}^{n}\frac{a_n}{i}\right|\leqslant\varepsilon\sum_{i=1}^{n}\frac{a_n}{i}.$$

Ceci montre que $s_n\underset{n\to\infty}{\sim}-a_n\sum\limits_{i=1}^{n}\frac{1}{i}$. Nous savons que $a_n\underset{n\to\infty}{\sim}\frac{1}{\ln n}$ et que $\sum\limits_{i=1}^{n}\frac{1}{i}\underset{n\to\infty}{\sim}\ln n$. On en déduit que $\lim\limits_{n\to\infty}s_n=-1$, et donc que $\lim\limits_{n\to\infty}\prod\limits_{i=1}^{n}\left(1-\frac{a_n}{i}\right)=\frac{1}{\mathrm{e}}$. On obtient finalement

$$|f_n(a_n)| \underset{n\to\infty}{\sim} \frac{n!}{e \ln n}.$$

Nous avons démontré précédemment que $\sup_{x\in[1,n-1]} |f_n(x)| \leqslant \frac{1}{4}(n-1)!$. Ce terme est négligeable devant $|f_n(a_n)|$. On en déduit que pour n assez grand, on a $\|f_n\|_\infty = |f_n(a_n)|$ et donc que

$$\boxed{\|f_n\|_\infty \underset{n\to\infty}{\sim} \frac{n!}{e \ln n}.} \quad \triangleleft$$

Notre majoration de $|f_n|$ sur $[1, n-1]$ a été très grossière. En fait, on peut démontrer que $\|f_n\|_\infty$ est toujours égal à $|f_n(a_n)|$.

De manière générale, il est interdit de dériver des équivalents. Toutefois, avec une hypothèse de monotonie en plus, on obtient tout de même des résultats.

4.58. Dérivation d'équivalents

Soit $f : \mathbb{R} \to \mathbb{R}$ continue et croissante. On pose $\mathrm{F}(x) = \int_0^x f(t)\mathrm{d}t$.

1. On suppose qu'il existe $\alpha > 0$ tel que $\mathrm{F}(x) \sim \dfrac{x^\alpha}{\alpha}$ en $+\infty$. Montrer que $f(x) \sim x^{\alpha-1}$.

2. On suppose, toujours en $+\infty$, que $\mathrm{F}(x) = \dfrac{x^2}{2} + o(x)$. Montrer que $f(x) = x + o(\sqrt{x})$.

(École normale supérieure)

▷ **Solution.**

1. Remarquons qu'il s'agit de démontrer, avec une hypothèse supplémentaire de monotonie, la réciproque d'un résultat connu. En effet, si on suppose simplement f continue et $f(x) \underset{x\to+\infty}{\sim} x^{\alpha-1}$, alors la fonction f est positive au voisinage de $+\infty$ et n'est pas intégrable sur $\mathbb{R}_+$. Le théorème d'intégration des relations d'équivalence permet d'affirmer que $\mathrm{F}(x) \underset{x\to+\infty}{\sim} \int_0^x t^{\alpha-1}\mathrm{d}t = \dfrac{x^\alpha}{\alpha}$. La réciproque serait fausse sans l'hypothèse supplémentaire «f croissante» (considérer par exemple $f(x) = 1 + \sin x$).

Revenons à la question posée. L'idée de la démonstration est d'intégrer f sur un intervalle où elle prend des valeurs proches de $f(x)$ et d'utiliser la croissance de f. Soit $h \in]0, 1[$, que l'on choisira plus tard. On a, par croissance de f, pour $x > 0$.

$$\varphi(x) = \mathrm{F}(x(1+h)) - \mathrm{F}(x) = \int_x^{x(1+h)} f(t)\mathrm{d}t \geqslant xhf(x) \text{ et}$$
$$\psi(x) = \mathrm{F}(x) - \mathrm{F}(x(1-h) = \int_{x(1-h)}^{x} f(t)\mathrm{d}t \leqslant xhf(x),$$

d'où l'on tire l'encadrement $\dfrac{\psi(x)}{x^\alpha h} \leqslant \dfrac{f(x)}{x^{\alpha-1}} \leqslant \dfrac{\varphi(x)}{x^\alpha h}$. Comme par hypothèse $\mathrm{F}(x) \sim \dfrac{x^\alpha}{\alpha}$ en $+\infty$, on déduit

$$\lim_{x\to+\infty} \frac{\varphi(x)}{x^\alpha h} = \frac{(1+h)^\alpha - 1}{\alpha h} \text{ et } \lim_{x\to+\infty} \frac{\psi(x)}{x^\alpha h} = \frac{1-(1-h)^\alpha}{\alpha h}.$$

Notons ensuite que

$$\lim_{h\to 0} \frac{(1+h)^\alpha - 1}{\alpha h} = \lim_{h\to 0} \frac{1-(1-h)^\alpha}{\alpha h} = 1.$$

Soit $\varepsilon > 0$. Fixons $h \in]0,1[$ tel que $\dfrac{(1+h)^\alpha - 1}{\alpha h}$ et $\dfrac{1-(1-h)^\alpha}{\alpha h}$ appartiennent à $[1-\varepsilon, 1+\varepsilon]$. Puisqu'alors $\lim\limits_{x\to+\infty} \dfrac{\varphi(x)}{x^\alpha h}$ et $\lim\limits_{x\to+\infty} \dfrac{\psi(x)}{x}$ appartiennent à $[1-\varepsilon, 1+\varepsilon]$, on aura, pour x assez grand, $\dfrac{\varphi(x)}{x}$ et $\dfrac{\psi(x)}{x}$ et donc $\dfrac{f(x)}{x^{\alpha-1}}$ dans $[1-2\varepsilon, 1+2\varepsilon]$. Ceci démontre que $\lim\limits_{x\to+\infty} \dfrac{f(x)}{x^{\alpha-1}} = 1$.

2. La méthode précédente ne conduit pas au résultat. Remplaçons $x(1+h)$ et $x(1-h)$ par $x+\eta(x)$ et $x-\eta(x)$, où η est une fonction positive à choisir telle que $\eta(x) = o(x)$ en $+\infty$. On obtient, comme précédemment

$$\varphi(x) = \mathrm{F}(x+\eta(x)) - \mathrm{F}(x) = \int_x^{x+\eta(x)} f(t)\mathrm{d}t \geqslant \eta(x)f(x),$$
$$\psi(x) = \mathrm{F}(x) - \mathrm{F}(x-\eta(x)) = \int_{x-\eta(x)}^{x} f(t)\mathrm{d}t \leqslant \eta(x)f(x)$$

et l'encadrement $\dfrac{\psi(x)}{\eta(x)} \leqslant f(x) \leqslant \dfrac{\varphi(x)}{\eta(x)}$. Par hypothèse, il existe une fonction ε telle que, $\mathrm{F}(x) = \dfrac{x^2}{2} + x\varepsilon(x)$ et $\lim\limits_{x\to+\infty} \varepsilon(x) = 0$. On en déduit que

$$\frac{\varphi(x)}{\eta(x)} = \frac{1}{\eta(x)}\left(\frac{1}{2}((x+\eta(x))^2 - x^2) + (x+\eta(x))\varepsilon(x+\eta(x)) - x\varepsilon(x)\right)$$
$$= x + \frac{\eta(x)}{2} + \varepsilon(x+\eta(x)) + \frac{x}{\eta(x)}\left(\varepsilon(x+\eta(x)) - \varepsilon(x)\right).$$

On obtient, de même,

$$\frac{\psi(x)}{\eta(x)} = x - \frac{\eta(x)}{2} + \varepsilon(x-\eta(x)) + \frac{x}{\eta(x)}\left(\varepsilon(x) - \varepsilon(x-\eta(x))\right).$$

Nous voulons que chacun de ces termes s'écrive $x + o(\sqrt{x})$. Pour cela il faut que $\eta(x) = o(\sqrt{x})$, $\dfrac{\sqrt{x}}{\eta(x)}\,(\varepsilon(x+\eta(x)) - \varepsilon(x)) = o(1)$ et que $\dfrac{\sqrt{x}}{\eta(x)}\,(\varepsilon(x) - \varepsilon(x-\eta(x))) = o(1)$. Posons $m(x) = \sup_{t\in[x-\sqrt{x},+\infty[} |\varepsilon(t)|$ et montrons que $\eta(x) = \sqrt{xm(x)}$ convient. En effet, on a $\lim_{+\infty} m = 0$, puisque $\lim_{+\infty} \varepsilon = 0$. On en déduit que $\eta(x) = o(\sqrt{x})$,

$$\left|\frac{\sqrt{x}}{\eta(x)}\,(\varepsilon(x+\eta(x)) - \varepsilon(x))\right| \leqslant \frac{2\sqrt{x}m(x)}{\eta(x)} \leqslant 2\sqrt{m(x)} = o(1) \text{ et}$$
$$\left|\frac{\sqrt{x}}{\eta(x)}\,(\varepsilon(x) - \varepsilon(x-\eta(x)))\right| \leqslant \frac{2\sqrt{x}m(x)}{\eta(x)} \leqslant 2\sqrt{m(x)} = o(1),$$

car pour x assez grand, on a x, $x+\eta(x)$ et $x-\eta(x)$ dans $[x-\sqrt{x},+\infty[$. Puisque $f(x)$ est encadré par deux termes de la forme $x + o(\sqrt{x})$, on peut conclure que

$$\boxed{f(x) \underset{x\to+\infty}{=} x + o(\sqrt{x}).} \quad \triangleleft$$

4.59. Intégration d'équivalents

1. Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ de classe $\mathcal{C}^1$, $\ell > 0$ et P un polynôme de degré n. On suppose que $f'(x)\mathrm{P}(f(x))$ tend vers ℓ quand x tend vers $+\infty$. Donner un équivalent de f en $+\infty$.

2. Soit $h : \mathbb{R} \to \mathbb{R}_+$ continue, $\ell > 0$, n un entier naturel. On suppose que $h(x)\displaystyle\int_0^x h^n(t)\mathrm{d}t$ tend vers ℓ lorsque x tend vers $+\infty$. Donner un équivalent de h en $+\infty$.

(École polytechnique)

▷ **Solution.**

1. Soit Q une primitive de P. L'hypothèse s'énonce $\lim_{+\infty}(\mathrm{Q}\circ f)' = \ell$. Montrons que si g est une fonction de classe $\mathcal{C}^1$ sur $\mathbb{R}_+$ telle que $\lim_{+\infty} g' = \ell \in \mathbb{R}$, alors $\lim_{x\to+\infty} \dfrac{g(x)}{x} = \ell$.

Soit $\varepsilon > 0$ et x_0 tel que $\ell - \varepsilon \leqslant g'(x) \leqslant \ell + \varepsilon$ si $x \geqslant x_0$. D'après le théorème des accroissements finis, on en déduit que, si $x > x_0$, on a

$$\ell - \varepsilon \leqslant \frac{g(x) - g(x_0)}{x - x_0} \leqslant \ell + \varepsilon$$

et donc

$$(\ell-\varepsilon)\frac{x-x_0}{x}+\frac{g(x_0)}{x}\leqslant\frac{g(x)}{x}\leqslant(\ell+\varepsilon)\frac{x-x_0}{x}+\frac{g(x_0)}{x}.$$

Puisque

$$\lim_{x\to+\infty}(\ell-\varepsilon)\frac{x-x_0}{x}+\frac{g(x_0)}{x}=\ell-\varepsilon$$

et

$$\lim_{x\to+\infty}(\ell+\varepsilon)\frac{x-x_0}{x}+\frac{g(x_0)}{x}=\ell+\varepsilon,$$

il existe $x_1\geqslant x_0$ tel que, pour $x\geqslant x_1$, on ait $(\ell-\varepsilon)\dfrac{x-x_0}{x}+\dfrac{g(x_0)}{x}\geqslant\ell-2\varepsilon$ et $(\ell+\varepsilon)\dfrac{x-x_0}{x}+\dfrac{g(x_0)}{x}\leqslant\ell+2\varepsilon$. On en déduit que, pour $x\geqslant x_1$, on a $\ell-2\varepsilon\leqslant\dfrac{g(x)}{x}\leqslant\ell+2\varepsilon$, ce qui démontre le résultat voulu.

On a, d'après ce qui précède, $\lim\limits_{x\to+\infty}\dfrac{(\mathrm{Q}\circ f)(x)}{x}=\ell>0$. On en déduit que $\lim\limits_{+\infty}\mathrm{Q}\circ f=+\infty$. Montrons que $\lim\limits_{+\infty}f=+\infty$. Soit $\mathrm{A}>0$. La fonction Q étant continue, est bornée sur le compact $[0,\mathrm{A}]$; il existe $\mathrm{C}>0$ tel que $\mathrm{Q}(f(x))\leqslant\mathrm{C}$ si $0\leqslant f(x)\leqslant\mathrm{A}$. D'autre part, il existe $x_0>0$ tel que $\mathrm{Q}(f(x))>\mathrm{C}$ si $x>x_0$. On en déduit, puisque f est positive, que $f(x)>\mathrm{A}$ si $x>x_0$, ce qui démontre le résultat annoncé.

Le polynôme Q est de degré $n+1$; si a est le coefficient dominant de P, on a $\mathrm{Q}(x)\underset{x\to+\infty}{\sim}\dfrac{ax^{n+1}}{n+1}$. On a donc, quand x tend vers $+\infty$, $\mathrm{Q}(f(x))\sim\dfrac{a(f(x))^{n+1}}{n+1}\sim\ell x$. On en déduit que

$$\boxed{f(x)\underset{x\to+\infty}{\sim}\left(\frac{\ell(n+1)}{a}x\right)^{\frac{1}{n+1}}.}$$

2. On se ramène à la question précédente en posant, $f(x)=\int_0^x h^n(t)\mathrm{d}t$, pour $x\geqslant0$. L'application f va de $\mathbb{R}_+$ dans $\mathbb{R}_+$ et est de classe $\mathcal{C}^1$. L'hypothèse entraîne que $\lim\limits_{x\to+\infty}h^n(x)\left(\int_0^x h(t)\mathrm{d}t\right)^n=\ell^n$, c'est-à-dire que $\lim\limits_{x\to+\infty}f'(x)(f(x))^n=\ell^n$. La fonction f vérifie les mêmes hypothèses que dans la question 1, avec $\mathrm{P}=\mathrm{X}^n$, ℓ étant remplacé par ℓ^n. Il résulte de la question 1 que $f(x)\underset{x\to+\infty}{\sim}(\ell^n(n+1)x)^{\frac{1}{n+1}}$. Sachant que $h(x)\underset{x\to+\infty}{\sim}\dfrac{\ell}{f(x)}$, on obtient finalement

$$\boxed{h(x)\underset{x\to+\infty}{\sim}\left(\frac{\ell}{(n+1)x}\right)^{\frac{1}{n+1}}.}\ \triangleleft$$

Index

A

Abel (transformation d'), 135
Abel-Dirichlet (théorème d'), 194
accéleration de convergence, 185

B

Baire (théorème de), 55, 248
Bertrand (séries de), 136, 149
Bolzano-Weierstrass (théorème de), 74
Borel (théorème de), 279
Borel-Lebesgue (théorème de), 47, 49, 225
Briggs (algorithme de), 18

C

Cantor
  procédé diagonal de, 12
  théorème de, 13
Carleman (inégalité de), 207
Cauchy
  critère de, 140, 144
  critère de condensation de, 147
  produit de, 198
  suite de, 236
  tranche de, 41, 72, 141, 144, 146, 149
Cauchy-Lipschitz (théorème de), 97
Cauchy-Schwarz (inégalité de), 134
Cesàro (théorème de), 63–72, 97, 99, 101, 102, 152, 190, 199
continuité uniforme, 245–248
convexité, 283–291
corde, 229
Croft (lemme de), 247

D

Darboux (théorème de), 254, 268
développement décimal d'un réel, 7–14
Dirichlet (théorème de), 79

E

Engel (série de), 16
équation fonctionnelle, 42, 113, 236–245
Euler
  constante d', 157, 162, 170
  méthode d', 98
  transformation d', 182

F

famille sommable, 200–212, 253
fonction
  à variation bornée, 291
  convexe, 283–291
  lipschitzienne, 254, 292
  propre, 219
  pseudo-dérivable, 249
  réglée, 224
  sur-additive, 222
  uniformément continue, 245–248
formule des accroissements finis, 252, 254, 255, 272, 273, 292

G

Glaeser (théorème de), 270

H

Hardy
  inégalité de, 210
  séries de, 192
  théorème taubérien de, 70
Heine (théorème de), 245
Hermite (interpolation d'), 264
homéomorphisme, 233, 239
homographie, 43

I

inégalité
  arithmético-géométrique, 134, 208, 274
  de Carleman, 207
  de Cauchy-Schwarz, 134, 210
  de convexité, 285–286
  de Hardy, 210
  de Kolmogorov, 274
  de Taylor-Lagrange, 274
  de Tchebychev, 33
  des accroissements finis, 75, 87, 89, 193, 252
  des pentes croissantes, 288
  du réordonnement, 31, 150
interversion de limites, 195, 196, 207

K

Kolmogorov (inégalité de), 274

L

Lebesgue (mesure extérieure de), 47
Lévy (théorème de), 229
limite inférieure, supérieure d'une suite réelle, 78
Liouville (nombres de), 25, 81

M

Möbius (formule d'inversion de), 168

P

Pisot (nombres de), 27, 175, 186
Poincaré (nombre de rotation de), 235
pseudo-dérivée, 249

R

rationalité et irrationalité, 21–25
Riemann
  séries de, 136, 144, 148, 157
  somme de, 165, 292
  théorème de, 38, 217
Rolle (théorème de), 252, 265, 267
Russell (paradoxe de), 13

S

Sard (lemme de), 253
Sarkowski (théorème de), 92
Schwob (suites de), 122
segments emboîtés (théorème des), 55
séries alternées (théorème des), 179, 181, 183, 185
Shapiro (équation fonctionnelle de), 42
sommation par paquets, 147–149, 185, 201
sous-groupes additifs de $\mathbb{R}$, 28
splines cubiques, 260
Steinitz (théorème de), 38
Stirling (formule de), 104, 132, 164, 197, 198, 208
système dynamique discret, 69, 85–103, 110

T

Taylor-Lagrange
  formule de, 21, 131, 265–276
  inégalité de, 274
Taylor-Young (formule de), 258, 270
Tchebychev (inégalité de), 33
transcendant (nombre), 25

U

universal chord theorem, 229

V

valeur d'adhérence d'une suite, 74–81
valeurs intermédiaires (théorème des), 228–232, 254, 259, 268

W

Whitney (théorème de), 282

IMPRIMÉ EN GRANDE-BRETAGNE
PAR CAMBRIDGE UNIVERSITY PRESS
DÉPÔT LÉGAL NOVEMBRE 2007

Le recueil d'exercices résolus des oraux des Écoles normales supérieures et de l'École polytechnique de Serge Francinou, Hervé Gianella et Serge Nicolas comprendra six volumes : trois consacrés à l'algèbre et trois à l'analyse.

Le présent volume est consacré aux bases de l'analyse : nombres réels et complexes, suites, séries, fonctions d'une variable réelle. Hormis les séries, ces sujets sont traités dès la première année de classes préparatoires.

Les auteurs se sont attachés à dégager les idées qui se trouvent à la source des solutions fournies, sans pour autant omettre le détail des vérifications et des calculs. À côté d'exercices très classiques, le lecteur trouvera des développements sur des questions plus originales, notamment dans le chapitre 1. Un soin tout particulier a été apporté au texte de présentation qui accompagne les exercices, groupés par thèmes. La présentation historique qui ponctue la succession des énoncés montrera au lecteur que l'élaboration des concepts de l'Analyse – qui apparaît aujourd'hui comme un édifice achevé – n'a pas été sans erreurs, hésitations, retours en arrière. D'autre part, certains points du programme parfois négligés par les candidats font l'objet d'utiles rappels.

Ce livre s'adresse naturellement aux élèves des classes préparatoires, de première et de deuxième année, mais il sera également très utile aux candidats à l'agrégation qui y trouveront de nombreux développements pour leur oral. Ces exercices constituent aussi un excellent complément à la préparation à l'écrit du CAPES.

Dans cette seconde édition, qui voit le jour alors que les deux derniers volumes (Algèbre 3 et Analyse 3) restent à paraître, figurent 26 exercices nouveaux.

Collection enseignement des mathématiques

ISBN 978-2-84225-135-2

Graphisme : Massin

exercices de mathématiques

oraux x-ens

analyse 2

Serge Francinou
Hervé Gianella
Serge Nicolas

CASSINI

EXERCICES DE MATHÉMATIQUES
ORAUX X-ENS

Enseignement des mathématiques

1. J.-Y. Ouvrard, *Probabilités I*
2. J. Hubbard, B. West, *Équations différentielles et systèmes dynamiques*
3. M. Cottrell, V. Genon-Catalot, Ch. Duhamel, Th. Meyre, *Exercices de probabilités*
4. F. Rouvière, *Petit guide de calcul différentiel à l'usage de la licence et de l'agrégation*
5. J.-Y. Ouvrard, *Probabilités II*
6. G. Zémor, *Cours de cryptographie*
7. A. Szpirglas, *Exercices d'algèbre*
8. B. Perrin-Riou, *Algèbre, arithmétique et Maple*
10. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 1*
11. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 1*
12. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 2*
13. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Analyse 2*
14. S. Francinou, H. Gianella, S. Nicolas, *Exercices des oraux X-ENS, Algèbre 3*
15. H. Krivine, *Exercices de mathématiques pour physiciens*
16. J. Jacod, Ph. Protter, *L'essentiel en théorie des probabilités*
17. M. Willem, *Analyse fonctionnelle élémentaire*
18. É. Amar, É. Matheron, *Analyse complexe*
19. B. Randé, *Problèmes corrigés. Concours 2002 et 2003 (MP)*
20. D. Perrin, *Mathématiques d'école*
21. B. Randé, *Problèmes corrigés. Concours 2004 (MP)*
22. P. Bourgade, *Olympiades internationales de mathématiques 1976-2005*
23. V. Prasolov, *Problèmes et théorèmes d'algèbre linéaire*

SERGE FRANCINOU
HERVÉ GIANELLA
SERGE NICOLAS

Exercices de mathématiques des oraux de l'École polytechnique et des Écoles normales supérieures

Analyse. Tome II

Deuxième édition, revue et augmentée

CASSINI

SERGE FRANCINOU, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Marcelin Berthelot.

HERVÉ GIANELLA, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Saint-Louis.

SERGE NICOLAS, ancien élève de l'École normale supérieure et agrégé de Mathématiques est actuellement professeur en classe préparatoire au lycée Henri IV.

2e édition
ISBN 978-2-84225-141-3
© Cassini, Paris, 2009.

(1ère édition. 2004, ISBN-10 : 2-84225-033-8, ISBN-13 : 978-2-84225-033-I)

Table des matières

Introduction 1

Chapitre 1. Intégration 5

1.1. Un calcul d'intégrale 6
1.2. Une inégalité (1) 7
1.3. Une inégalité (2) 7
1.4. Inégalité de Young 8
1.5. Cas d'égalité dans le théorème de majoration 12
1.6. Une inégalité (3) 13
1.7. Inégalité de Van der Corput 15
1.8. Majoration de l'erreur dans la méthode des trapèzes 17
1.9. Méthode des trapèzes pour une fonction convexe 19
1.10. Recherche d'un minimum 21
1.11. Recherche d'une borne inférieure 22
1.12. Inégalité de Hardy 24
1.13. Recherche d'un minimum 25
1.14. Étude d'une suite 26
1.15. Inégalité de Wirtinger 27
1.16. Fonctions dont les deux premiers moments sont nuls 29
1.17. Fonctions dont les n premiers moments sont nuls 30
1.18. Recherche d'orthogonal 31
1.19. Méthode de Gauss 32
1.20. Une somme de Riemann (1) 34
1.21. Une somme de Riemann (2) 35
1.22. Étude asymptotique des sommes de Riemann 36
1.23. Découpage en morceaux de même aire 38
1.24. Estimation d'une suite de coefficients binomiaux 39
1.25. Lemme de Riemann-Lebesgue 41
1.26. Deuxième formule de la moyenne 43
1.27. Suites équiréparties : généralités 45
1.28. Suites équiréparties : critère de Weyl 47
1.29. Suites équiréparties : un exemple 50
1.30. Calcul d'une intégrale 50
1.31. Mesure du domaine d'annulation d'une fonction 51
1.32. Limites de normes intégrales 53
1.33. Équivalent d'une primitive 54
1.34. Recherche de limite 55
1.35. Équivalent des intégrales de Wallis 56

1.36. Recherche d'un équivalent (1) 57
1.37. Recherche d'un équivalent (2) 58
1.38. Recherche d'un équivalent (3) 60
1.39. Développement asymptotique (1) 62
1.40. Développement asymptotique (2) 64
1.41. Méthode de Laplace (1) 66
1.42. Méthode de Laplace (2) 69
1.43. Développement en série entière 73
1.44. Lemme de Gronwall (1) 76
1.45. Lemme de Gronwall (2) 76
1.46. Moyenne spatiale 77
1.47. Une équation intégrale (1) 78
1.48. Une équation intégrale (2) 79
1.49. Une équation intégrale (3) 80
1.50. Une équation intégrale (4) 81
1.51. Indice d'une courbe fermée 83
1.52. Une preuve du théorème de d'Alembert-Gauss 87

Chapitre 2. Suites et séries de fonctions **91**

2.1. Majoration sur une demi-droite 92
2.2. Une décomposition en série de la valeur absolue 94
2.3. Fonctions continues presque additives 96
2.4. Contrôle uniforme de séries alternées 97
2.5. Sommation au sens de Riemann 99
2.6. Convergence uniforme des séries de Dirichlet 102
2.7. Convergence d'une série trigonométrique 104
2.8. Étude de la convergence d'une série de fonctions 106
2.9. Suite ne convergeant uniformément sur aucun ouvert 108
2.10. Continuité et convergence de fonctions 110
2.11. Série normalement convergente de somme non dérivable . . 111
2.12. Régularité d'une série trigonométrique 112
2.13. Série de primitives successives 113
2.14. Équivalent d'une série de fonctions 114
2.15. Courbe de Bolzano 117
2.16. Interversion série-intégrale 121
2.17. Sur le théorème d'intégration d'une série de fonctions . . . 123
2.18. Polynômes de Bernstein 125
2.19. Limites uniformes de polynômes strictement croissants . . . 128
2.20. Un théorème de Walsh 130
2.21. Théorème de Chudnovsky 131
2.22. Algèbre des fonctions presque périodiques 133
2.23. Théorème de Korovkin (1953) 136
2.24. Approximation de Laguerre 138

2.25. Convergence d'une suite de polynômes 140
2.26. Développement eulérien de la cotangente 142
2.27. Produits infinis (1) 146
2.28. Produits infinis (2) 148
2.29. Développement eulérien du sinus sur $\mathbb{C}$ 152
2.30. Théorème de Dini 156
2.31. Un théorème de point fixe 158
2.32. Suite de fonctions lipschitziennes sur un compact 160
2.33. Convergence uniforme de suites de fonctions convexes . . . 161
2.34. Théorème d'Ascoli 162
2.35. Un cas particulier du théorème d'Ascoli 165
2.36. Théorème de sélection de Helly 166
2.37. Suite de fonctions d'un sous-espace de dimension finie de $\mathcal{C}^0([0,1])$ 167
2.38. Critère de convergence uniforme 169

Chapitre 3. Séries entières **171**

3.1. Rayon de convergence (1) 172
3.2. Rayon de convergence (2) 172
3.3. Rayon de convergence (3) 173
3.4. Rayon de convergence (4) 174
3.5. Étude d'une série entière sur le cercle de convergence 175
3.6. Le théorème d'Abel-Dirichlet 178
3.7. Calcul de la somme d'une série entière 181
3.8. Calcul de la somme de séries de fonctions 182
3.9. Somme d'une série entière extraite 184
3.10. Calcul d'une somme 186
3.11. Calculs de sommes de séries alternées 187
3.12. Développement en série entière des fractions rationnelles . 190
3.13. Signe des coefficients d'un développement en série entière . 192
3.14. Séries génératrices de deux suites récurrentes 195
3.15. Partitions d'un entier en parts fixées 197
3.16. Série génératrice du nombre de partitions de n 199
3.17. Signe des dérivées successives d'une fonction 201
3.18. Développement en série entière (1) 203
3.19. Développement en série entière (2) 204
3.20. Développement en série entière (3) 206
3.21. Limite simple de polynômes à coefficients positifs 208
3.22. Recherche d'un équivalent 210
3.23. Étude asymptotique au bord du disque de convergence . . 212
3.24. Un théorème de Gauss 214
3.25. Série de Lambert 217
3.26. Théorèmes taubériens 220

3.27. Condition suffisante d'injectivité d'une série entière 226
3.28. Théorème de Bieberbach dans le cas réel 227
3.29. Série de Laurent à valeurs entières 229
3.30. Caractérisation des fonctions réelles analytiques 231
3.31. Étude d'analycité 233
3.32. Un théorème de Bernstein 236
3.33. Principe des zéros isolés 237
3.34. Nullité sur un arc du cercle de convergence 238
3.35. Série entière à coefficients entiers, bornée sur son disque de convergence 239
3.36. Fonctions entières de partie réelle bornée 240
3.37. Fonction entière dominée par une exponentielle 242
3.38. Relation entre $\sqrt[n]{|a_n|}$ et $\dfrac{\ln|f(z)|}{|z|}$ 243
3.39. Principe du maximum pour une fraction rationnelle 245
3.40. Principe du maximum 247
3.41. Extrema d'une fonction entière 250
3.42. Critère de convergence simple d'une suite de séries entières 252
3.43. Complétude de l'espace des fonctions continues sur $\overline{\mathrm{D}}$, développables en série entière sur D 255
3.44. Logarithme d'une fonction entière qui ne s'annule pas . . . 258
3.45. Un théorème de Fejér. 260
3.46. Fonction entière telle que $f(2^n) = (-1)^n$ 266

Chapitre 4. Séries de Fourier **271**

4.1. Calcul d'intégrales 274
4.2. Majoration de la norme de la dérivation 275
4.3. Inégalité de Bernstein (1912) 278
4.4. Fonction 2π-périodique orthogonale aux polynômes trigonométriques de degré $< n$ 280
4.5. Une inégalité 281
4.6. Inégalité de Hilbert (1) 282
4.7. Inégalité de Hilbert (2) 283
4.8. Calcul de sommes de séries 286
4.9. Développement eulérien de la fonction cotangente 289
4.10. Développement en série de Fourier (1) 291
4.11. Calcul d'intégrales par développement en série de Fourier . 293
4.12. Développement en série de Fourier (2) 294
4.13. Calcul de la somme d'une série trigonométrique 297
4.14. Inégalité entre normes 300
4.15. Recherche d'équivalent 301
4.16. Formule sommatoire de Poisson 304
4.17. Un cas particulier de la formule sommatoire de Poisson . . 306

4.18. Série génératrice des nombres de Bernoulli 308
4.19. Série de Fourier d'une primitive 311
4.20. Inégalité de Wirtinger 313
4.21. Solutions périodiques de $x' = \mathrm{F}(x)$ 315
4.22. Résolution d'une équation différentielle 319
4.23. Étude de convergence uniforme 323
4.24. Condition suffisante d'analycité 328
4.25. Le phénomène de Gibbs 329
4.26. Théorème de Weierstrass trigonométrique 332
4.27. Théorème de Jackson (1911) 334
4.28. Un théorème de Bernstein (1912) 338
4.29. Noyau de Poisson 340
4.30. Fonctions à coefficients de Fourier positifs 343
4.31. Théorème de Fejér (1904) 345
4.32. Complétude pour la norme quadratique 348
4.33. Fonctions dont la série des coefficients de Fourier converge absolument 349
4.34. Convergence de la série de Fourier d'une fonction α-höldérienne 352
4.35. Dual topologique 355
4.36. Opérateurs de convolution 357

Table des matières **363**

Index **369**

Introduction

Cet ouvrage est le second tome d'analyse d'un recueil d'exercices de mathématiques destiné à la préparation des oraux des concours d'entrée aux Écoles normales supérieures et à l'École polytechnique. Il comportera six tomes, trois d'algèbre et trois d'analyse.

La vocation première des Écoles normales est de former des chercheurs ou des enseignants-chercheurs. Le concours d'entrée vise donc à détecter les qualités scientifiques du candidat, son aptitude à la recherche. À l'oral, on jugera avant tout la capacité de prendre des initiatives, d'utiliser une indication, de mener à bien une démarche. On ne sera pas surpris que les exercices posés aient un contenu mathématique riche, qu'ils soient très éloignés du simple exercice technique, d'application du cours, qu'ils soient souvent difficiles. Ils visent la plupart du temps à la démonstration d'un résultat mathématique significatif. Ils pourraient apparaître excessivement difficiles, si on perdait de vue le déroulement concret de l'épreuve. L'oral des ENS est un long dialogue (l'épreuve dure environ cinquante minutes, comme d'ailleurs à l'École polytechnique) entre le candidat et l'examinateur, qui tout au long de l'épreuve fournit des indications, quand c'est nécessaire, pour relancer la réflexion du candidat et tester ses réactions. Il est d'ailleurs impossible de rendre pleinement compte dans un recueil d'exercices du caractère oral de l'épreuve.

L'École polytechnique, quant à elle, est plus généraliste. Les exercices posés au concours sont de facture plus classique et, en règle générale, l'examinateur intervient moins. C'est au candidat de montrer sa maîtrise du programme dans la résolution d'un exercice dont la difficulté est cependant très variable. Certains sont proches des exercices d'ENS. Les énoncés circulent d'ailleurs d'un concours à l'autre, ou peuvent même être repris d'exercices d'Olympiades.

Les énoncés qui figurent dans ce recueil ont été donnés entre 1996 et 2008. Ils sont extraits pour l'essentiel des listes publiées chaque année par la RMS (*Revue des mathématiques de l'enseignement supérieur* aux éditions Vuibert jusqu'en 2003 et désormais *Revue de la filière Mathématiques* aux éditions e.net) dont nous remercions les auteurs pour l'aide précieuse qu'ils apportent ainsi aux élèves et aux professeurs des classes préparatoires. Il s'agit de versions communiquées par les étudiants, reflétant la compréhension que ceux-ci ont eue de l'exercice et le déroulement conjoncturel de leur oral, comme le montrent les variations d'une année à l'autre pour un même exercice. Nous n'avons pas hésité à les modifier, pour rectifier des erreurs, compléter un énoncé

quand manifestement l'exercice s'est arrêté avant que le résultat que l'examinateur avait en vue ne soit atteint, ou ajouter des indications.

Nous avons choisi de laisser quelques énoncés « bruts », ceux pour lesquels nous estimons qu'une démarche naturelle (qui peut être longue et ardue) permet de conduire à la solution. Pour d'autres exercices, nous avons pris la liberté de rajouter des questions intermédiaires, qui auraient pu être celles posées par l'examinateur. Quitte à perdre en concision, nous avons tenu à rédiger les solutions les plus pédagogiques possible, essayant d'exposer clairement les idées et démarches des raisonnements sans pour autant escamoter les détails ou calculs qui peuvent paraître évidents. On évite autant que possible l'introduction d'une astuce ou d'un objet *ad hoc* permettant d'atteindre rapidement la solution. S'il n'y a pas moyen d'expliquer l'origine de cette astuce, c'est que l'exercice est peu intéressant et que l'étudiant en tirera peu de profit.

À l'intérieur de chaque chapitre, les exercices ont été regroupés thématiquement, et à l'intérieur de chaque thème, souvent par ordre de difficulté croissante. Ainsi regroupés, ils apparaîtront plus accessibles, car plongés dans leur contexte mathématique, éclairés par d'autres exercices voisins. Les introductions historiques qui ouvrent chaque chapitre, outre leur intérêt propre, visent au même but. Enfin, nous avons agrémenté les énoncés de quelques remarques préliminaires. Sans faire de rappels de cours systématiques, nous avons énoncé, voire redémontré certains résultats : lemmes classiques, intervenant dans la résolution d'un grand nombre d'exercices, ou résultats au contraire à la lisière du programme, mais utiles, pour lesquels des éclaircissements étaient nécessaires. On trouvera aussi des remarques de synthèse ou des généralisations qui, nous l'espérons, pourront amener le candidat curieux à approfondir ses connaissances. Les quelques indications bibliographiques ont le même objectif.

Le lecteur ne tirera profit de ce livre d'exercices que s'il cherche des solutions personnelles avant d'en étudier les corrigés. Une bonne connaissance du cours est indispensable. En effet, les théorèmes du programme fournissent bon nombre de schémas de démonstration. Rappelons aussi quelques démarches générales qui peuvent faciliter l'appréhension des exercices difficiles :

▷ ne pas hésiter à faire un calcul formel, par exemple avec des séries entières, pour se faire une idée du résultat et justifier ensuite la convergence.

▷ pour des questions asymptotiques concernant les intégrales il faut souvent essayer de repérer le (où les) endroit(s) où la contribution de la fonction est la plus importante et/ou remplacer la fonction intégrée par une fonction qui lui est équivalente et qui conduit à une intégrale calculable.

▷ ne pas perdre de vue les liens étroits qu'il y a entre les séries de Fourier et les séries entières. L'une des deux théories peut servir à résoudre un exercice posé *a priori* dans l'autre.

Au-delà des étudiants en classe préparatoire, ces ouvrages intéresseront aussi les candidats au CAPES et à l'Agrégation, qui y trouveront matière à réviser les principales notions du programme, ainsi que des exemples pour nourrir un développement pour leur oral. Leur connaissance de l'Analyse Complexe pourra éclairer ou raccourcir certaines solutions du chapitre sur les séries entières.

Voyons maintenant plus précisément le contenu de ce tome 2 d'analyse. Le premier chapitre, un peu à part, concerne l'intégrale sur un segment. C'est le seul qui peut être étudié en première année. Nous avons choisi de regrouper les exercices concernant l'intégrale généralisée dans le troisième tome d'analyse. Il est toutefois possible, à l'occasion d'un exercice, de rencontrer le théorème de convergence dominée. Les trois autres chapitres sont centrés sur les séries, et concernent des notions au cœur du programme de seconde année : suites et séries de fonctions, séries entières et séries de Fourier.

Comme dans les autres tomes, les exercices sont classés par thème. La difficulté est toutefois plutôt croissante : les chapitres commencent par des questions techniques ou des savoir-faire indispensables (calculs d'intégrales, de rayon de convergence, de développement en série de Fourier,...) et se terminent souvent par des exercices difficiles qui ont pour objet de démontrer des théorèmes du niveau licence ou maîtrise (citons par exemple le théorème d'Ascoli, le théorème de Fejér,...).

Le troisième et dernier tome d'analyse portera sur l'intégrale généralisée, la topologie, les équations différentielles et les fonctions de plusieurs variables.

Nous remercions André et Catherine Bellaïche, René Cori, Rached Mneimné, Bernard Randé, Yves Duval et Joon Kwon pour leur relecture approfondie de l'ouvrage et leurs nombreuses suggestions, tant sur le fond que sur la forme.

Enfin, si vous souhaitez nous contacter pour nous faire part de vos remarques, vous pouvez envoyer un courrier à l'adresse *fgn.cassini@free.fr*.

Chapitre 1

Intégration

À partir de la deuxième moitié du XVII^{e} *siècle et de la mise en place du calcul infinitésimal, s'impose le concept d'intégration, traité comme le concept inverse de celui de dérivation. Leibniz donne une méthode permettant, en principe, le calcul des primitives d'une fonction rationnelle. Euler montre comment ramener à ce calcul celui de nombreuses primitives comme* $\int \mathrm{R}(x, \sqrt{\mathrm{P}(x)})\,dx$, *où* R *est une fraction rationnelle et* P *un polynôme de degré inférieur ou égal à* 2. *Il développe aussi la théorie des intégrales doubles. Vers 1730, il introduit la fonction* Γ *comme fonction interpolatrice des factorielles, sous forme de produit infini puis sous forme d'intégrale. Il commence l'étude des intégrales elliptiques dont Lagrange introduit la forme générale (1784). L'étude en est approfondie par Legendre, puis Abel et Jacobi (vers 1830). Laplace est le premier à démontrer que* $\displaystyle\int_{-\infty}^{+\infty} e^{-t^2}\,dt = \sqrt{\pi}$. *En liaison avec le calcul des probabilités, vers 1778, il obtient des formules asymptotiques pour les intégrales du type* $\displaystyle\int_a^b g(t)h(t)^n\,dt$ *lorsque* n *tend vers* $+\infty$. *Il considère le point* c *où* h *atteint son maximum et utilise les développements de Taylor de* g *et* h *au point* c *(méthode de Laplace). En 1749, Euler pose le problème de développer une fonction* 2π*-périodique arbitraire sous la forme* $\displaystyle\sum_{k=0}^{+\infty} a_k \cos kx + b_k \sin kx$, *et note les relations d'orthogonalité* $\displaystyle\int_0^{\pi} \cos mx \cos nx\,dx = 0$. *Clairaut donne la forme générale des coefficients.*

Cauchy, vers 1820, définit l'intégrale d'une fonction continue sur un segment comme la limite quand $\max(x_i - x_{i-1})$ *tend vers* 0 *de* $\displaystyle\sum_{i=1}^{n}(x_i - x_{i-1})f(x_i)$, *en utilisant implicitement la continuité uniforme de la fonction. Il faudra attendre 1872 pour que Heine introduise la notion de continuité uniforme et montre qu'une fonction définie sur un segment y est uniformément continue. En 1854, dans un mémoire intitulé «* La possibilité de représenter une fonction par une série trigonométrique *», Riemann pose les bases de ce qui allait devenir l'intégrale de Riemann. Il définit l'intégrale d'une fonction bornée sur un segment comme la limite de ses sommes de Riemann* $\displaystyle\sum_{i=1}^{n}(x_i - x_{i-1})f(\lambda_i)$ *où* λ_i *est un point*

quelconque de $[x_{i-1}, x_i]$. Il démontre un critère d'intégrabilité (qui en langage moderne se ramène à : l'ensemble des points de discontinuité de la fonction est de mesure nulle), montre que les fonctions monotones sont intégrables, donne un exemple de fonction monotone ayant une infinité de discontinuités et enfin, revenant aux séries trigonométriques, montre que toute fonction intégrable a des coefficients de Fourier qui tendent vers 0. Vers 1860, Weierstrass montre que si une série de fonctions continues converge uniformément sur $[a, b]$, on peut intégrer cette série terme à terme, opération que l'on ne s'était pas privé de faire depuis au moins un siècle et demi sans aucune justification.

L'intégrale de Riemann domine jusqu'aux travaux de Lebesgue. Celui-ci complète les travaux de Borel sur la théorie de la mesure et construit ce qui s'appelle depuis la mesure de Lebesgue sur $\mathbb{R}^n$. Pour définir l'intégrale il remplace les fonctions en escalier utilisées par Riemann par des fonctions étagées (fonctions dont l'image est finie mais où l'image réciproque de chaque point est un ensemble mesurable quelconque et pas forcément un intervalle). Lebesgue montre immédiatement la fécondité de ses conceptions par les théorèmes fondamentaux de passage à la limite pour les suites de fonctions. puis par le renouvellement qu'il apporte à la théorie des séries de Fourier (1903).

1.1. Un calcul d'intégrale

Calculer $\displaystyle\int_0^{\pi/4} \ln(1+\tan x)\mathrm{d}x$.

(École polytechnique)

▷ **Solution.**

Notons I l'intégrale à calculer. Il s'agit d'un exemple classique, et il n'y en a pas beaucoup, où la primitive de la fonction intégrée n'est pas calculable, mais où il est possible de s'en sortir par une petite astuce. On remplace $\tan x$ par $\dfrac{\sin x}{\cos x}$. On obtient

$$\mathrm{I} = \int_0^{\pi/4} \ln(\sin x + \cos x)\mathrm{d}x - \int_0^{\pi/4} \ln\cos x\,\mathrm{d}x.$$

On a $\sin x + \cos x = \sqrt{2}\cos\left(x - \dfrac{\pi}{4}\right)$, de sorte que

$$\int_0^{\pi/4} \ln(\sin x + \cos x)\mathrm{d}x = \frac{\pi}{4}\ln\sqrt{2} + \int_0^{\pi/4} \ln\cos\left(x - \frac{\pi}{4}\right)\mathrm{d}x.$$

Or, par le changement de variable $u = \dfrac{\pi}{4} - x$, on a

$$\int_0^{\pi/4} \ln\cos\left(x-\frac{\pi}{4}\right)\mathrm{d}x = \int_0^{\pi/4} \ln\cos u\,\mathrm{d}u.$$

On obtient donc finalement

$$\boxed{\mathrm{I} = \frac{\pi}{4}\ln\sqrt{2} = \frac{\pi}{8}\ln 2}. \quad \triangleleft$$

Les prochains exercices sont consacrés à des inégalités intégrales. Le sujet est vaste et les techniques employées très variées. L'inégalité qui suit est une simple utilisation de la positivité.

1.2. Une inégalité (1)

Soit $f : [0,1] \longrightarrow \mathbb{R}$ continue et telle que $\displaystyle\int_0^1 f(t)\mathrm{d}t = 0$. On pose $\alpha = \min_{x\in[0,1]} f(x)$ et $\beta = \max_{x\in[0,1]} f(x)$. Montrer que $\displaystyle\int_0^1 f(t)^2\mathrm{d}t \leqslant -\alpha\beta$.

(École polytechnique)

▷ **Solution.**

Les fonctions $f-\alpha$ et $\beta - f$ sont positives, si bien que

$$0 \leqslant \int_0^1 (f-\alpha)(\beta-f) = (\beta+\alpha)\int_0^1 f - \alpha\beta - \int_0^1 f^2 = -\alpha\beta - \int_0^1 f^2,$$

et donc

$$\boxed{\int_0^1 f^2 \leqslant -\alpha\beta}. \quad \triangleleft$$

1.3. Une inégalité (2)

Soit $f : [0,1] \longrightarrow \mathbb{R}$ une fonction concave et continue telle que $f(0) = 1$. Montrer que

$$\int_0^1 xf(x)\mathrm{d}x \leqslant \frac{2}{3}\left(\int_0^1 f(x)\mathrm{d}x\right)^2.$$

(École polytechnique)

▷ **Solution.**

Posons $A = \int_0^1 f(x)\,dx$ et $B = \int_0^1 xf(x)\,dx$. Comme f est continue, $F : x \longmapsto \int_0^x f$ est une primitive de f et une intégration par parties donne

$$B = \Big[xF(x)\Big]_0^1 - \int_0^1 F(x)\,dx = A - \int_0^1 \left(\int_0^x f(t)\,dt\right)dx.$$

Soit $x > 0$. Pour tout $t \in [0, x]$ on a, par concavité de f,

$$f(t) \geqslant \frac{t}{x}f(x) + \frac{x-t}{x}f(0) = \frac{f(x)-1}{x}t + 1,$$

et en intégrant, il vient

$$\int_0^x f(t)\,dt \geqslant \frac{f(x)-1}{x}\frac{x^2}{2} + x = \frac{xf(x)+x}{2}.$$

Ainsi,

$$A - B = \int_0^1 \left(\int_0^x f(t)\,dt\right)dx \geqslant \int_0^1 \frac{xf(x)+x}{2}dx$$

et on obtient $A - B \geqslant \dfrac{B}{2} + \dfrac{1}{4}$, ou encore $B \leqslant \dfrac{2}{3}\left(A - \dfrac{1}{4}\right)$. Or, on a $A - \dfrac{1}{4} \leqslant A^2$ car $(2A-1)^2 \geqslant 0$, si bien que $\boxed{B \leqslant \dfrac{2}{3}A^2}$. ◁

1.4. Inégalité de Young

1. Soit $f : \mathbb{R}_+ \to \mathbb{R}_+$ une fonction de classe $\mathcal{C}^1$ telle que $f(0) = 0$, $f' > 0$ et $\lim\limits_{x\to+\infty} f(x) = +\infty$. Montrer que pour $a \geqslant 0$,

$$af(a) = \int_0^a f(t)\,dt + \int_0^{f(a)} f^{-1}(t)\,dt,$$

puis que, pour $b \geqslant 0$,

$$ab \leqslant \int_0^a f(t)\,dt + \int_0^b f^{-1}(t)\,dt$$

(*inégalité de Young*).

2. Établir que l'inégalité de la question précédente reste valable si on suppose seulement f continue strictement croissante avec $f(0) = 0$ et $\lim\limits_{x\to+\infty} f(x) = +\infty$.

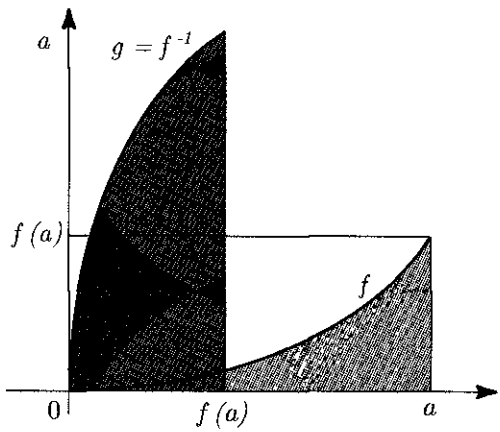
(École polytechnique)

▷ **Solution.**

1. Posons $\mathrm{F}(a) = \int_0^a f(t)\,\mathrm{d}t + \int_0^{f(a)} f^{-1}(t)\,\mathrm{d}t - af(a)$. Par hypothèse, f établit un $\mathcal{C}^1$-difféomorphisme de $\mathbb{R}_+$ sur lui-même. Il en résulte que f^{-1} est de classe $\mathcal{C}^1$ et en particulier continue, et donc que F est de classe $\mathcal{C}^1$ sur $\mathbb{R}_+$. On a pour tout $a \geqslant 0$,

$$\mathrm{F}'(a) = f(a) + f^{-1}(f(a)) \times f'(a) - f(a) - af'(a) = 0.$$

Comme $\mathrm{F}' = 0$, F est constante et comme $\mathrm{F}(0) = 0$, elle est identiquement nulle. Cela montre la première égalité qui devient très claire si on interprète les intégrales en terme d'aires : la somme des deux aires grisées de la figure est égale à l'aire du rectangle.



Soient a, b deux réels positifs. Pour montrer l'inégalité, supposons d'abord $b \geqslant f(a)$. On a d'après l'identité précédente,

$$\int_0^a f + \int_0^b f^{-1} - ab = af(a) + \int_{f(a)}^b f^{-1} - ab = \int_{f(a)}^b (f^{-1} - a) \geqslant 0$$

car f^{-1} étant croissante, elle est minorée par a sur l'intervalle $[f(a), b]$. Lorsque $b \leqslant f(a)$ on fait le même travail mais en faisant jouer à f^{-1} le rôle de f. On écrit

$$\begin{aligned}\int_0^a f + \int_0^b f^{-1} - ab &= \int_0^{f^{-1}(b)} f + \int_{f^{-1}(b)}^a f + \int_0^b f^{-1} - ab \\ &= \int_{f^{-1}(b)}^a f + bf^{-1}(b) - ab = \int_{f^{-1}(b)}^a (f - b) \geqslant 0.\end{aligned}$$

2. La preuve de l'inégalité de Young à partir de l'égalité obtenue dans la première question n'utilise pas du tout le caractère dérivable de

f. Il nous suffit donc de prouver que l'égalité

$$af(a) = \int_0^a f(t)\,\mathrm{d}t + \int_0^{f(a)} f^{-1}(t)\,\mathrm{d}t$$

est encore vraie avec f seulement continue et bien entendu strictement monotone avec $f(0) = 0$ et $\lim_{x\to+\infty} f(x) = +\infty$. Notons que ces hypothèses permettent de dire que f est un homéomorphisme strictement croissant de $\mathbb{R}_+$ sur lui-même. Cela justifie l'existence de f^{-1} ainsi que sa continuité.

On va prouver cette égalité de plusieurs manières. Soit $a \geqslant 0$.

• La vision de l'égalité en terme d'aires donnée ci-dessus fait bien sentir que la dérivabilité de f n'est pas utile. Considérons l'ensemble $\mathrm{A} = \{(x,y),\ 0 \leqslant x \leqslant a,\ 0 \leqslant y \leqslant f(x)\}$. L'aire de A vaut $\int_0^a f(x)\mathrm{d}x$. La transformation $s : (x,y) \mapsto (y,x)$ conserve les aires (c'est simplement la symétrie par rapport à la droite d'équation $y = x$). L'aire de A est donc égale à l'aire de son image $\mathrm{A}' = s(\mathrm{A})$. Or,

$$\begin{aligned}\mathrm{A}' &= \{(x,y),\ 0 \leqslant y \leqslant a,\ 0 \leqslant x \leqslant f(y)\}\\ &= \{(x,y),\ 0 \leqslant y \leqslant a,\ 0 \leqslant f^{-1}(x) \leqslant y\}\\ &= \{(x,y),\ 0 \leqslant x \leqslant f(a).\ f^{-1}(x) \leqslant y \leqslant a\}.\end{aligned}$$

Ainsi, l'aire de A′ vaut $\int_0^{f(a)} (a - f^{-1}(x))\,\mathrm{d}x = af(a) - \int_0^{f(a)} f^{-1}(x)\,\mathrm{d}x$. L'égalité des deux aires fournit le résultat.

• Voici une autre solution purement analytique. Il suffit de prouver que la fonction F introduite dans la question 1 est toujours identiquement nulle avec la seule hypothèse f continue. Pour cela, on va montrer que F est dérivable de dérivée nulle, mais en revenant à la définition de la dérivabilité. Soit $a \geqslant 0$. Le taux d'accroissement $\dfrac{\mathrm{F}(a+h) - \mathrm{F}(a)}{h}$ vaut

$$\frac{1}{h}\left(\int_a^{a+h} f + \int_{f(a)}^{f(a+h)} f^{-1} - (a+h)f(a+h) + af(a)\right).$$

Le rapport $\dfrac{1}{h}\int_a^{a+h} f$ tend vers $f(a)$ lorsque h tend vers 0 car f est continue. Le terme $f(a+h)$ tend également vers $f(a)$ par continuité de f. Il reste donc à prouver que

$$\psi(h) = \frac{1}{h}\int_{f(a)}^{f(a+h)} f^{-1} - a\frac{f(a+h) - f(a)}{h} = \frac{1}{h}\int_{f(a)}^{f(a+h)} (f^{-1} - a)$$

tend vers 0 lorsque h tend vers 0. Comme f^{-1} est continue, la première formule de la moyenne assure l'existence d'un réel c_h compris entre $f(a)$

et $f(a+h)$ tel que

$$\psi(h) = \frac{(f(a+h)-f(a))(f^{-1}(c_h)-a)}{h}.$$

Par croissance de f^{-1}, le rapport $\dfrac{f^{-1}(c_h)-a}{h}$ est dans $[0,1]$, et $\psi(h)$ tend donc vers 0 quand h tend vers 0. Ainsi, F est dérivable en a avec $\mathrm{F}'(a)=0$. On peut alors conclure comme dans la première question.

• Pour terminer, nous proposons une troisième démonstration de l'égalité de Young qui fait appel aux sommes de Riemann : prenons une subdivision régulière $0 = x_0 < x_1 < \cdots < x_n = a$ de l'intervalle $[0,a]$. Posons $y_i = f(x_i)$ pour $0 \leqslant i \leqslant n$. Comme f est continue, la quantité $\sum_{i=1}^{n} y_i(x_i - x_{i-1})$ converge vers $\int_0^a f$ lorsque n tend vers l'infini. Or, nous avons

$$\begin{aligned}\sum_{i=1}^{n} y_i(x_i - x_{i-1}) &= \sum_{i=1}^{n} y_i x_i - \sum_{i=0}^{n-1} y_{i+1}x_i = y_n x_n - \sum_{i=0}^{n}(y_i - y_{i+1})x_i \\ &= af(a) - \sum_{i=0}^{n}(y_i - y_{i+1})x_i. \qquad (*)\end{aligned}$$

Comme f^{-1} est également continue, la dernière somme est une somme de Riemann qui va converger vers $\int_0^{f(a)} f^{-1}$ sous réserve que le pas de la subdivision des y_i converge aussi vers 0. C'est bien le cas : en effet, d'après le théorème de Heine, f est uniformément continue sur le compact $[0,a]$ et si $\varepsilon > 0$ est donné, il existe $\eta > 0$ tel que si $|x-y| \leqslant \eta$, on a $|f(x)-f(y)| \leqslant \varepsilon$. Pour n suffisamment grand, $\dfrac{a}{n} \leqslant \eta$ et pour tout $i \in [\![1,n]\!]$, $x_i - x_{i-1} = \dfrac{a}{n}$ et $0 \leqslant y_i - y_{i-1} \leqslant \varepsilon$. Le pas de la subdivision des y_i est donc inférieur ou égal à ε. En passant à la limite dans $(*)$, il vient

$$\int_0^a f(x)\mathrm{d}x = af(a) - \int_0^{f(a)} f^{-1}(x)\mathrm{d}x. \quad \lhd$$

En appliquant l'inégalité de Young avec la fonction $f(t) = t^{p-1}$ pour $p > 1$ on obtient l'inégalité de Hölder $ab \leqslant \dfrac{a^p}{p} + \dfrac{b^q}{q}$ où q est l'exposant conjugué de p, i.e. l'unique réel tel que $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

Le théorème de majoration dit que $\left|\int_a^b f\right| \leqslant \int_a^b |f|$. Il peut se voir comme une version continue de l'inégalité triangulaire. Lorsque f est continue à valeurs réelles, il est aisé de voir que l'inégalité ci-dessus est

une égalité si et seulement si f garde un signe constant[1]*. L'exercice suivant étudie le cas d'égalité dans $\mathbb{R}^n$, la norme euclidienne remplaçant la valeur absolue.*

1.5. Cas d'égalité dans le théorème de majoration

Soit $f : [a,b] \longrightarrow \mathbb{R}^n$ continue, $\mathbb{R}^n$ étant muni de sa structure euclidienne canonique. Donner une condition nécessaire et suffisante pour que

$$\left\| \int_a^b f \right\| = \int_a^b \|f\|.$$

(École polytechnique)

▷ **Solution.**

Montrons que la condition cherchée est que f prenne ses valeurs dans une demi-droite issue de l'origine, autrement dit qu'il existe un vecteur e de norme 1 et une fonction $\varphi : [a,b] \to \mathbb{R}_+$ tels que, pour tout $t \in [a,b]$, $f(t) = \varphi(t)e$. Dans le cas $n = 1$ on retrouve ainsi le fait que f garde un signe constant.

• Montrons pour commencer que cette condition est suffisante. En effet, on a alors,

$$\int_a^b f(t)\,\mathrm{d}t = \left(\int_a^b \varphi(t)\,\mathrm{d}t\right)e$$

et donc

$$\left\| \int_a^b f(t)\,\mathrm{d}t \right\| = \left| \int_a^b \varphi(t)\,\mathrm{d}t \right| = \int_a^b \varphi(t)\,\mathrm{d}t,$$

car φ est positive. Comme, pour tout $t \in [a,b]$, $\|f(t)\| = \varphi(t)$, on a le résultat voulu.

• Supposons réciproquement que $\left\| \int_a^b f \right\| = \int_a^b \|f\|$. Si $\int_a^b \|f\| = 0$, alors $\|f\| = 0$, puisque cette fonction est positive et continue, et $f = 0$. On prend dans ce cas e unitaire quelconque et $\varphi = 0$.

Sinon, on pose, comme le montre le sens direct $e = \dfrac{\int_a^b f}{\int_a^b \|f\|}$. On a alors

$$\left\| \int_a^b f \right\| = \left\langle e, \int_a^b f \right\rangle = \int_a^b \langle e, f(t)\rangle\,\mathrm{d}t \leqslant \int_a^b \|f(t)\|\,\mathrm{d}t = \left\| \int_a^b f \right\|,$$

1. Si $\int_a^b |f| = \varepsilon \int_a^b f$ avec $\varepsilon = \pm 1$, l'intégrale de la fonction continue positive $|f| - \varepsilon f$ est nulle, et donc $|f| = \varepsilon f$ *i.e.* $f = \varepsilon |f|$.

d'après l'inégalité de Cauchy-Schwarz. Ainsi l'inégalité est une égalité et $\int_a^b (\|f(t)\| - \langle e, f(t)\rangle)\,\mathrm{d}t = 0$. Comme la fonction qu'on intègre est continue et positive, elle est nulle. On a donc, pour tout $t \in [a,b]$, $\langle e, f(t)\rangle = \|f(t)\|$. On en déduit que $f(t) \in \mathbb{R}_+ e$, ce qui donne le résultat voulu si on pose, pour tout $t \in [a,b]$, $\varphi(t) = \|f(t)\|$. ◁

Pour $f : [a,b] \longrightarrow \mathbb{C}$, continue, l'égalité $\left|\int_a^b f\right| = \int_a^b |f|$ est donc réalisée si et seulement si l'argument de $f(t)$ est constant, i.e. s'il existe $\theta \in \mathbb{R}$ et $\rho : [a,b] \longrightarrow \mathbb{R}_+$ tels que $f(t) = \rho(t)e^{i\theta}$ pour tout $t \in [a,b]$.

1.6. Une inégalité (3)

Soit $\mathrm{E} = \mathcal{C}^1([a,b], \mathbb{R})$.

1. Montrer que, pour tout $\varepsilon > 0$, il existe $\mathrm{C} > 0$ tel que pour tout $f \in \mathrm{E}$ et tout $x \in [a,b]$,

$$|f(x)^2 - f(a)^2| \leqslant \mathrm{C}\int_a^b f^2 + \varepsilon \int_a^b f'^2.$$

2. Montrer que, pour tout $\varepsilon > 0$, il existe $\mathrm{D} > 0$ tel que, pour tout $f \in \mathrm{E}$,

$$\sup_{x\in[a,b]} f^2(x) \leqslant \mathrm{D}\int_a^b f^2 + \varepsilon \int_a^b f'^2.$$

(École polytechnique)

▷ **Solution.**

1. On écrit, pour tout $x \in [a,b]$,

$$|f(x)^2 - f(a)^2| = \left|\int_a^x 2f(t)f'(t)\,\mathrm{d}t\right| \leqslant \int_a^x 2|f(t)||f'(t)|\,\mathrm{d}t$$
$$\leqslant \int_a^b 2|f(t)||f'(t)|\,\mathrm{d}t.$$

On pense à la majoration $2|f(t)||f'(t)| \leqslant f^2(t) + f'^2(t)$. L'appliquer directement ne fait pas apparaître ε. On écrit donc

$$2|f(t)||f'(t)| = 2\left|\frac{1}{\sqrt{\varepsilon}}f(t)\right||\sqrt{\varepsilon}f'(t)| \leqslant \frac{1}{\varepsilon}f^2(t) + \varepsilon f'^2(t)$$

et on obtient

$$|f(x)^2 - f(a)^2| \leqslant \frac{1}{\varepsilon}\int_a^b f^2 + \varepsilon \int_a^b f'^2.$$

L'astuce consistant à majorer le produit $2ab$ *par* $\frac{1}{\varepsilon}a^2 + \varepsilon b^2$ *est très utilisée dans la théorie des équations aux dérivées partielles, théorie dont est issu cet énoncé.*

2. Soit $x \in [a,b]$. Pour tout $t \in [a,b]$, on a

$$f^2(x) = f^2(t) + \int_t^x 2f(u)f'(u)\,\mathrm{d}u.$$

On en déduit que

$$f^2(x) \leqslant f^2(t) + \int_a^b 2|f(u)||f'(u)|\mathrm{d}u.$$

En intégrant sur $[a,b]$ par rapport à t, on obtient

$$(b-a)f^2(x) \leqslant \int_a^b f^2(t)\,\mathrm{d}t + (b-a)\int_a^b 2|f(u)||f'(u)|\mathrm{d}u$$

et donc

$$f^2(x) \leqslant \frac{1}{b-a}\int_a^b f^2(t)\,\mathrm{d}t + \int_a^b 2|f(u)||f'(u)|\mathrm{d}u.$$

En raisonnant comme dans la première question, on en déduit que

$$\begin{aligned} f^2(x) &\leqslant \frac{1}{b-a}\int_a^b f^2(t)\,\mathrm{d}t + \frac{1}{\varepsilon}\int_a^b f^2 + \varepsilon\int_a^b f'^2 \\ &\leqslant \left(\frac{1}{b-a} + \frac{1}{\varepsilon}\right)\int_a^b f^2 + \varepsilon\int_a^b f'^2. \end{aligned}$$

Cela étant vrai pour tout $x \in [a,b]$, on a donc

$$\sup_{x\in[a,b]} f^2(x) \leqslant \left(\frac{1}{b-a} + \frac{1}{\varepsilon}\right)\int_a^b f^2 + \varepsilon\int_a^b f'^2. \quad \lhd$$

L'exercice suivant s'intéresse à des intégrales de la forme $\int_a^b e^{i\varphi(t)}\,dt$. *Il est clair que le module de cette intégrale est majoré par* $b-a$. *Toutefois, lorsque la fonction* φ *varie assez rapidement, on peut espérer obtenir une majoration meilleure puisque les complexes de module* 1 *vont avoir tendance à se compenser. L'inégalité de Van der Corput quantifie cette idée et donne une majoration remarquable qui ne dépend pas de la longueur de l'intervalle.*

1.7. Inégalité de Van der Corput

Montrer qu'il existe une suite $(c_k)_{k\geqslant 1}$ de réels positifs vérifiant
(*i*) pour tout $a<b$, tout $\varphi\in C^2([a,b],\mathbb{R})$ vérifiant $|\varphi'|\geqslant 1$ et φ' monotone, et tout $\lambda>0$

$$\left|\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x\right|\leqslant\frac{c_1}{\lambda};$$

(*ii*) pour tout $a<b$, tout $k\in\mathbb{N}^*$ avec $k\geqslant 2$, toute fonction $\varphi\in C^{k+1}([a,b],\mathbb{R})$ vérifiant $|\varphi^{(k)}|\geqslant 1$, tout $\lambda>0$

$$\left|\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x\right|\leqslant\frac{c_k}{\lambda^{1/k}}.$$

(**École normale supérieure**)

▷ **Solution.**

Nous allons démontrer l'existence des c_k par récurrence sur k.

• Commençons par prouver l'existence de c_1. Quitte à prendre le conjugué de l'intégrale à majorer (ce qui revient à remplacer φ par $-\varphi$) on peut faire l'hypothèse que $\varphi'\geqslant 1$. Comme φ' ne s'annule pas, on peut écrire

$$\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x=\int_a^b\frac{i\lambda\varphi'(x)e^{i\lambda\varphi(x)}}{i\lambda\varphi'(x)}\mathrm{d}x,$$

ce qui donne en intégrant par parties

$$\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x=\left[\frac{e^{i\lambda\varphi(x)}}{i\lambda\varphi'(x)}\right]_a^b+\int_a^b\frac{\varphi''(x)e^{i\lambda\varphi(x)}}{i\lambda\varphi'(x)^2}\mathrm{d}x.$$

En passant au module, on obtient

$$\left|\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x\right|\leqslant\frac{1}{\lambda}\left(\frac{1}{\varphi'(a)}+\frac{1}{\varphi'(b)}+\int_a^b\frac{|\varphi''(x)|}{\varphi'(x)^2}\mathrm{d}x\right).$$

Comme φ'' est de signe constant, on a

$$\int_a^b\frac{|\varphi''(x)|}{\varphi'(x)^2}\mathrm{d}x=\left|\int_a^b\frac{\varphi''(x)}{\varphi'(x)^2}\mathrm{d}x\right|=\left|\frac{1}{\varphi'(a)}-\frac{1}{\varphi'(b)}\right|\leqslant\frac{1}{\varphi'(a)}+\frac{1}{\varphi'(b)}.$$

Compte-tenu de l'inégalité $\varphi'\geqslant 1$, on majore finalement par

$$\left|\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x\right|\leqslant\frac{4}{\lambda}.$$

On peut donc prendre $c_1 = 4$.

• Supposons $k \geqslant 2$. Quitte à changer φ en son opposé, on peut à nouveau supposer $\varphi^{(k)} \geqslant 1$ (en particulier $\varphi^{(k-1)}$ est strictement croissante et s'annule au plus une fois).

⋆ Supposons tout d'abord que la fonction $\varphi^{(k-1)}$ s'annule en un point c de $]a, b[$. Pour $\alpha > 0$ tel que $c+\alpha \leqslant b$, on peut écrire, pour $x \in [c+\alpha, b]$, $\varphi^{(k-1)}(x) = \int_c^x \varphi^{(k)} \geqslant x - c \geqslant \alpha$ et donc $\dfrac{\varphi^{(k-1)}(x)}{\alpha} \geqslant 1$. Par hypothèse de récurrence, on a donc

$$\left|\int_{c+\alpha}^{b} e^{i\lambda\varphi(x)}\mathrm{d}x\right| = \left|\int_{c+\alpha}^{b} e^{\lambda\alpha i\frac{\varphi(x)}{\alpha}}\mathrm{d}x\right| \leqslant \frac{c_{k-1}}{(\alpha\lambda)^{1/(k-1)}}.$$

De même, si $c - \alpha \geqslant a$, on a $-\varphi^{(k-1)}(x) = \int_x^c \varphi^{(k)} \geqslant c - x \geqslant \alpha$ et donc $\left|\dfrac{\varphi^{(k-1)}(x)}{\alpha}\right| \geqslant 1$ pour $x \in [a, c-\alpha]$. On peut majorer de façon identique

$$\left|\int_{a}^{c-\alpha} e^{i\lambda\varphi(x)}\mathrm{d}x\right| = \left|\int_{a}^{c-\alpha} e^{\lambda\alpha i\frac{\varphi(x)}{\alpha}}\mathrm{d}x\right| \leqslant \frac{c_{k-1}}{(\alpha\lambda)^{1/(k-1)}}.$$

Comme $\left|\int_{c-\alpha}^{c+\alpha} e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \int_{c-\alpha}^{c+\alpha} 1 = 2\alpha$, on a donc

$$\left|\int_{a}^{b} e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant 2\alpha + \frac{2c_{k-1}}{(\alpha\lambda)^{1/(k-1)}}.$$

Remarquons que cette inégalité est vraie en fait pour tout $\alpha > 0$: en effet, si $c + \alpha > b$ (resp. $c - \alpha < a$) , on a alors $\left|\int_c^b e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \alpha$ (resp. $\left|\int_a^c e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \alpha$). Il suffit alors de prendre $\alpha = \dfrac{1}{\lambda^{1/k}}$ et il vient

$$\left|\int_{a}^{b} e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \frac{2}{\lambda^{1/k}} + \frac{2c_{k-1}}{(\lambda^{(1-1/k)})^{1/(k-1)}} = \frac{2 + 2c_{k-1}}{\lambda^{1/k}}.$$

⋆ Supposons maintenant que $\varphi^{(k-1)}$ ne s'annule pas sur $]a, b[$. Par continuité, elle reste de signe constant. Dans le cas où elle est positive, on a $\varphi^{(k-1)}(a) \geqslant 0$. Pour $\alpha > 0$ avec $a + \alpha < b$, on peut écrire que si $x \in [a + \alpha, b]$, $\varphi^{(k-1)}(x) = \int_a^x \varphi^{(k)} + \varphi^{(k-1)}(a) \geqslant (x - a) + 0 \geqslant \alpha$. En majorant comme précédemment, on trouve

$$\left|\int_{a}^{b} e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \int_{a}^{a+\alpha} 1 + \left|\int_{a+\alpha}^{b} e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \alpha + \frac{c_{k-1}}{(\alpha\lambda)^{1/(k-1)}},$$

inégalité qui reste évidemment valable si $a+\alpha \geqslant b$. Toujours en prenant $\alpha = \frac{1}{\lambda^{1/k}}$, il vient

$$\left|\int_a^b e^{i\lambda\varphi(x)}\mathrm{d}x\right| \leqslant \frac{1+c_{k-1}}{\lambda^{1/k}} \leqslant \frac{2+2c_{k-1}}{\lambda^{1/k}}.$$

Dans le cas où $\varphi^{(k-1)}$ reste négative, on considère un découpage de l'intégrale en $b-\alpha$ et on arrive à la même majoration.

Il suffit donc de poser $c_{k+1} = 2(1+c_k)$. ◁

Les méthodes usuelles de calcul approché de l'intégrale d'une fonction f, disons continue, sur un segment $[a,b]$, consistent à remplacer f par une fonction plus simple (constante, affine, polynomiale,...) sur chaque intervalle de la subdivision régulière du segment en n morceaux, puis de faire tendre n vers l'infini. On obtient alors une suite qui converge vers $\int_a^b f$. Pour estimer la vitesse de convergence de cette suite on est souvent amené à faire une hypothèse de régularité plus forte sur la fonction f. L'exercice suivant propose une telle estimation pour une fonction de classe $\mathcal{C}^2$ dans le cadre de la méthode des trapèzes.

1.8. Majoration de l'erreur dans la méthode des trapèzes

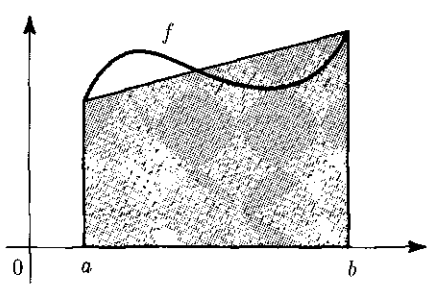
Soit $f \in \mathcal{C}^2([a,b],\mathbb{R})$. Montrer que

$$\left|\int_a^b f(t)\mathrm{d}t - \frac{b-a}{2}\Big(f(a)+f(b)\Big)\right| \leqslant \frac{(b-a)^3}{12}\|f''\|_\infty.$$

(École polytechnique)

▷ **Solution.**

Il est bon de remarquer que la quantité à majorer mesure la différence entre l'aire définie entre la courbe et l'axe des x et l'aire du trapèze défini par la sécante à la courbe en a et b et l'axe des x.



Si on pose pour $x \in [a,b]$, $g(x) = f(x) - \left[\dfrac{f(b)-f(a)}{b-a}(x-a) + f(a)\right]$ (on retire à f l'équation de la sécante, on a g de classe C^2, $g'' = f''$, $g(a) = g(b) = 0$ et

$$\int_a^b g = \int_a^b f - \frac{b-a}{2}\Big(f(a)+f(b)\Big).$$

Afin de faire apparaître la dérivée seconde de g, nous allons intégrer deux fois par parties : nous allons dériver g deux fois et intégrer deux fois la constante 1. Cependant, afin de supprimer les termes entre crochets, nous choisirons judicieusement les constantes des primitives. On a donc

$$\begin{aligned}\int_a^b g &= [(t-m)g]_a^b - \int_a^b (t-m)g'(t)\mathrm{d}t = -\int_a^b (t-m)g'(t)\mathrm{d}t \\ &= -\left[\left(\frac{t^2}{2} - mt + h\right)\right]_a^b + \int_a^b \left(\frac{t^2}{2} - mt + h\right) g''(t)\mathrm{d}t\end{aligned}$$

où m et h sont des constantes à choisir. Prenons donc m et h dans $\mathbb{R}$ tels que $\left(\dfrac{t^2}{2} - mt + h\right) = \dfrac{(t-a)(t-b)}{2}$. Le crochet est alors nul et $\int_a^b g = \int_a^b \dfrac{(t-a)(t-b)}{2} g''(t)\mathrm{d}t = \int_a^b \dfrac{(t-a)(t-b)}{2} f''(t)\mathrm{d}t$. Comme f'' est continue sur le compact $[a,b]$, elle est bornée et

$$\begin{aligned}\left|\int_a^b g\right| &\leqslant \|f''\|_\infty \int_a^b \frac{(t-a)(b-t)}{2}\mathrm{d}t \\ &\leqslant \frac{\|f''\|_\infty}{2}\left(\left[\frac{(t-a)^2}{2}(b-t)\right]_a^b + \int_a^b \frac{(t-a)^2}{2}\mathrm{d}t\right)\end{aligned}$$

ce qui donne $\left|\int_a^b g\right| \leqslant \dfrac{\|f''\|_\infty}{2}\left(0 + \left[\dfrac{(t-a)^3}{6}\right]_a^b\right) = \dfrac{\|f''\|_\infty}{12}(b-a)^3$. On conclut donc

$$\boxed{\left|\int_a^b f(t)\mathrm{d}t - \frac{b-a}{2}(f(a)+f(b))\right| \leqslant \frac{(b-a)^3}{12}\|f''\|_\infty}. \quad \lhd$$

On en déduit une majoration de l'erreur dans la méthode des trapèzes. On considère une subdivision $a = x_0 < x_1 < \cdots < x_n = b$ de l'intervalle $[a,b]$ en n intervalles égaux. Pour $k \in [\![0, n-1]\!]$, on remplace $\int_{x_k}^{x_{k+1}} f(t)\,dt$ par l'aire du trapèze défini par la sécante à la courbe en x_k et x_{k+1} et l'axe des abscisses. On obtient $\frac{1}{2}f(x_0) + \sum_{k=1}^{n-1} f(x_k) + \frac{1}{2}f(x_n)$ comme valeur approchée de $\int_a^b f(t)\,dt$. L'erreur commise est majorée par $\dfrac{\|f''\|_\infty (b-a)^3}{12n^2}$.

On dit que la méthode des trapèzes est du deuxième ordre. Le lecteur pourra se reporter à l'exercice 1.19 pour d'autres résultats concernant le calcul approché d'intégrales.

L'exercice concerne une estimation du même type que celle de l'énoncé précédent, mais avec une fonction convexe.

1.9. Méthode des trapèzes pour une fonction convexe

Soit f une fonction convexe de $\mathcal{C}^1(\mathbb{R}_+,\mathbb{R})$ et $n \in \mathbb{N}$, $n \geqslant 2$. Montrer que

$$0 \leqslant \frac{1}{2}f(0)+f(1)+\cdots+f(n-1)+\frac{1}{2}f(n)-\int_0^n f \leqslant \frac{1}{8}(f'(n)-f'(0)).$$

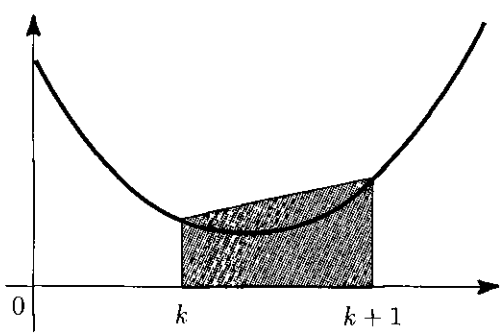
(École polytechnique)

▷ **Solution.**

On observe que la quantité $\dfrac{f(k)+f(k+1)}{2}$ représente l'aire du trapèze délimité par l'axe des abscisses, les droites d'équation $x = k$ et $x = k+1$ et la corde du graphe de f qui joint les points $(k, f(k))$ et $(k+1, f(k+1))$. Ainsi,

$$S_n = \frac{1}{2}f(0) + f(1) + \cdots + f(n-1) + \frac{1}{2}f(n) = \sum_{k=0}^{n-1} \frac{f(k)+f(k+1)}{2}$$

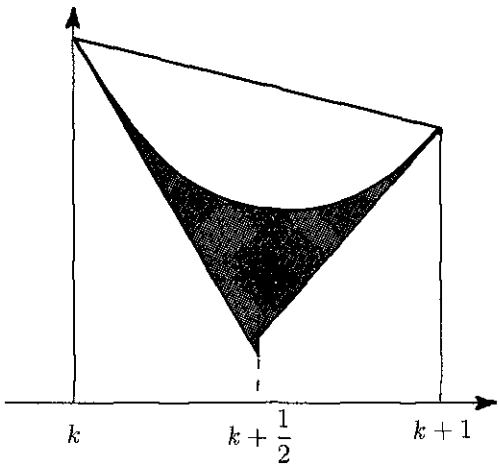
est la somme des aires de ces trapèzes.



Comme la fonction f est convexe, il est déjà clair que $\int_0^n f \leqslant S_n$. Cela provient de ce que sur chaque intervalle $[k, k+1]$ le graphe de f se trouve en-dessous de la corde qui joint les points $(k, f(k))$ et $(k+1, f(k+1))$.

On cherche maintenant à majorer l'erreur $S_n - \int_0^n f$. On se place sur un intervalle $[k, k+1]$ et on majore $\dfrac{f(k)+f(k+1)}{2} - \int_k^{k+1} f$.

Géométriquement, l'idée consiste à minorer f par sa tangente en k sur $[k, k+\frac{1}{2}]$ et par sa tangente en $k+1$ sur $[k+\frac{1}{2}, k+1]$.



On a donc pour tout $t \in [k, k+\frac{1}{2}]$, $f(t) \geqslant f(k) + f'(k)(t-k)$ et pour tout $t \in [k+\frac{1}{2}, k+1]$, $f(t) \geqslant f(k+1) + f'(k+1)(t-k-1)$. Calculons

$$\int_k^{k+\frac{1}{2}} [f(k) + f'(k)(t-k)]\mathrm{d}t = \frac{f(k)}{2} + f'(k)\left[\frac{(t-k)^2}{2}\right]_k^{k+\frac{1}{2}}$$
$$= \frac{f(k)}{2} + \frac{f'(k)}{8}$$

et

$$\int_{k+\frac{1}{2}}^{k+1} [f(k+1) + f'(k+1)(t-k-1)]\mathrm{d}t$$
$$= \frac{f(k+1)}{2} - f'(k+1)\left[\frac{(t-k-1)^2}{2}\right]_{k+\frac{1}{2}}^{k+1}$$
$$= \frac{f(k)}{2} - \frac{f'(k+1)}{8}.$$

Ainsi, $\int_k^{k+1} f$ est minoré par $\dfrac{f(k)+f(k+1)}{2} + \dfrac{1}{8}(f'(k) - f'(k+1))$. On a donc $\dfrac{f(k)+f(k+1)}{2} - \int_k^{k+1} f \leqslant \dfrac{1}{8}(f'(k+1) - f'(k))$. En sommant cela pour k entre 0 et $n-1$ on obtient la majoration demandée,

$$\boxed{0 \leqslant S_n - \int_0^n f \leqslant \frac{1}{8}(f'(n) - f'(0))}. \quad \triangleleft$$

Pour établir la majoration, on peut aussi reprendre la méthode de l'exercice précédent et intégrer par parties, en considérant la primitive $t \longmapsto t - (k + \frac{1}{2})$ *de* 1. *On obtient*

$$\frac{1}{2}(f(k) + f(k+1)) - \int_k^{k+1} f(t)\,dt = \int_k^{k+1} \left(t - \left(k + \frac{1}{2}\right)\right) f'(t)\,dt.$$

Comme f' *est croissante, on en déduit*

$$\int_k^{k+\frac{1}{2}} \left(t - \left(k + \frac{1}{2}\right)\right) f'(t)\,dt \leqslant f'(k) \int_k^{k+\frac{1}{2}} \left(t - \left(k + \frac{1}{2}\right)\right) = -\frac{f'(k)}{8},$$

$$\int_{k+\frac{1}{2}}^{k+1} \left(t - \left(k + \frac{1}{2}\right)\right) f'(t)\,dt \leqslant f'(k+1) \int_{k+\frac{1}{2}}^{k+1} \left(t - \left(k + \frac{1}{2}\right)\right) dt = \frac{f'(k+1)}{8}$$

et donc $\frac{1}{2}(f(k) + f(k+1)) - \int_k^{k+1} f(t)\,dt \leqslant \frac{f'(k+1) - f'(k)}{8}$.

1.10. Recherche d'un minimum

Trouver le minimum de $\int_{-1}^{1} |t^2 + at + b|\,dt$ lorsque a et b décrivent $\mathbb{C}$.

(École polytechnique)

▷ **Solution.**

Posons, pour $(a, b) \in \mathbb{C}^2$, $f(a, b) = \int_{-1}^{1} |t^2 + at + b|\,dt$. On constate que

$$\begin{aligned} f(a,b) &= \int_{-1}^{1} \sqrt{(t^2 + \mathrm{Re}(a)t + \mathrm{Re}(b))^2 + (\mathrm{Im}(a)t + \mathrm{Im}(b))^2}\,dt \\ &\geqslant \int_{-1}^{1} |t^2 + \mathrm{Re}(a)t + \mathrm{Re}(b)|\,dt = f(\mathrm{Re}(a), \mathrm{Re}(b)). \end{aligned}$$

On peut donc se limiter à $(a, b) \in \mathbb{R}^2$, ce qu'on fait dans la suite de l'exercice.

L'observation essentielle est que la fonction $(a, b) \mapsto \int_{-1}^{1} |t^2 + at + b|dt$ est convexe sur $\mathbb{R}^2$. En effet, si (a, b) et (a', b') sont deux couples de $\mathbb{R}^2$ et λ un réel de $[0, 1]$, par l'inégalité triangulaire, on a pour $t \in [-1, 1]$,

$$|t^2 + ((1-\lambda)a + \lambda a')t + (1-\lambda)b + \lambda b')| \leqslant (1-\lambda)|t^2 + at + b| + \lambda|t^2 + a't + b'|.$$

En intégrant cela par rapport à t, on obtient

$$f((1-\lambda)a+\lambda a', (1-\lambda)b+\lambda b') \leqslant (1-\lambda)f(a,b)+\lambda f(a',b').$$

On peut même observer que f est strictement convexe, c'est-à-dire que l'inégalité ci-dessus est stricte si $(a,b) \neq (a',b')$ et $\lambda \in]0,1[$. Comme il est clair que f est continue et que $f(a,b)$ tend vers $+\infty$ lorsque $|a|+|b|$ tend vers $+\infty$, on en déduit que f admet un minimum sur $\mathbb{R}^2$ et que celui-ci est atteint en un point unique. On va déterminer ce point.

Si on effectue le changement de variable $u=-t$, on obtient

$$f(a,b)=\int_{-1}^{1}|t^2+at+b|\mathrm{d}t=\int_{-1}^{1}|u^2-au+b|\mathrm{d}u=f(-a,b).$$

Ainsi, si b est fixé, la fonction $a \mapsto f(a,b)$ est paire. Or, d'après ce qui précède, il s'agit d'une fonction convexe d'une variable réelle : elle atteint donc son minimum en 0 (car $f(0) \leqslant \dfrac{f(x)+f(-x)}{2} = f(x)$ pour tout x). Le minimum de f est donc nécessairement atteint lorsque $a=0$.

Il ne reste plus qu'à étudier le minimum de $f(0,b)=\int_{-1}^{1}|t^2+b|\mathrm{d}t$. On calcule cette intégrale en discutant selon la position de b par rapport à -1 et 0 pour enlever la valeur absolue.

- Si $b \geqslant 0$, $f(0,b)=\dfrac{2}{3}+2b$.
- Si $b \leqslant -1$, $f(0,b)=-2b-\dfrac{2}{3}$.
- Enfin si $-1<b<0$, on pose $b=-c^2$ avec $0<c<1$ et on a

$$f(0,b)=2\int_0^1|t^2-c^2|\mathrm{d}t=2\int_0^c(c^2-t^2)\mathrm{d}t+2\int_c^1(t^2-c^2)\mathrm{d}t.$$

Après simplification, on obtient $f(0,b)=\dfrac{2}{3}(4c^3-3c^2+1)$. Une étude de fonction montre que le polynôme $4c^3-3c^2+1$ atteint sur l'intervalle $[0,1]$ son minimum en $\dfrac{1}{2}$.

On peut donc conclure que la fonction f atteint son minimum pour $a=0$ et $b=-\dfrac{1}{4}$. En ce point elle vaut $\dfrac{1}{2}$. ◁

1.11. Recherche d'une borne inférieure

Soit F l'ensemble des fonctions $f \in \mathcal{C}^1([0,1],\mathbb{R})$ vérifiant $f(0)=0$ et $f(1)=1$. Déterminer $\inf\limits_{f\in F}\int_0^1|f'-f|$.

(École normale supérieure, École polytechnique)

▷ **Solution.**

La quantité $f' - f$ invite à introduire la fonction $g : x \mapsto f(x)e^{-x}$. En effet, g est de classe $\mathcal{C}^1$ et on a $g'(x) = (f'(x) - f(x))e^{-x}$ pour tout $x \in [0,1]$. On remplace donc $f'(x) - f(x)$ par $e^x g'(x)$. Lorsque f parcourt l'ensemble F la fonction g parcourt l'ensemble G des fonctions de classe $\mathcal{C}^1$ qui prennent la valeur 0 en 0 et la valeur $\dfrac{1}{e}$ en 1. Et $h = g'$ parcourt donc l'ensemble H des fonctions continues dont l'intégrale sur $[0,1]$ vaut $\dfrac{1}{e}$. On a donc

$$\inf_{f\in \mathrm{F}} \int_0^1 |f' - f| = \inf_{g\in \mathrm{G}} \int_0^1 e^x |g'(x)| \mathrm{d}x = \inf_{h\in \mathrm{H}} \int_0^1 e^x |h(x)| \mathrm{d}x.$$

Ce changement de variable permet de résoudre facilement le problème posé. En effet, si $h \in \mathrm{H}$ on a

$$\int_0^1 e^x |h(x)| \mathrm{d}x \geqslant \int_0^1 |h(x)| \mathrm{d}x \geqslant \int_0^1 h(x) \mathrm{d}x = \frac{1}{e}.$$

Montrons que $\dfrac{1}{e}$ est exactement la borne inférieure recherchée. Pour cela, on va construire une suite de fonctions $(h_n)_{n\geqslant 1}$ telle que $\displaystyle\int_0^1 e^x |h_n(x)| \mathrm{d}x$ tende vers $\dfrac{1}{e}$. On choisit des fonctions positives, affines par morceaux qui concentrent la masse au voisinage de 0. Plus précisément, on prend h_n nulle sur $[1/n, 1]$ et affine sur $[0, 1/n]$ avec $h_n(0) = \dfrac{2n}{e}$ et $h_n(1/n) = 0$. Il est facile de voir que $h_n \in \mathrm{H}$ et

$$\int_0^1 e^x |h_n(x)| \mathrm{d}x = \int_0^{1/n} e^x h_n(x) \mathrm{d}x \leqslant e^{1/n} \int_0^1 h_n(x) \mathrm{d}x = \frac{e^{1/n}}{e}.$$

Le majorant tend bien vers $\dfrac{1}{e}$ lorsque n tend vers l'infini ce qui prouve le résultat.

Conclusion. On a $\boxed{\displaystyle\inf_{f\in \mathrm{F}} \int_0^1 |f' - f| = \frac{1}{e}}$. ◁

Les exercices qui suivent concernent encore des inégalités et ils reposent sur l'importante inégalité de Cauchy-Schwarz.

1.12. Inégalité de Hardy

Soit $f : [0,1] \longrightarrow \mathbb{R}$ une fonction continue. On pose pour tout $x \in]0,1]$, $\mathrm{F}(x) = \frac{1}{x}\int_0^x f(t)\mathrm{d}t$ et $\mathrm{F}(0) = f(0)$. Montrer que

$$\int_0^1 \mathrm{F}^2(x)\mathrm{d}x \leqslant 4\int_0^1 f^2(x)\mathrm{d}x.$$

(École polytechnique)

▷ **Solution.**

Comme f est continue, F est de classe $\mathcal{C}^1$ sur $]0,1]$. De plus la fonction F est continue en 0, puisque $\mathrm{F}(x)$ est le taux d'accroissement entre 0 et x d'une primitive de f. Cela justifie donc l'existence de l'intégrale de F^2. Pour tout x dans $]0,1]$ on a

$$\mathrm{F}'(x) = \frac{-1}{x^2}\int_0^x f + \frac{f(x)}{x}$$

autrement dit $x\mathrm{F}'(x) = f(x) - \mathrm{F}(x)$. En multipliant par $\mathrm{F}(x)$, on obtient $\mathrm{F}^2(x) = f(x)\mathrm{F}(x) - x\mathrm{F}(x)\mathrm{F}'(x)$. On va intégrer cette relation entre $\varepsilon > 0$ et 1. On obtient

$$\int_\varepsilon^1 \mathrm{F}^2 = \int_\varepsilon^1 f\mathrm{F} - \int_\varepsilon^1 x\mathrm{F}(x)\mathrm{F}'(x)\mathrm{d}x = \int_\varepsilon^1 f\mathrm{F} - \left[x\frac{\mathrm{F}^2(x)}{2}\right]_\varepsilon^1 + \frac{1}{2}\int_\varepsilon^1 \mathrm{F}^2.$$

En faisant tendre ε vers 0, on en déduit donc que

$$\int_0^1 \mathrm{F}^2 = 2\int_0^1 f\mathrm{F} - \mathrm{F}^2(1) \leqslant 2\int_0^1 f\mathrm{F}.$$

On en déduit, grâce à l'inégalité de Cauchy-Schwarz, que

$$\int_0^1 \mathrm{F}^2 \leqslant 2\int_0^1 f\mathrm{F} \leqslant 2\sqrt{\int_0^1 f^2}\sqrt{\int_0^1 \mathrm{F}^2}.$$

Si $\int_0^1 \mathrm{F}^2 > 0$, en élevant au carré, il reste $\left(\int_0^1 \mathrm{F}^2\right)^2 \leqslant 4\int_0^1 f^2 \int_0^1 \mathrm{F}^2$ et en divisant par l'intégrale de F^2, on trouve la majoration voulue

$$\boxed{\int_0^1 \mathrm{F}^2 \leqslant 4\int_0^1 f^2}.$$

Cette majoration reste clairement valable lorsque $\int_0^1 \mathrm{F}^2 = 0$. ◁

Le résultat est un cas particulier des inégalités de Hardy. Le cas général consiste en une majoration de la la norme p de F *en fonction de celle de* f : $\|\mathrm{F}\|_p \leqslant q\|f\|_p$, *où* $\|f\|_p = \left(\int_0^1 |f|^p\right)^{\frac{1}{p}}$ *et* $\frac{1}{p} + \frac{1}{q} = 1$ *(ici on a le cas* $p = q = 2$*).*

1.13. Recherche d'un minimum

Déterminer le minimum de $\int_0^1 (f''(t))^2 \mathrm{d}t$ dans l'ensemble des fonctions de classe C^2 de $[0,1]$ dans $\mathbb{R}$ vérifiant $f(0) = f(1) = 0$ et $f'(0) = a$ où a est un réel donné.

(École polytechnique)

▷ **Solution.**

Le carré présent dans l'intégrale nous invite à essayer d'utiliser l'inégalité de Cauchy-Schwarz. Soit f une fonction dans l'ensemble considéré par l'énoncé. On a des informations sur f en deux points distincts à savoir 0 et 1. On va estimer $f(1)$ à partir des informations en 0 en utilisant la formule de Taylor avec reste intégral. On a

$$f(1) = f(0) + f'(0) + \int_0^1 (1-t)f''(t)\mathrm{d}t,$$

ce qui, compte-tenu des hypothèses vérifiées par f, donne

$$-a = \int_0^1 (1-t)f''(t)\mathrm{d}t.$$

L'inégalité de Cauchy-Schwarz donne alors,

$$a^2 \leqslant \int_0^1 (1-t)^2\mathrm{d}t \int_0^1 (f''(t))^2\mathrm{d}t = \frac{1}{3}\int_0^1 (f''(t))^2\mathrm{d}t.$$

Cette inégalité est une égalité si et seulement si f'' est proportionnelle à la fonction $t \longmapsto 1-t$.

Si $f''(t) = \lambda(1-t)$, f est de la forme $f(t) = \frac{\lambda}{6}(1-t)^3 + \alpha(1-t) + \beta$. La condition $f(1) = 0$ impose $\beta = 0$. La condition $f(0) = 0$ donne $\lambda = -6\alpha$ et enfin la condition $f'(0) = a$ conduit à $\alpha = \frac{a}{2}$.

Conclusion. Le minimum de $\int_0^1 (f''(t))^2\mathrm{d}t$ est égal à $3a^2$ et il n'est atteint que pour la fonction $t \longmapsto \frac{a}{2}((1-t)-(1-t)^3) = \frac{a}{2}(1-t)t(2-t)$. ◁

1.14. Étude d'une suite

Soient f et g continues de $[0,1]$ dans $]0,+\infty[$. On pose, pour tout n de $\mathbb{N}$, $u_n = \int_0^1 g(x)f(x)^n \mathrm{d}x$. Étudier la suite $\left(\frac{u_{n+1}}{u_n}\right)_{n\in\mathbb{N}}$.

(École polytechnique)

▷ **Solution.**

Les fonctions f et g sont continues à valeurs dans $]0,+\infty[$. La suite $(u_n)_{n\in\mathbb{N}}$ est donc à termes strictement positifs. Posons, pour tout entier naturel n, $v_n = \frac{u_{n+1}}{u_n}$. Pour tout n de $\mathbb{N}$, on a

$$u_{n+1} = \int_0^1 g(x)(f(x))^{n+1}\mathrm{d}x = \int_0^1 \left(g(x)^{\frac{1}{2}} f(x)^{\frac{n+2}{2}}\right)\left(g(x)^{\frac{1}{2}} f(x)^{\frac{n}{2}}\right)\mathrm{d}x.$$

En appliquant l'inégalité de Cauchy-Schwarz, on obtient

$$(u_{n+1})^2 \leqslant \int_0^1 g(x)f(x)^{n+2}\mathrm{d}x \int_0^1 g(x)f(x)^n \mathrm{d}x,$$

c'est-à-dire $(u_{n+1})^2 \leqslant u_{n+2}u_n$. On a donc $\frac{u_{n+1}}{u_n} \leqslant \frac{u_{n+2}}{u_{n+1}}$, soit encore $v_n \leqslant v_{n+1}$. La suite $(v_n)_{n\in\mathbb{N}}$ est donc croissante. Montrons qu'elle est majorée. Posons $\mathrm{M} = \sup_{[0,1]} f$. Pour tout entier naturel n et tout x de $[0,1]$, on a $g(x)f(x)^{n+1} \leqslant \mathrm{M}g(x)f(x)^n$. En intégrant cette inégalité on en déduit que $u_{n+1} \leqslant \mathrm{M}u_n$ et donc que $v_n \leqslant \mathrm{M}$.

La suite $(v_n)_{n\in\mathbb{N}}$ est croissante et majorée par M donc converge. On note ℓ sa limite. C'est un réel strictement positif car $(v_n)_{n\in\mathbb{N}}$ est croissante. La suite $\ln v_n = \ln u_{n+1} - \ln u_n$ converge vers $\ln \ell$ donc d'après le théorème de Cesàro la suite $\frac{\ln u_n}{n}$ converge aussi vers $\ln \ell$. En repassant à l'exponentielle on en déduit que la suite $u_n^{1/n}$ converge vers ℓ. Or la limite de cette suite peut se calculer directement (c'est une question classique que le lecteur retrouvera dans l'exercice 1.32). Précisons cela.

La borne supérieure de f sur le compact $[0,1]$ est atteinte : il existe α dans $[0,1]$ tel que $f(\alpha) = \mathrm{M}$. Soit $\varepsilon > 0$; il existe un intervalle $[a,b]$, avec $a < b$, inclus dans $[0,1]$, tel que pour tout x de $[a,b]$, on ait l'encadrement $\mathrm{M} - \varepsilon \leqslant f(x) \leqslant \mathrm{M}$. On a alors pour tout entier naturel n,

$$\int_0^1 g(x)f(x)^n\mathrm{d}x \geqslant \int_a^b g(x)(\mathrm{M}-\varepsilon)^n\mathrm{d}x \geqslant (\mathrm{M}-\varepsilon)^n \int_a^b g(x)\mathrm{d}x$$

$$\text{et} \quad \int_0^1 g(x)f(x)^n\mathrm{d}x \leqslant \mathrm{M}^n \int_0^1 g(x)\mathrm{d}x.$$

On a donc $(M-\varepsilon)\left(\int_a^b g(x)\mathrm{d}x\right)^{\frac{1}{n}} \leqslant u_n^{1/n} \leqslant M\left(\int_0^1 g(x)\mathrm{d}x\right)^{\frac{1}{n}}$. Sachant que

$$\lim_{n\to\infty}(M-\varepsilon)\left(\int_a^b g(x)\mathrm{d}x\right)^{\frac{1}{n}} = M-\varepsilon \text{ et } \lim_{n\to\infty} M\left(\int_0^1 g(x)\mathrm{d}x\right)^{\frac{1}{n}} = M,$$

on en déduit qu'il existe un entier naturel n_0 tel que, pour tout entier $n \geqslant n_0$,

$$M-2\varepsilon \leqslant u_n^{1/n} \leqslant M+2\varepsilon.$$

On a donc $\lim_{n\to\infty} u_n^{1/n} = M$ et par conséquent $\ell = M$.

Conclusion. La suite $\left(\frac{u_{n+1}}{u_n}\right)_{n\in\mathbb{N}}$ converge vers $M = \sup_{[0,1]} f$. ◁

L'énoncé suivant établit l'inégalité de Wirtinger. Le lecteur trouvera une autre démonstration de cette inégalité à l'aide des séries de Fourier (cf. exercice 4.20).

1.15. Inégalité de Wirtinger

Soit E l'ensemble des fonctions $f : [0,1] \to \mathbb{R}$ de classe $\mathcal{C}^1$ telles que $f(0) = f(1) = 0$.

1. Soit $f \in$ E. Montrer que $\mathrm{I}_1 = \int_0^1 f(t)f'(t)\operatorname{cotan}(\pi t)\mathrm{d}t$ et $\mathrm{I}_2 = \int_0^1 \frac{f(t)^2}{\tan^2 \pi t}(1+\tan^2 \pi t)\mathrm{d}t$ existent. Comparer I_1 et I_2.

2. Montrer que pour tout $f \in$ E, on a

$$\int_0^1 f'^2 \geqslant \pi^2 \int_0^1 f^2$$

(*inégalité de Wirtinger*).

3. Quels sont les cas d'égalité ?

(**École polytechnique**)

Solution.

1. Commençons par justifier l'existence de I_1. On a *a priori* un problème en 0 et en 1. Comme $f(0) = 0$ et que f est de classe $\mathcal{C}^1$, on a en 0 le développement limité suivant : $f(t) = f'(0)t + o(t)$. On sait par ailleurs que, toujours en 0, $\operatorname{cotan} \pi t \sim \frac{1}{\pi t}$. On a donc

$$f(t)f'(t)\operatorname{cotan} \pi t \xrightarrow[t\to 0]{} \frac{f'(0)^2}{\pi}.$$

La fonction intégrée se prolonge donc par continuité en 0. Il en est de même en 1 de sorte que I_1 est parfaitement défini. Pour l'intégrale I_2 c'est pareil : on vérifie que la fonction intégrée que l'on peut peut écrire $t \longmapsto f(t)^2(1+\operatorname{cotan}^2 \pi t)$ se prolonge par continuité en 0, 1 et $\frac{1}{2}$.

Pour comparer I_1 et I_2, on effectue une intégration par parties dans I_1. Pour tout $\varepsilon > 0$, $\int_\varepsilon^{1-\varepsilon} f(t)f'(t)\operatorname{cotan} \pi t dt$ vaut

$$\left[\frac{1}{2}f^2(t)\operatorname{cotan} \pi t\right]_\varepsilon^{1-\varepsilon} + \frac{\pi}{2}\int_\varepsilon^{1-\varepsilon} f^2(t)(1+\operatorname{cotan}^2 \pi t)dt.$$

Si on fait tendre ε vers 0 il vient d'après ce qui précède $\boxed{I_1 = \frac{\pi}{2}I_2}$.

2. Soit $f \in E$ non nulle (si f est nulle l'inégalité est triviale). D'après l'inégalité de Cauchy-Schwarz, on a

$$I_1^2 \leqslant \int_0^1 (f'(t))^2 dt \int_0^1 (f(t))^2 \operatorname{cotan}^2 \pi t dt.$$

Posons alors $I_2 = A + B$ avec $A = \int_0^1 f^2$ et $B = \int_0^1 (f(t))^2 \operatorname{cotan}^2 \pi t dt$. Alors, $I_2^2 = (A+B)^2 \geqslant 4AB$ (inégalité qui provient tout simplement du fait que $(A-B)^2 \geqslant 0$). On a donc

$$\int_0^1 f'^2 \geqslant \frac{I_1^2}{B} = \frac{\pi^2}{4}\frac{I_2^2}{B} \geqslant \pi^2 A = \pi^2 \int_0^1 f^2.$$

C'est le résultat demandé.

3. On a égalité si et seulement si il y a égalité dans l'inégalité de Cauchy-Schwarz et si $A = B$. Soit $f \in E$ une fonction non nulle réalisant l'égalité. Il existe un réel λ tel que $f'(t) = kf(t)\operatorname{cotan} \pi t$ pour tout $t \in]0,1[$. Alors, on a

$$k^2 \int_0^1 f(t)^2 \operatorname{cotan}^2 \pi t dt = \int_0^1 f'^2 = \pi^2 \int_0^1 f^2 = \pi^2 \int_0^1 f(t)^2 \operatorname{cotan}^2 \pi t dt$$

et on en déduit que $k^2 = \pi^2$.

• Si $k = \pi$, les solutions de l'équation $f'(t) = kf(t)\operatorname{cotan} \pi t$ sont les fonctions de la forme $t \longmapsto A \sin \pi t$, $A \in \mathbb{R}$. Elles sont dans E et vérifient toutes le cas d'égalité.

• Si $k = -\pi$, les solutions de l'équation différentielle sont les fonctions $t \longmapsto \dfrac{A}{\sin \pi t}$. Il ne s'agit pas d'éléments de E. ◁

De la linéarité de l'intégrale, il résulte notamment que l'application $(f,g) \mapsto \int_a^b fg$ *est une forme bilinéaire symétrique positive. C'est même*

un produit scalaire si on se restreint à un sous-espace de $C^0([a,b],\mathbb{R})$. Les exercices qui suivent s'interprètent alors géométriquement comme la recherche de l'orthogonal de certains sous-espaces au sens de ce produit scalaire.

1.16. Fonctions dont les deux premiers moments sont nuls

Soit $a < b$ deux réels et E l'ensemble des fonctions continues de $[a,b]$ dans $\mathbb{R}$. On pose

$$\mathrm{F} = \{g \in C^2([a,b],\mathbb{R}),\ g(a) = g(b) = g'(a) = g'(b) = 0\}.$$

1. Soit $f \in \mathrm{E}$. Montrer l'équivalence :

$$\exists g \in \mathrm{F},\quad g'' = f \iff \int_a^b f(t)\mathrm{d}t = \int_a^b tf(t)\mathrm{d}t = 0.$$

2. Soit $h \in \mathrm{E}$ telle que $\int_a^b h(t)g''(t)\mathrm{d}t = 0$ pour tout $g \in \mathrm{F}$. Montrer que h est affine.

(École polytechnique)

▷ **Solution.**

1. Notons G l'ensemble des fonctions f de E qui s'écrivent sous la forme $f = g''$ avec $g \in \mathrm{F}$. C'est un sous-espace vectoriel de E (comme image de F par l'application linéaire $g \mapsto g''$) et la question consiste à prouver que G est exactement l'orthogonal, au sens du produit scalaire $(f,g) \mapsto \int_a^b fg$, du sous-espace A formé des fonctions affines.

• Soit $f \in \mathrm{G}$ et $g \in \mathrm{F}$ tel que $f = g''$. On a alors

$$\int_a^b f(t)\mathrm{d}t = \int_a^b g''(t)\mathrm{d}t = [g'(t)]_a^b = 0,$$

et

$$\int_a^b tf(t)\mathrm{d}t = \int_a^b tg''(t)\mathrm{d}t = [tg'(t)]_a^b - \int_a^b g'(t)\mathrm{d}t = [tg'(t) - g(t)]_a^b = 0$$

par définition de F.

• Réciproquement, si f appartient à E et vérifie

$$\int_a^b f(t)\mathrm{d}t = \int_a^b tf(t)\mathrm{d}t = 0,$$

posons, pour tout $x \in [a,b]$,

$$f_1(x) = \int_a^x f(t)\mathrm{d}t \text{ et } g(x) = \int_a^x f_1(t)\mathrm{d}t.$$

La fonction g est $\mathcal{C}^2$ sur $[a,b]$ et vérifie $g'' = f$. On a, de plus

$$g'(a) = f_1(a) = 0\ ; \quad g'(b) = f_1(b) = \int_a^b f(t)\mathrm{d}t = 0\ ;$$
$$g(a) = 0\ ; \quad g(b) = [tf_1(t)]_a^b - \int_a^b tf(t)\mathrm{d}t = 0.$$

Donc g appartient à F et on a l'équivalence voulue *i.e.* $\mathrm{A}^\perp = \mathrm{G}$.

2. Cette seconde question demande de prouver que l'orthogonal de G est le sous-espace A formé des fonctions affines, autrement dit que $(\mathrm{A}^\perp)^\perp = \mathrm{A}$. L'inclusion $\mathrm{A} \subset (\mathrm{A}^\perp)^\perp = \mathrm{G}^\perp$ est banale. L'énoncé demande l'autre inclusion qui en dimension infinie n'est pas automatique. Elle est vérifiée ici car A est un sous-espace de E de dimension finie.

Puisque A est de dimension finie, la projection orthogonale sur A est bien définie. On prend $f \in (\mathrm{A}^\perp)^\perp$ qu'on écrit

$$f = u + v \quad \text{avec} \quad u \in \mathrm{A}, v \in \mathrm{A}^\perp.$$

On obtient $0 = (v, f) = (v, v)$, d'où on déduit $v = 0$ et $f \in \mathrm{A}$. On a l'inclusion voulue. ◁

Dans l'énoncé suivant, on s'intéresse aux fonctions orthogonales aux polynômes de degré inférieur ou égal à n, puis à l'orthogonal de l'espace des polynômes. Ce sont des questions très classiques.

1.17. Fonctions dont les n premiers moments sont nuls

Soit $f : [0,1] \longrightarrow \mathbb{R}$ une fonction continue

1. On suppose que pour tout $k \in [\![0, n]\!]$, $\int_0^1 f(x)x^k\mathrm{d}x = 0$. Montrer que f possède au moins $n+1$ zéros dans $[0,1]$.

2. On suppose que pour tout $k \in \mathbb{N}$, $\int_0^1 f(x)x^k\mathrm{d}x = 0$. Montrer que f est identiquement nulle.

(École polytechnique)

▷ **Solution.**

1. Par linéarité de l'intégrale, on a $\int_0^1 f\mathrm{P} = 0$ pour tout polynôme P de degré inférieur ou égal à n. Raisonnons par l'absurde et supposons que f ait moins de n zéros dans $[0,1]$. Entre deux zéros consécutifs, le signe de f reste constant d'après le théorème des valeurs intermédiaires puisque f est continue. Notons $\alpha_1 < \alpha_2 < \cdots < \alpha_p$ les zéros de f qui

correspondent à un changement de signe pour f. Nécessairement, $p \leqslant n$. Soit $\mathrm{P} = (\mathrm{X}-\alpha_1)(\mathrm{X}-\alpha_2)\dots(\mathrm{X}-\alpha_p) \in \mathbb{R}[\mathrm{X}]$. On a $\deg \mathrm{P} \leqslant n$ et P change de signe en même temps que f en les α_k. Par conséquent, le produit $f\mathrm{P}$ reste de signe constant. Comme $f\mathrm{P}$ est continue et $\int_0^1 f\mathrm{P} = 0$, $f\mathrm{P} = 0$. Pour tout $x \in [0,1]$, distinct des α_k, $\mathrm{P}(x) \neq 0$ et donc $f(x) = 0$. C'est absurde puisque f a moins de n zéros.

Couclusiou. La fonction f a au moins $n+1$ zéros dans $[0,1]$.

2. On utilise le théorème de Weierstrass : il existe une suite de polynôme $(\mathrm{P}_n)_{n\geqslant 0}$ qui converge uniformément vers f sur $[0,1]$ (on en trouvera une preuve dans l'exercice 2.18). Par linéarité de l'intégrale, $\int_0^1 f\mathrm{P}_n = 0$ pour tout n. Or, f étant bornée car continue sur un segment, la suite $(f\mathrm{P}_n)_{n\geqslant 0}$ converge uniformément vers f^2 sur $[0,1]$. En passant à la limite sous le signe intégral on a donc $\int_0^1 f^2 = 0$. Comme f^2 est continue et positive, cela impose $f^2 = 0$ et donc $f = 0$. ◁

Le résultat de la seconde question signifie que l'orthogonal du sous-espace des fonctions polynomiales pour le produit scalaire $(f,g) \mapsto \int_0^1 fg$ est réduit à $\{0\}$: c'est donc un exemple de sous-espace d'un espace préhilbertien qui n'a pas de supplémentaire orthogonal.

L'énoncé suivant est encore une variation sur le même thème. On y utilise le résultat de la question 2 mais sur un segment $[a,b]$ quelconque.

1.18. Recherche d'orthogonal

Soit $y : [a,b] \longrightarrow \mathbb{R}$ une fonction continue.

1. On suppose que pour toute fonction u de classe $\mathcal{C}^\infty$ telle que $\int_a^b u(t)\mathrm{d}t = 0$, on a $\int_a^b u(t)y(t)\mathrm{d}t = 0$. Montrer que y est constante.

2. Que dire si $\int_a^b u(t)y(t)\mathrm{d}t = 0$ pour toute fonction u de classe $\mathcal{C}^\infty$ telle que $u(a) = u(b)$?

(École polytechnique)

▷ **Solution.**

Notons $\mathrm{E} = \mathcal{C}^\infty([a,b],\mathbb{R})$ et remarquons que si l'on suppose que $\int_a^b uy = 0$ pour toute fonction $u \in \mathrm{E}$, c'est-à-dire si on supprime les conditions restrictives sur u, on a clairement $y = 0$. En effet, l'hypothèse est en particulier vraie pour les polynômes. D'après le théorème de Weierstrass, il existe une suite $(\mathrm{P}_n)_{n\geqslant 0}$ de polynômes convergeant uniformément vers y sur $[a,b]$. Puisque y est bornée, la suite $(\mathrm{P}_n y)_{n\geqslant 0}$

converge uniformément vers y^2 et on a donc $\int_a^b y^2 = \lim_{n\to+\infty} \int_a^b yP_n = 0$, par hypothèse, puisqu'un polynôme est $\mathcal{C}^\infty$. On en déduit que $y = 0$ puisqu'elle est continue. On va se ramener à ce cas dans chacune des questions.

1. Considérons les formes linéaires sur E

$$I : u \longmapsto \int_a^b u \quad \text{et} \quad I_y : u \longmapsto \int_a^b yu.$$

L'hypothèse se traduit par $\operatorname{Ker} I \subset \operatorname{Ker} I_y$. Cela implique que I_y est proportionnel à I. Soit $\lambda \in \mathbb{R}$ tel que $I_y = \lambda I$ et $z = y - \lambda$. On a alors, pour tout $u \in E$, $\int_a^b uz = I_y(u) - \lambda I(u) = 0$. On en déduit d'après la remarque initiale que $z = 0$ et donc $y = \lambda$: la fonction y est constante. Il est clair, réciproquement, que toute fonction constante a la propriété voulue.

2. La condition $u(a) = u(b)$ s'écrit $\int_a^b u' = 0$. Soit $v \in E$ tel que $\int_a^b v(t)dt = 0$, et soit u la primitive de v qui s'annule en a (et donc en b). On a par hypothèse $\int_a^b uy = 0$. En considérant une primitive Y de y, on obtient, en intégrant par parties,

$$\int_a^b uy = [uY]_a^b - \int_a^b vY = -\int_a^b vY = 0.$$

Ceci est vrai pour toute fonction $v \in E$, telle que $\int_a^b v(t)dt = 0$, donc d'après la question précédente, la fonction Y est constante et sa dérivée y est nulle. ◁

1.19. Méthode de Gauss

Soit $n \geqslant 1$ et $L_n(X)$ le polynôme $\dfrac{d^n}{dX^n}(X^2-1)^n$.

1. Montrer que

$$\forall Q \in \mathbb{R}_{n-1}[X], \quad \int_{-1}^{1} Q(x)L_n(x)dx = 0.$$

2. Montrer que L_n admet n racines simples $x_1, \ldots, x_n$ qui se trouvent dans l'intervalle $]-1, 1[$.

3. Montrer qu'il existe $(\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n$ tel que

$$\forall Q \in \mathbb{R}_{2n-1}[X], \quad \int_{-1}^{1} Q(x)dx = \sum_{i=1}^{n} \alpha_i Q(x_i).$$

(École normale supérieure)

▷ **Solution.**

1. Posons $P_n = (X^2-1)^n$ et soit $Q \in \mathbb{R}_{n-1}[X]$. On a $L_n = P_n^{(n)}$. On calcule l'intégrale par parties. On a

$$\int_{-1}^{1} QL_n = \left[\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n Q(x)\right]_{-1}^{1} - \int_{-1}^{1} Q'(x)\frac{d^{n-1}}{dx^{n-1}}(x^2-1)^n dx.$$

Or, 1 et -1 sont des racines d'ordre n de $P_n = (X^2-1)^n$, donc sont racines simples de $P_n^{(n-1)} = \frac{d^{n-1}}{dX^{n-1}}(X^2-1)^n$ et le crochet est nul. On intègre encore $n-1$ fois par parties pour obtenir finalement

$$\int_{-1}^{1} Q(x)L_n(x)dx = -\int_{-1}^{1} Q'(x)P_n^{(n-1)} = \int_{-1}^{1} Q''(X)P_n^{(n-2)}$$
$$= \cdots = (-1)^n \int_{-1}^{1} Q^{(n)}P_n = 0$$

car $Q^{(n)} = 0$.

2. On peut résoudre cette question en appliquant le théorème de Rolle successivement à $P_n, P_n', \ldots, P_n^{(n-1)}$. Cependant, on peut aussi utiliser la question précédente. Notons $x_1 < x_2 < \cdots < x_p$ les racines de L_n qui sont dans l'intervalle $]-1,1[$ et de multiplicité impaire. Posons $Q(X) = (X-x_1)\ldots(X-x_p)$. La fonction polynôme $x \longmapsto Q(x)L_n(x)$ garde alors un signe constant sur l'intervalle $]-1,1[$ et ne s'annule qu'en les x_i. Si $p \leqslant n-1$ cela induit une contradiction avec la question précédente car $\int_{-1}^{1} QL_n$ est non nul. Comme $\deg L_n = n$ on a donc $p = n$ ce qui signifie que L_n admet n racines dans $]-1,1[$ qui sont toutes de multiplicité 1, c'est-à-dire simples.

3. Les applications $g_k : Q \longmapsto Q(x_k)$ sont des formes linéaires sur $\mathbb{R}_{2n-1}[X]$ de même que l'application $f : Q \longmapsto \int_{-1}^{1} Q$. La question consiste à prouver que $f \in \mathrm{Vect}(g_1, \ldots, g_n)$. Pour cela, il suffit que f soit nulle sur le sous-espace $\bigcap_{i=1}^{n} \mathrm{Ker}\, g_i$. Or, si Q est un polynôme de $\bigcap_{i=1}^{n} \mathrm{Ker}\, g_i$, on a $Q(x_i) = 0$ pour tout i, de sorte que Q est divisible

par le polynôme $(X-x_1)\dots(X-x_n)$ et donc par L_n qui lui est associé (d'après la question 2). On peut donc écrire $Q(X) = L_n(X)R(X)$ avec $\deg R \leqslant n-1$ et $f(Q) = \int_{-1}^{1} Q = \int_{-1}^{1} L_n R = 0$. ◁

La méthode de Gauss exposée dans l'exercice est une méthode de calcul approchée des intégrales. Pour toute fonction f continue sur $[-1,1]$ on prend $\sum_{k=1}^{n} a_k f(x_k)$ comme valeur approchée de $\int_{-1}^{1} f(x)\,dx$. On obtient la valeur exacte pour tout polynôme de degré $\leqslant 2n-1$. On remarque que la famille (P_n) est une base de $\mathbb{R}[X]$ formée de polynômes orthogonaux, car P_n est de degré n. Le polynôme $\frac{1}{2^n n!} P_n$ est appelé n-ième polynôme de Legendre.

La méthode se généralise au calcul de $\int_a^b f(x)w(x)\,dx$ où la fonction w est continue et strictement positive sur $]a,b[$. On considère le produit scalaire défini sur $\mathbb{R}[X]$ par $\langle P, Q\rangle = \int_a^b P(x)Q(x)w(x)\,dx$ et on construit une suite orthonormale de polynômes $(P_n)_{n\in\mathbb{N}}$ tels que $\deg P_n = n$. On peut démontrer que chaque P_n possède n racines distinctes sur $]a,b[$ et que, pour tout $n \in \mathbb{N}^*$, il existe des réels $a_1,\dots,a_n$ tels que $\int_a^b Q(x)w(x)\,dx = \sum_{k=1}^{n} a_k Q(x_k)$ pour tout polynôme Q de degré $\leqslant 2n-1$. On prend $\sum_{k=1}^{n} a_k f(x_k)$ comme valeur approchée de $\int_a^b f(x)w(x)\,dx$. L'erreur commise est donnée, pour une fonction de classe C^{2n}, par la formule $\int_a^b f(x)w(x)\,dx = \sum_{k=1}^{n} a_k f(x_k) + \frac{f^{(2n)}(c)}{(2n)!C_n^2}$, où C_n est le coefficient de X^n dans P_n et $c \in]a,b[$.

Les cinq exercices suivants font intervenir des sommes de Riemann.

1.20. Une somme de Riemann (1)

Calculer $\lim_{n\to+\infty} \frac{1}{n\sqrt{n}} \sum_{k=1}^{n} E(\sqrt{k})$ où E est la fonction partie entière.

(École polytechnique)

▷ **Solution.**

Pour tout $k \geqslant 0$, on a $\sqrt{k} - 1 \leqslant E(\sqrt{k}) \leqslant \sqrt{k}$. Posons pour $n \geqslant 0$, $u_n = \frac{1}{n\sqrt{n}} \sum_{k=1}^{n} E(\sqrt{k})$. On a

$$\frac{1}{n\sqrt{n}}\sum_{k=1}^{n}(\sqrt{k}-1)\leqslant u_n\leqslant\frac{1}{n\sqrt{n}}\sum_{k=1}^{n}\sqrt{k}$$

$$\frac{1}{n}\left(\sum_{k=1}^{n}\sqrt{\frac{k}{n}}\right)-\frac{1}{\sqrt{n}}\leqslant u_n\leqslant\frac{1}{n}\left(\sum_{k=1}^{n}\sqrt{\frac{k}{n}}\right).$$

Comme $x\longmapsto\sqrt{x}$ est continue sur $[0,1]$, nous reconnaissons dans $\frac{1}{n}\left(\sum_{k=1}^{n}\sqrt{\frac{k}{n}}\right)$ une somme de Riemann qui converge vers $\int_0^1\sqrt{t}dt=\frac{2}{3}$ lorsque n tend vers l'infini. Nous avons donc encadré u_n par deux suites qui convergent vers $\frac{2}{3}$. D'après le théorème d'encadrement,

$$\boxed{\lim_{n\to+\infty}u_n=\frac{2}{3}}.\ \triangleleft$$

1.21. Une somme de Riemann (2)

Soit $f:\mathbb{R}\to\mathbb{R}$ de classe $\mathcal{C}^2$. Quelle est la limite de

$$u_n=\frac{1}{n+1}\sum_{k=0}^{n}f\left(\frac{k}{n+1}\right)f'\left(\frac{k+1}{n+1}\right)\ ?$$

(École normale supérieure)

▷ **Solution.**

Il est naturel de comparer u_n à la suite v_n définie pour $n\in\mathbb{N}$ par

$$v_n=\frac{1}{n+1}\sum_{k=0}^{n}f\left(\frac{k}{n+1}\right)f'\left(\frac{k}{n+1}\right)$$

dans laquelle on reconnaît une somme de Riemann qui converge vers $\int_0^1 ff'=\frac{f(1)^2-f(0)^2}{2}$. Considérons la différence

$$u_n-v_n=\frac{1}{n+1}\sum_{k=0}^{n}f\left(\frac{k}{n+1}\right)\left(f'\left(\frac{k+1}{n+1}\right)-f'\left(\frac{k}{n+1}\right)\right).$$

Les fonctions f et f'' sont continues sur le compact $[0,1]$, donc bornées respectivement par un réel M_0 et un réel M_2. En vertu de l'inégalité des accroissements finis, il vient

$$|u_n-v_n|\leqslant\frac{1}{n+1}\sum_{k=0}^{n}\left|f\left(\frac{k}{n+1}\right)\right|\frac{M_2}{n+1}\leqslant\frac{1}{n+1}\sum_{k=0}^{n}\frac{M_0M_2}{n+1}\leqslant\frac{M_0M_2}{n+1}.$$

On en déduit que $u_n - v_n$ converge vers 0 et par conséquent

$$\lim_{n\to+\infty} \frac{1}{n+1} \sum_{k=0}^{n} f\left(\frac{k}{n+1}\right) f'\left(\frac{k+1}{n+1}\right) = \frac{f(1)^2 - f(0)^2}{2}. \quad \triangleleft$$

Le cours démontre la convergence des sommes de Riemann vers l'intégrale d'une fonction continue. Un problème intéressant consiste à chercher un développement asymptotique de ces sommes de Riemann en mettant une hypothèse de régularité plus forte sur f. L'exercice suivant, classique, donne les deux premiers termes du développement lorsque la fonction est supposée de classe $\mathcal{C}^2$.

1.22. Étude asymptotique des sommes de Riemann

Soit $f : [a,b] \longrightarrow \mathbb{R}$ une fonction de classe $\mathcal{C}^1$. On pose pour $n \in \mathbb{N}^*$,

$$u_n = \frac{b-a}{n} \sum_{k=1}^{n} f\left(a + k\frac{b-a}{n}\right).$$

1. Étudier la limite de $v_n = n\left(u_n - \int_a^b f(t)\mathrm{d}t\right)$.

2. On suppose maintenant f de classe $\mathcal{C}^2$. Trouver un développement asymptotique de u_n en $+\infty$ à la précision $o\left(\frac{1}{n^2}\right)$.

(École normale supérieure, École polytechnique)

▷ **Solution.**

1. On remarque que, d'après le cours, la suite $(u_n)_{n\geqslant 1}$ converge vers $\int_a^b f(t)\mathrm{d}t$, puisqu'il s'agit d'une somme de Riemann. On a donc affaire à une forme indéterminée.

Pour tout $n \geqslant 1$ et tout $k \geqslant 1$, posons $a_k = a + kh$ avec $h = \frac{b-a}{n}$. Transformons l'expression de v_n à l'aide d'une primitive F de la fonction continue f :

$$v_n = n \sum_{k=1}^{n} \left(hf(a_k) - \int_{a_{k-1}}^{a_k} f(t)\mathrm{d}t\right) = n \sum_{k=1}^{n} (h\mathrm{F}'(a_k) - \mathrm{F}(a_k) + \mathrm{F}(a_{k-1}))$$

$$= n \sum_{k=1}^{n} (\mathrm{F}(a_{k-1}) - \mathrm{F}(a_k) + h\mathrm{F}'(o_k)) = n \sum_{k=1}^{n} \frac{h^2}{2}\mathrm{F}''(c_k)$$

où $c_k \in]a_{k-1}, a_k[$ vérifie

$$\mathrm{F}(a_{k-1}) = \mathrm{F}(a_k) + \mathrm{F}'(a_k)(-h) + \mathrm{F}''(c_k)\frac{(-h)^2}{2},$$

l'existence de c_k étant justifiée par la formule de Taylor-Lagrange (car $a_{k-1} - a_k = -h$). Ainsi, on a

$$v_n = \frac{(b-a)^2}{2n}\sum_{k=1}^{n} f'(c_k) \xrightarrow[n\to+\infty]{} \frac{b-a}{2}\int_a^b f'(t)\mathrm{d}t,$$

puisqu'on reconnaît une somme de Riemann. Comme f est de classe $\mathcal{C}^1$, il reste

$$\boxed{\lim_{n\to+\infty} v_n = \frac{(b-a)(f(b)-f(a))}{2}}.$$

2. Dans la question précédente, on a obtenu le développement asymptotique suivant de la somme de Riemann u_n :

$$u_n = \int_a^b f(t)\mathrm{d}t + (f(b)-f(a))\frac{b-a}{2n} + o\left(\frac{1}{n}\right).$$

Pour obtenir le terme suivant du développement, on va pousser à l'ordre suivant la formule de Taylor-Lagrange appliquée à F entre a_k et a_{k-1} : il existe pour tout k, $d_k \in]a_{k-1}, a_k[$ tel que

$$\mathrm{F}(a_{k-1}) = \mathrm{F}(a_k) - h\mathrm{F}'(a_k) + \mathrm{F}''(a_k)\frac{h^2}{2} - \mathrm{F}'''(d_k)\frac{h^3}{6}.$$

Cela s'écrit aussi

$$hf(a_k) = \int_{a_{k-1}}^{a_k} f + f'(a_k)\frac{h^2}{2} - f''(d_k)\frac{h^3}{6}.$$

En sommant de $k = 1$ à n, on obtient

$$u_n = \int_a^b f + \frac{(b-a)}{2n}\left(\frac{b-a}{n}\sum_{k=1}^{n} f'(a_k)\right) - \frac{(b-a)^2}{6n^2}\left(\frac{b-a}{n}\sum_{k=1}^{n} f''(d_k)\right).$$

Or, d'après la question précédente,

$$\frac{b-a}{n}\sum_{k=1}^{n} f'(a_k) = \int_a^b f' + (f'(b)-f'(a))\frac{b-a}{2n} + o\left(\frac{1}{n}\right).$$

Ainsi, il vient

$$u_n = \int_a^b f + (f(b)-f(a))\frac{b-a}{2n} + (f'(b)-f'(a))\frac{(b-a)^2}{4n^2} - \frac{(b-a)^2}{6n^2}\int_a^b f'' + o\left(\frac{1}{n^2}\right),$$

ce qui donne en regroupant les termes en $\frac{1}{n^2}$,

$$u_n = \int_a^b f + (f(b) - f(a))\frac{b-a}{2n} + (f'(b) - f'(a))\frac{(b-a)^2}{12n^2} + o\left(\frac{1}{n^2}\right). \quad \triangleleft$$

La formule d'Euler-Mac Laurin permet d'obtenir un développement asymptotique des sommes de Riemann à un ordre arbitraire. Ce développement fait intervenir les nombres de Bernoulli.

1.23. Découpage en morceaux de même aire

Soit f une fonction continue et > 0 sur $[a, b]$ et $n \in \mathbb{N}^*$.

1. Montrer qu'il existe une unique subdivision $(x_0, x_1, \ldots, x_n)$ de $[a, b]$ telle que, pour k de $[\![1, n]\!]$,

$$\int_{x_{k-1}}^{x_k} f(t)\mathrm{d}t = \frac{1}{n}\int_a^b f(t)\mathrm{d}t.$$

2. Trouver la limite, quand n tend vers $+\infty$, de $\frac{1}{n}\sum_{k=0}^{n} f(x_k)$.

(École polytechnique)

▷ **Solution.**

1. Soit F la fonction définie sur $[a, b]$ par $\mathrm{F}(x) = \int_a^x f(t)\mathrm{d}t$. Elle a pour dérivée f qui est strictement positive par hypothèse. Donc F est une bijection strictement croissante de $[a, b]$ sur $[0, \mathrm{I}]$, où $\mathrm{I} = \int_a^b f$.

On veut $x_0 = a$, $x_n = b$ et $\mathrm{F}(x_k) - \mathrm{F}(x_{k-1}) = \frac{1}{n}\mathrm{I}$ pour tout entier $k \in [\![1, n]\!]$. Cela équivaut à $\mathrm{F}(x_k) = \frac{k}{n}\mathrm{I}$ pour tout k de $[\![1, n]\!]$. Comme $\frac{k}{n}\mathrm{I}$ appartient à $[0, \mathrm{I}]$, il existe, par bijectivité de F, un unique réel $x_k \in [a, b]$ tel que $\mathrm{F}(x_k) = \frac{k}{n}\mathrm{I}$. Ceci montre l'existence et l'unicité de la subdivision $(x_0, x_1, \ldots, x_n)$.

2. Posons $u_n = \frac{1}{n}\sum_{k=0}^{n} f(x_k)$. On obtient

$$u_n = \frac{1}{n}\sum_{k=0}^{n} f\left(\mathrm{F}^{-1}\left(\frac{k}{n}\mathrm{I}\right)\right) = \frac{1}{n}\sum_{k=0}^{n}(f \circ \mathrm{F}^{-1})\left(\frac{k}{n}\mathrm{I}\right).$$

On reconnaît dans u_n une somme de Riemann. La fonction $f \circ \mathrm{F}^{-1}$ étant continue sur $[0, \mathrm{I}]$, on en déduit que

$$\lim_{n\to+\infty} u_n = \frac{1}{1}\int_0^1 (f\circ \mathrm{F}^{-1})(t)\mathrm{d}t.$$

Par le changement de variable $u = \mathrm{F}^{-1}(t)$, on obtient $\int_0^1 f\circ\mathrm{F}^{-1} = \int_a^b f^2$, et finalement

$$\boxed{\lim_{n\to+\infty} u_n = \frac{\int_a^b f^2}{\int_a^b f}}. \quad \lhd$$

L'énoncé suivant, qui fait aussi appel aux sommes de Riemann, est nettement plus technique que les précédents.

1.24. Estimation d'une suite de coefficients binomiaux

1. Soit f une fonction continue de $[0,1]$ dans $\mathbb{R}$. Déterminer

$$\lim_{n\to+\infty} \prod_{k=1}^{n} \left(1 + \frac{1}{n} f\left(\frac{k}{n}\right)\right).$$

2. Soit (n_i) une suite d'entiers pairs, (μ_i) une suite d'entiers. On suppose que ces deux suites tendent vers $+\infty$, que $\mu_i \leqslant n_i$ pour tout i et que $\dfrac{\mu_i - \frac{n_i}{2}}{\sqrt{n_i}}$ a une limite $\lambda > 0$ quand i tend vers $+\infty$. Trouver un équivalent de $\mathrm{C}_{n_i}^{\mu_i}$.

(École normale supérieure)

Solution.

1. Posons $u_n = \prod_{k=1}^{n} \left(1 + \frac{1}{n} f\left(\frac{k}{n}\right)\right)$. Pour $n > \|f\|_\infty$, tous les termes du produit sont strictement positifs et on peut donc considérer $\ln u_n = \sum_{k=1}^{n} \ln\left(1 + \frac{1}{n} f\left(\frac{k}{n}\right)\right)$. Comme $\left|\frac{1}{n} f\left(\frac{k}{n}\right)\right| \leqslant \frac{\|f\|_\infty}{n}$, pour tout $k \in [\![1, n]\!]$ et que $\ln(1+x) \underset{x\to 0}{\sim} x$, on compare $\ln u_n$ à $v_n = \frac{1}{n}\sum_{k=1}^{n} f\left(\frac{k}{n}\right)$. On reconnaît dans v_n une somme de Riemann de f et on a donc $\lim_{n\to+\infty} v_n = \int_0^1 f(t)\,\mathrm{d}t$.

Pour tout $|x| \leqslant \frac{1}{2}$, la formule de Taylor-Lagrange à l'ordre 2 donne l'existence de $c \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ tel que $\ln(1+x) = x + \frac{x^2}{2} \times \frac{-1}{(c+1)^2}$, ce qui implique $|\ln(1+x) - x| \leqslant 2x^2$. On en déduit que, si $n \geqslant 2\|f\|_\infty$,

$$|\ln u_n - v_n| \leqslant \sum_{k=1}^{n} \left| \ln\left(1 + \frac{1}{n} f\left(\frac{k}{n}\right)\right) - \frac{1}{n} f\left(\frac{k}{n}\right) \right|$$
$$\leqslant \sum_{k=1}^{n} \frac{2}{n^2} \left(f\left(\frac{k}{n}\right)\right)^2 \leqslant \frac{2\|f\|_\infty^2}{n}.$$

Ainsi, $\ln u_n - v_n$ tend vers 0 et $\ln u_n$ a même limite que v_n donc

$$\lim_{n \to +\infty} u_n = \exp\left(\int_0^1 f(t)\,\mathrm{d}t\right).$$

2. Posons, pour tout i, $m_i = \frac{\mu_i}{2}$. Puisque μ_i est équivalent à m_i et qu'il est aisé d'obtenir un équivalent de $\mathrm{C}_{2m_i}^{m_i}$ par la formule de Stirling, comparons $\mathrm{C}_{2m_i}^{\mu_i}$ à $\mathrm{C}_{2m_i}^{m_i}$. Pour i assez grand, on a $\mu_i > m_i$, puisque $\mu_i - m_i \underset{i \to \infty}{\sim} \lambda\sqrt{2m_i}$. On a donc $m_i > 2m_i - \mu_i$ et on obtient

$$\frac{\mathrm{C}_{2m_i}^{\mu_i}}{\mathrm{C}_{2m_i}^{m_i}} = \frac{(m_i!)^2}{\mu_i!(2m_i - \mu_i)!} = \frac{m_i(m_i-1)\dots(2m_i - \mu_i + 1)}{(m_i+1)(m_i+2)\dots\mu_i}$$
$$= \frac{\displaystyle\prod_{k=0}^{\mu_i - m_i - 1} (m_i - k)}{\displaystyle\prod_{k=1}^{\mu_i - m_i} (m_i + k)} = \frac{\displaystyle\prod_{k=1}^{\mu_i - m_i - 1} \left(1 - \frac{k}{m_i}\right)}{\displaystyle\prod_{k=1}^{\mu_i - m_i} \left(1 + \frac{k}{m_i}\right)}.$$

Comme $\mu_i - m_i \underset{i \to \infty}{\sim} \lambda\sqrt{2m_i}$, le numérateur est approximativement égal à $\displaystyle\prod_{k=1}^{n} \left(1 - \frac{k}{m_i}\right)$ ou encore à $\displaystyle\prod_{k=1}^{n} \left(1 - \frac{2\lambda^2 k}{n^2}\right) = \prod_{k=1}^{n} \left(1 - \frac{1}{n}\,\frac{2\lambda^2 k}{n}\right)$, où $n = \mathrm{E}(\lambda\sqrt{2m_i})$ et on retrouve une expression semblable à celle de la question 1, avec $f : x \longmapsto -2\lambda^2 x$. Il en serait de même pour le dénominateur. Plutôt que de justifier ces approximations successives, il est plus simple de reprendre directement la démarche de la première question. On pose

$$\mathrm{N}_i = \prod_{k=1}^{\mu_i - m_i - 1} \left(1 - \frac{k}{m_i}\right) \quad \text{et} \quad \mathrm{D}_i = \prod_{k=1}^{\mu_i - m_i} \left(1 + \frac{k}{m_i}\right).$$

Pour i assez grand, on a

$$\ln(\mathrm{N}_i) = \sum_{k=1}^{\mu_i - m_i - 1} \ln\left(1 - \frac{k}{m_i}\right).$$

On obtient, comme dans la question 1, que si i est assez grand pour que l'on ait $\dfrac{\mu_i - m_i - 1}{m_i} \leqslant \dfrac{1}{2}$,

$$\left|\ln(\mathrm{N}_i) + \sum_{k=1}^{\mu_i - m_i - 1} \frac{k}{m_i}\right| \leqslant \sum_{k=1}^{\mu_i - m_i - 1} \frac{2k^2}{m_i^2}$$

$$\leqslant \frac{2(\mu_i - m_i - 1)^3}{m_i^2} \underset{i\to\infty}{\sim} \frac{2\lambda^3(2m_i)^{\frac{3}{2}}}{m_i^3} \xrightarrow[i\to\infty]{} 0.$$

Comme

$$\sum_{k=1}^{\mu_i - m_i - 1} \frac{k}{m_i} = \frac{(\mu_i - m_i - 1))(\mu_i - m_i)}{2m_i} \underset{i\to\infty}{\sim} \frac{2\lambda^2 m_i}{2m_i} \xrightarrow[i\to\infty]{} \lambda^2,$$

on en déduit que $\lim_{i\to\infty} \mathrm{N}_i = e^{-\lambda^2}$. On trouve de même $\lim_{i\to\infty} \mathrm{D}_i = e^{\lambda^2}$ et donc

$$\lim_{i\to+\infty} \frac{\mathrm{C}_{2m_i}^{\mu_i}}{\mathrm{C}_{2m_i}^{m_i}} = e^{-2\lambda^2}.$$

Reste à trouver un équivalent de $\mathrm{C}_{2m_i}^{m_i} = \dfrac{(2m_i)!}{(m_i!)^2}$. La formule de Stirling $n! \underset{n\to+\infty}{\sim} n^n e^{-n}\sqrt{2\pi n}$ donne

$$\mathrm{C}_{2m_i}^{m_i} \underset{i\to\infty}{\sim} \frac{(2m_i)^{2m_i} e^{-2m_i}\sqrt{4\pi m_i}}{m_i^{2m_i} e^{-2m_i}(2\pi m_i)} \underset{i\to\infty}{\sim} \frac{2^{2m_i}}{\sqrt{\pi m_i}}.$$

En remplaçant m_i par $\dfrac{n_i}{2}$, on obtient

$$\boxed{\mathrm{C}_{n_i}^{\mu_i} \underset{i\to\infty}{\sim} \frac{e^{-2\lambda^2} 2^{n_i + \frac{1}{2}}}{\sqrt{\pi n_i}}}. \quad \triangleleft$$

Les deux exercices suivants exploitent la densité dans l'espace des fonctions continues sur $[a,b]$ de l'espace des fonctions en escalier. On commence par démontrer le résultat pour une fonction en escalier.

1.25. Lemme de Riemann-Lebesgue

Soit f continue de $[a,b]$ dans $\mathbb{C}$. Montrer que $\mathrm{I}_n = \displaystyle\int_a^b f(t)e^{int}\,dt$ tend vers 0.

(École polytechnique)

Solution.

Nous commençons par démontrer le résultat pour une fonction f en escalier. Nous l'étendrons ensuite aux fonctions continues par un argu-

ment de densité. Si f est la fonction caractéristique d'un intervalle $[\alpha, \beta]$ inclus dans $[a, b]$, on a, pour tout $n \in \mathbb{N}$,

$$\mathrm{I}_n = \int_\alpha^\beta e^{int}\,\mathrm{d}t = \frac{1}{n}(e^{in\beta} - e^{in\alpha}) \quad \text{et} \quad |\mathrm{I}_n| \leqslant \frac{2}{n}.$$

La suite $(\mathrm{I}_n)_{n\geqslant 0}$ tend bien vers 0.

Si f est une fonction en escalier, elle s'écrit comme combinaison linéaire de fonctions caractéristiques d'intervalles et la propriété résulte de la linéarité de l'intégrale et de la limite. Enfin, si f est supposée seulement continue, pour tout $\varepsilon > 0$, il existe une fonction en escalier g telle que $\|f - g\|_\infty \leqslant \varepsilon$. On a alors, pour tout $n \in \mathbb{N}$,

$$\begin{aligned}
|\mathrm{I}_n| &\leqslant \left|\int_a^b (f(t) - g(t))e^{int}\,\mathrm{d}t\right| + \left|\int_a^b g(t)e^{int}\,\mathrm{d}t\right| \\
&\leqslant \int_a^b |f(t) - g(t)|\,\mathrm{d}t + \left|\int_a^b g(t)e^{int}\,\mathrm{d}t\right| \\
&\leqslant (b-a)\varepsilon + \left|\int_a^b g(t)e^{int}\,\mathrm{d}t\right|.
\end{aligned}$$

Le deuxième terme tendant vers 0, on aura $|\mathrm{I}_n| \leqslant 2(b-a)\varepsilon$ pour n assez grand et donc $\lim\limits_{n\to+\infty} \mathrm{I}_n = 0$. ◁

La propriété est vérifiée, avec la même démonstration, pour toute fonction réglée puisqu'une telle fonction est encore limite uniforme d'une suite de fonctions en escalier (cf. exercice 4.6 du tome 1).

Notons que lorsque f est de classe $\mathcal{C}^1$, on peut démontrer le résultat directement en écrivant

$$\mathrm{I}_n = \left[f(t)\frac{e^{int}}{in}\right]_a^b - \frac{1}{in}\int_a^b f'(t)e^{int}\,dt$$

et donc

$$|\mathrm{I}_n| \leqslant \frac{1}{n}\left(2\|f\|_\infty + \int_a^b |f'(t)|\,dt\right).$$

À partir de là, le théorème de Weierstrass, par un argument de densité, permet d'étendre le résultat aux fonctions continues, ce qui fournit une autre solution de l'exercice.

1.26. Deuxième formule de la moyenne

Soient $f, g : [a,b] \longrightarrow \mathbb{R}$ continues par morceaux avec f positive et décroissante. Prouver l'existence de $c \in [a,b]$ tel que

$$\int_a^b f(t)g(t)\mathrm{d}t = f(a^+)\int_a^c g(t)\mathrm{d}t,$$

où $f(a^+)$ désigne la limite à droite en a de f.

(École polytechnique)

▷ **Solution.**

• Notons $\mathrm{G} : x \in [a,b] \longmapsto \int_a^x g$. La fonction G est continue (car lipschitzienne de rapport $\|g\|_\infty$). Il suffit de prouver d'après le théorème des valeurs intermédiaires que

$$mf(a^+) \leqslant \int_a^b fg \leqslant \mathrm{M}f(a^+),$$

où $m = \min \mathrm{G}$ et $\mathrm{M} = \max \mathrm{G}$ (G est continue donc est bornée et atteint ses bornes sur le compact $[a,b]$).

• Nous allons d'abord le vérifier lorsque f est une fonction en escalier. On considère alors $a = x_0 < x_1 < \cdots < x_n = b$ subdivision compatible avec f et on note f_i la valeur constante de f sur $]x_{i-1}, x_i[$ $(1 \leqslant i \leqslant n)$. Remarquons dès maintenant que $f_1 = f(a^+)$. On a alors

$$\begin{aligned}
\int_a^b fg &= \sum_{i=1}^n \int_{x_{i-1}}^{x_i} f_i g = \sum_{i=1}^n f_i \int_{x_{i-1}}^{x_i} g \\
&= \sum_{i=1}^n f_i(\mathrm{G}(x_i) - \mathrm{G}(x_{i-1})) = \sum_{i=1}^n f_i \mathrm{G}(x_i) - \sum_{i=1}^n f_i \mathrm{G}(x_{i-1}) \\
&= \boxed{j = i-1} \sum_{i=1}^n f_i \mathrm{G}(x_i) - \sum_{j=0}^{n-1} f_{j+1}\mathrm{G}(x_j) \\
&= \sum_{i=1}^{n-1} (f_i - f_{i+1})\mathrm{G}(x_i) - f_1\mathrm{G}(a) + f_n\mathrm{G}(b) \\
&= \sum_{i=1}^{a-1} (f_i - f_{i+1})\mathrm{G}(x_i) + f_n\mathrm{G}(b).
\end{aligned}$$

Comme pour tout $1 \leqslant i \leqslant n-1$, $f_i - f_{i+1} \geqslant 0$ (f est décroissante), on a $m(f_i - f_{i+1}) \leqslant (f_i - f_{i+1})\mathrm{G}(x_i) \leqslant \mathrm{M}(f_i - f_{i+1})$. Comme f est positive, $f_n \geqslant 0$ et

$$m\left(\sum_{i=1}^{n-1}(f_i - f_{i+1}) + f_n\right) \leqslant \int_a^b fg \leqslant \mathrm{M}\left(\sum_{i=1}^{a-1}(f_i - f_{i+1}) + f_a\right).$$

Les f_i s'éliminant en diagonale, il reste

$mf_1 \leqslant \int_a^b fg \leqslant \mathrm{M}f_1$, ce qui s'écrit $mf(a^+) \leqslant \int_a^b fg \leqslant \mathrm{M}f(a^+)$.

• Passons au cas général : f est supposée continue par morceaux. On considère une subdivision $a = a_0 < a_1 < \cdots < a_n = b$ compatible avec f et on note f_i l'application continue sur $[a_{i-i}, a_i]$ qui coïncide avec f sur $]a_{i-1}, a_i[$ $(1 \leqslant i \leqslant n)$. On se donne $\varepsilon > 0$. D'après le théorème de Heine, chaque f_i est uniformément continue. Comme elles sont en nombre fini, il existe η un module d'uniforme continuité pour ε valable pour toutes les applications f_i. On prend dans chaque intervalle $[a_{i-1}, a_i]$ une subdivision de pas inférieur ou égal à η. En en prenant la réunion, on obtient une subdivision $\mathrm{S} = \{a = x_0 < x_1 < \cdots < x_n = b\}$ de l'intervalle $[a, b]$ qui est de pas inférieur ou égal à η.

On considère l'application $h : [a, b] \longrightarrow \mathbb{R}$ en escalier qui vaut $f\left(\frac{x_{i-1} + x_i}{2}\right)$ sur $]x_{i-1}, x_i[$ et $f(x_i)$ en x_i. Alors h est décroissante et positive, et par construction, $\|f - h\|_\infty \leqslant \varepsilon$.

D'après le cas déjà étudié, on a

$$mh(a^+) \leqslant \int_a^b hg \leqslant \mathrm{M}h(a^+).$$

Or, $\int_a^b |f - h|g \leqslant \varepsilon \|g\|_\infty (b - a)$, si bien qu'on a

$$mh(a^+) - \varepsilon\|g\|_\infty(b-a) \leqslant \int_a^b fg \leqslant \mathrm{M}h(a^+) + \varepsilon\|g\|_\infty(b-a).$$

D'autre part, $h(a^+) = f\left(\frac{x_0 + x_1}{2}\right)$ et donc, pour tout $x \in]x_0, x_1[$, on a par uniforme continuité $|h(a^+) - f(x)| \leqslant \varepsilon$, ce qui donne en faisant tendre x vers a, $|h(a^+) - f(a^+)| \leqslant \varepsilon$. On en déduit que

$$mf(a^+) - \varepsilon(m + \|g\|_\infty(b-a)) \leqslant \int_a^b fg \leqslant \mathrm{M}f(a^+) + \varepsilon(\mathrm{M} + \|g\|_\infty(b-a)).$$

Cette inégalité étant vraie pour tout $\varepsilon > 0$, on peut faire tendre ε vers 0, ce qui donne l'inégalité attendue

$$mf(a^+) \leqslant \int_a^b fg \leqslant \mathrm{M}f(a^+).$$

Conclusion. Il existe donc $c \in [a, b]$ tel que

$$\boxed{\int_a^b f(t)g(t)\mathrm{d}t = f(a^+)\int_a^c g(t)\mathrm{d}t}. \lhd$$

Notons que la démonstration est considérablement simplifiée si f est de classe $\mathcal{C}^1$ et g est continue. On peut alors écrire en gardant les notations de l'exercice,

$$\int_a^b f(t)g(t)\,dt = \mathrm{G}(b)f(b) - \int_a^b \mathrm{G}(t)f'(t)\,dt.$$

Comme f et $-f'$ sont positives, on obtient $mf(b) \leqslant \mathrm{G}(b)f(b) \leqslant \mathrm{M}f(b)$ et

$$m(f(a)-f(b)) = m\int_a^b (-f'(t))\,dt \leqslant \int_a^b (-\mathrm{G}(t)f'(t))\,dt \leqslant \mathrm{M}\int_a^b (-f'(t))\,dt = \mathrm{M}(f(a)-f(b))$$

et en additionnant $mf(a) \leqslant \int_a^b f(t)g(t)\,dt \leqslant \mathrm{M}f(a)$.

Les trois exercices suivants étudient la notion d'équirépartition d'une suite de réels du segment $[0,1]$. De manière informelle une telle suite est dite équirépartie si pour tout intervalle I inclus dans $[0,1]$ la probabilité pour un terme de la suite d'être dans I est égale à la longueur de I. Le premier exercice précise cette définition et étudie quelques généralités.

1.27. Suites équiréparties : généralités

Soit $(u_n)_{n\geqslant 1}$ une suite de points de $[0,1]$. Pour $0 \leqslant a < b \leqslant 1$, on pose

$$\mathrm{S}_n(a,b) = \mathrm{Card}\{k \in [\![1,n]\!],\ u_k \in [a,b[\}.$$

On dit que la suite $(u_n)_{n\geqslant 1}$ est *équirépartie* si $\dfrac{\mathrm{S}_n(a,b)}{n}$ tend vers $b-a$ pour tout $0 \leqslant a < b \leqslant 1$.

1. Montrer que si $(u_n)_{n\geqslant 1}$ est équirépartie l'ensemble des u_n est dense dans $[0,1]$. La réciproque est-elle vraie ?

2. Pour tout $n \in \mathbb{N}^*$, on pose

$$\mathrm{D}_n = \sup_{0\leqslant a<b\leqslant 1} \left|\frac{\mathrm{S}_n(a,b)}{n} - (b-a)\right| \quad \text{et} \quad \mathrm{D}_n^* = \sup_{0<\alpha<1} \left|\frac{\mathrm{S}_n(0,\alpha)}{n} - \alpha\right|.$$

Montrer que $\mathrm{D}_n^* \leqslant \mathrm{D}_n \leqslant 2\mathrm{D}_n^*$.

3. Montrer que la suite $(u_n)_{n\geqslant 1}$ est équirépartie si et seulement si la suite $(\mathrm{D}_n)_{n\geqslant 1}$ tend vers 0.

(École polytechnique)

▷ **Solution.**

1. Supposons que $(u_n)_{n\geqslant 1}$ est équirépartie et considérons a, b deux réels tels que $0 \leqslant a < b \leqslant 1$. Comme $\dfrac{\mathrm{S}_n(a,b)}{n}$ tend vers $b-a > 0$, il existe

un rang N au-delà duquel $S_n(a,b) > 0$. Cela signifie qu'il existe au moins un indice k tel que u_k appartienne à $[a,b[$. L'ensemble $\{u_n,\ n \geqslant 1\}$ est donc dense dans le segment $[0,1]$. La réciproque n'est bien entendu pas vraie. Considérons par exemple la suite $(v_n)_{n\geqslant 1}$ définie par $v_{2n} = u_n$ et $v_{2n+1} = 0$, où $(u_n)_{n\geqslant 1}$ est toujours une suite équirépartie. L'ensemble des v_n contient $\{u_n,\ n \geqslant 1\}$ et est donc toujours dense dans $[0,1]$. En revanche

$$\frac{S_{2n}(0,1/2)}{2n} = \frac{n + S_n(0,1/2)}{2n} \xrightarrow[n\to+\infty]{} \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \neq \frac{1}{2},$$

et $(v_n)_{n\geqslant 1}$ n'est donc pas équirépartie.

Il n'est pas difficile de vérifier que la définition de l'équirépartition prise ici correspond bien à ce qui est annoncé avant l'exercice : si $(u_n)_{n\geqslant 1}$ est équirépartie et si I *est un intervalle quelconque (pas forcément semi-ouvert à droite), $\frac{1}{n}\,\mathrm{Card}\{k \in [\![1,n]\!], u_k \in \mathrm{I}\}$ tend vers la longueur de* I *quand n tend vers $+\infty$.*

2. L'inégalité $D_n^* \leqslant D_n$ est évidente : la borne supérieure est prise sur l'ensemble des intervalles du type $[0,\alpha[$ qui est inclus dans l'ensemble de tous les intervalles semi-ouverts. Prenons maintenant $0 \leqslant a < b \leqslant 1$. On a clairement $S_n(a,b) = S_n(0,b) - S_n(0,a)$. Ainsi,

$$\left|\frac{S_n(a,b)}{n} - (b-a)\right| \leqslant \left|\frac{S_n(0,b)}{n} - b\right| + \left|\frac{S_n(0,a)}{n} - a\right| \leqslant 2D_n^*.$$

En passant à la borne supérieure on a donc $D_n \leqslant 2D_n^*$.

3. Si D_n tend vers 0 la suite $\dfrac{S_n(a,b)}{n}$ tend vers $(b-a)$ pour tout $0 \leqslant a < b \leqslant 1$ et $(u_n)_{n\geqslant 1}$ est équirépartie.

Montrons la réciproque en raisonnant par l'absurde. On suppose donc que la suite (u_n) est équirépartie et que D_n ne tend pas vers 0. D'après la question précédente, D_n^* ne tend pas vers 0 non plus. On peut donc trouver $\varepsilon > 0$ et $\varphi : \mathbb{N} \to \mathbb{N}$ strictement croissante telle que $D^*_{\varphi(n)} \geqslant 2\varepsilon$ pour tout n. Par définition de $D^*_{\varphi(n)}$, on peut donc trouver un réel $\alpha_{\varphi(n)}$ de $]0,1[$ tel que

$$\left|\frac{S_{\varphi(n)}(0,\alpha_{\varphi(n)})}{\varphi(n)} - \alpha_{\varphi(n)}\right| \geqslant \varepsilon.$$

Par compacité du segment $[0,1]$ on peut, quitte à prendre une sous-suite de $(\alpha_{\varphi(n)})$, supposer directement que cette suite converge vers un réel α. Supposons $0 < \alpha < 1$ et soit $\eta > 0$ tel que $0 < \alpha - \eta < \alpha + \eta < 1$. Pour n assez grand, on a $\alpha - \eta \leqslant \alpha_{\varphi(n)} \leqslant \alpha + \eta$. On a forcément, pour les mêmes n,

$$S_{\varphi(n)}(0,\alpha-\eta) \leqslant S_{\varphi(n)}(0,\alpha_{\varphi(n)}) \leqslant S_{\varphi(n)}(0,\alpha+\eta).$$

On en déduit que pour n assez grand,

$$\frac{S_{\varphi(n)}(0,\alpha-\eta)}{\varphi(n)} \leqslant \frac{S_{\varphi(n)}(0,\alpha_{\varphi(n)})}{\varphi(n)} \leqslant \frac{S_{\varphi(n)}(0,\alpha+\eta)}{\varphi(n)}.$$

Comme le terme de gauche de cette inégalité tend vers $\alpha-\eta$ et le terme de droite vers $\alpha+\eta$ (car la suite est supposée équirépartie), on a pour n assez grand $\alpha-2\eta \leqslant \dfrac{S_{\varphi(n)}(0,\alpha_{\varphi(n)})}{\varphi(n)} \leqslant \alpha+2\eta$. Comme $\alpha-\eta \leqslant \alpha_{\varphi(n)} \leqslant \alpha+\eta$ on a finalement pour n assez grand,

$$-3\eta \leqslant \frac{S_{\varphi(n)}(0,\alpha_{\varphi(n)})}{\varphi(n)} - \alpha_{\varphi(n)} \leqslant 3\eta$$

et cela est contradictoire pour η assez petit (par exemple $\eta = \varepsilon/4$). On adapte sans mal ce raisonnement lorsque $\alpha = 0$ ou $\alpha = 1$. ◁

L'exercice suivant établit un critère d'équirépartition particulièrement efficace démontré par Weyl en 1916. On y utilise essentiellement la définition de l'intégrale d'une fonction continue.

1.28. Suites équiréparties : critère de Weyl

Soit $(u_n)_{n\geqslant 1}$ une suite de $[0,1]$. Pour $0 \leqslant a \leqslant b \leqslant 1$, on pose

$$X_n(a,b) = \mathrm{Card}\{k \in [\![1,n]\!], u_k \in [a,b]\}.$$

Prouver l'équivalence des propriétés suivantes :

(*i*) $\dfrac{X_n(a,b)}{n}$ tend vers $b-a$ pour tout couple (a,b) ;

(*ii*) pour toute fonction $f : [0,1] \longrightarrow \mathbb{R}$ continue, on a

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} f(u_k) = \int_0^1 f(t)\mathrm{d}t \ ;$$

(*iii*) pour tout $p \in \mathbb{N}^*$, on a $\lim\limits_{n\to+\infty} \dfrac{1}{n}\sum\limits_{k=1}^{n} e^{2i\pi p u_k} = 0$.

(**École normale supérieure**)

▷ **Solution.**

• Montrons que (*i*) $\Longrightarrow$ (*ii*). Notons que la quantité $\dfrac{X_n(a,b)}{n}$ est égale à $\dfrac{1}{n}\sum\limits_{k=1}^{n} \chi_{[a,b]}(u_k)$ où $\chi_{[a,b]}$ désigne la fonction caractéristique du segment

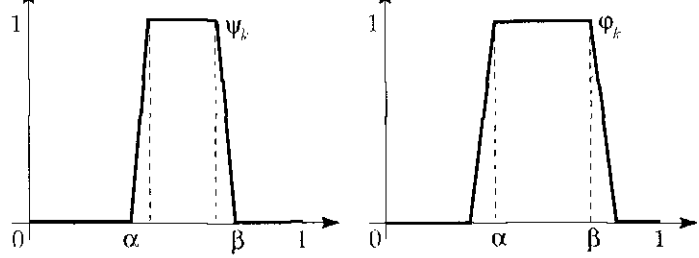
$[a,b]$ et que $\int_a^b \chi_{[a,b]} = b-a$. La propriété (ii) est donc vérifiée pour les fonctions caractéristiques d'un segment. En fait, toute fonction en escalier f sur $[0,1]$ est combinaison linéaire de fonctions caractéristiques de segments (éventuellement réduits à un point pour obtenir les valeurs de f aux points de discontinuité). Par linéarité, la propriété (ii) est alors vraie pour toute fonction en escalier. Montrons maintenant que (ii) est vérifiée pour les fonctions continues. Soit $f : [0,1] \longrightarrow \mathbb{R}$ une fonction continue et $\varepsilon > 0$. On sait qu'on peut trouver une fonction en escalier g qui approche f uniformément à ε près sur $[0,1]$, *i.e.* telle que $\|f-g\|_\infty \leqslant \varepsilon$. Grâce à l'inégalité triangulaire, pour tout $n \geqslant 1$, on peut majorer $\left|\frac{1}{n}\sum_{k=1}^{n} f(u_k) - \int_0^1 f\right|$ par

$$\left|\frac{1}{n}\sum_{k=1}^{n}(f(u_k)-g(u_k))\right| + \left|\frac{1}{n}\sum_{k=1}^{n} g(u_k) - \int_0^1 g\right| + \left|\int_0^1 g - \int_0^1 f\right|.$$

Le premier terme est majoré par ε de même que le dernier. Comme g est en escalier, le terme du milieu devient inférieur à ε au-delà d'un certain rang N. Bref, pour $n \geqslant \mathrm{N}$, $\left|\frac{1}{n}\sum_{k=1}^{n} f(u_k) - \int_0^1 f\right| \leqslant 3\varepsilon$ ce qui prouve (ii).

Bien entendu (ii) reste encore vraie pour une fonction continue par morceaux et même une fonction réglée.

• Montrons réciproquement que $(ii) \Longrightarrow (i)$. Une fonction caractéristique d'un segment I (distinct de $[0,1]$) présente au moins une discontinuité donc ne peut pas être obtenue comme limite uniforme d'une suite de fonctions continues. En fait on n'a pas besoin d'une approximation uniforme : il va suffire d'encadrer χ_{I} par deux suites de fonctions continues affines par morceaux qui convergent vers χ_{I} au sens de la norme intégrale. Prenons pour commencer un segment $\mathrm{I} = [\alpha,\beta]$ avec $0 < \alpha < \beta < 1$. On considère les suites de fonctions continues φ_k et ψ_k définies pour k suffisamment grand par les figures suivantes :



Ainsi, ψ_k est nulle sur les segments $[0,\alpha]$ et $[\beta,1]$ et vaut 1 sur le segment $[\alpha+1/k, \beta-1/k]$ et est affine sur les deux segments qui restent.

On observe que pour tout p assez grand, $\psi_p \leqslant \chi_1 \leqslant \varphi_p$. Il en résulte que pour tout $n \geqslant 1$, $\frac{1}{n}\sum_{k=1}^{n} \psi_p(u_k) \leqslant \frac{X_n(\alpha,\beta)}{n} \leqslant \frac{1}{n}\sum_{k=1}^{n} \varphi_p(u_k)$. Par hypothèse,

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \psi_p(u_k) = \int_0^1 \psi_p = \beta - \alpha - \frac{1}{p}$$

et

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} \varphi_p(u_k) = \int_0^1 \varphi_p = \beta - \alpha + \frac{1}{p}.$$

Soit $\varepsilon > 0$. Choisissons p tel que $\frac{1}{p} < \varepsilon$. Il existe N tel que pour $n \geqslant N$, $\left|\frac{X_n(\alpha,\beta)}{n} - \beta + \alpha\right| \leqslant 2\varepsilon$. Ainsi, $\frac{X_n(\alpha,\beta)}{n}$ converge vers $\beta - \alpha$ lorsque $0 < \alpha < \beta < 1$. Il est aisé d'adapter cela lorsque $\alpha = 0$ ou $\beta = 1$.

• La propriété (iii) résulte directement de (ii) puisque

$$\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} e^{2i\pi p u_k} = \int_0^1 \cos(2\pi p x)\mathrm{d}x + i\int_0^1 \sin(2\pi p x)\mathrm{d}x = 0.$$

• Montrons enfin que $(iii) \Longrightarrow (ii)$. Par linéarité, on a la propriété (ii) pour tout polynôme trigonométrique du type

$$x \longmapsto a_0 + \sum_{k=1}^{n} a_k \cos(2k\pi x) + \sum_{k=1}^{n} b_k \sin(2k\pi x).$$

D'après le théorème de Weierstrass trigonométrique, toute fonction continue $f : [0,1] \longrightarrow \mathbb{R}$ vérifiant $f(0) = f(1)$ est limite uniforme d'une suite de polynômes trigonométriques de ce type. Comme précédemment, on en déduit que (ii) est vérifiée pour une telle fonction continue. Si f ne vérifie pas la condition $f(0) = f(1)$ on peut, pour tout $\varepsilon > 0$, trouver une fonction continue g vérifiant $g(0) = g(1)$ et $\int_0^1 |f - g| \leqslant \varepsilon$. Comme dans l'implication $(ii) \Longrightarrow (i)$, cela suffit pour prouver que (ii) est aussi vraie pour f.

Les propositions (i), (ii) et (iii) sont donc toutes équivalentes. ◁

La proposition (i) correspond à l'équirépartition de la suite $(u_n)_{n\geqslant 1}$. Le critère de Weyl correspond à l'équivalence des assertions (i) et (iii).

L'exercice suivant en donne un exemple d'application avec l'étude des suites arithmétiques. On dit qu'une suite réelle $(x_n)_{n\geqslant 1}$ est équirépartie modulo 1 si la suite $u_n = x_n - \mathrm{E}(x_n)$ est équirépartie. Notons que comme les fonctions $e_p : x \longmapsto e^{2ip\pi x}$ sont toutes 1-périodiques, on a encore l'équivalence

$$(x_n)_{n\geqslant 1} \text{ équirépartie modulo } 1 \Longleftrightarrow \forall p \in \mathbb{N}^*,\ \lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^{n} e_p(x_k) = 0.$$

1.29. Suites équiréparties : un exemple

Soit $\theta > 0$. Montrer que la suite $(n\theta)_{n\geqslant 1}$ est équirépartie modulo 1 si et seulement si $\theta \notin \mathbb{Q}$.

(École polytechnique)

▷ **Solution.**

• Supposons $\theta \notin \mathbb{Q}$. Pour prouver que la suite $(n\theta)_{n\geqslant 1}$ est équirépartie on va utiliser le critère de Weyl, ce qui est facile ici car les sommes d'exponentielles se calculent (on a la somme des termes d'une progression géométrique). Plus précisément, avec les notations du préambule, on a pour tout $p \in \mathbb{N}^*$,

$$\frac{1}{n}\sum_{k=1}^{n} e_p(k\theta) = \frac{1}{n}\sum_{k=1}^{n} e_p(\theta)^k = \frac{e_p(\theta)}{n}\cdot\frac{e_p(\theta)^n - 1}{e_p(\theta) - 1} = \frac{e_p(\theta)}{n}\cdot\frac{e_p(n\theta) - 1}{e_p(\theta) - 1}$$

et ce calcul est possible car comme $\theta \notin \mathbb{Q}$, $e_p(\theta) \neq 1$ pour tout entier $p \in \mathbb{N}^*$. Comme $|e_p(n\theta)| = 1$ pour tout n, la suite $\dfrac{e_p(n\theta) - 1}{e_p(\theta) - 1}$ est bornée et $\dfrac{1}{n}\displaystyle\sum_{k=1}^{n} e_p(k\theta)$ tend vers 0 quand n tend vers $+\infty$. Donc si $\theta \notin \mathbb{Q}$, la suite $(n\theta)_{n\geqslant 1}$ est équirépartie modulo 1.

• Supposons que $\theta = \dfrac{p}{q} \in \mathbb{Q}$, avec p, q dans $\mathbb{N}^*$ premiers entre eux. On a, en notant $\{x\} = x - \mathrm{E}(x)$ la partie fractionnaire d'un réel x, $\{(n+q)\theta\} = \{n\theta + p\} = \{n\theta\}$. Donc la suite $\{n\theta\}$ est périodique de période q et n'est donc clairement pas équirépartie. ◁

Le lecteur trouvera une application de ce résultat dans l'exercice 1.13 du tome 1 d'algèbre qui étudie la distribution du premier chiffre de l'écriture en base 10 des puissances de 2. Notons que l'exercice ci-dessus a été posé comme seconde question d'un exercice qui demandait aussi de prouver le critère de Weyl. Il est possible de donner une preuve directe du résultat mais ce n'est pas facile.

Ici commence une série d'exercices sur le vaste thème des intégrales à paramètres. Le lecteur trouvera d'autres exercices sur ce thème, mais avec des intégrales généralisées, dans le troisième tome d'analyse.

1.30. Calcul d'une intégrale

Soient $a > 1$ et $b > 1$. Calculer $\displaystyle\int_0^{\pi} \ln\left(\frac{b - \cos x}{a - \cos x}\right) \mathrm{d}x$.

(École polytechnique)

Solution.

L'intégrande est bien continu sur $[0,\pi]$. Il semble vain de vouloir calculer une primitive directement. En revanche, on peut écrire

$$\mathrm{I} = \int_0^\pi \ln\left(\frac{b-\cos x}{a-\cos x}\right)\mathrm{d}x = \mathrm{F}(b) - \mathrm{F}(a),$$

où $\mathrm{F} : t \longmapsto \int_0^\pi \ln(t-\cos x)\mathrm{d}x$.

La fonction $(t,x) \in]1,+\infty[\times\mathbb{R} \longmapsto \ln(t-\cos x)$ étant de classe $\mathcal{C}^\infty$, on peut dériver sous le signe intégrale, si bien que pour $t > 1$,

$$\mathrm{F}'(t) = \int_0^\pi \frac{\mathrm{d}x}{t-\cos x}.$$

Cette nouvelle intégrale présente l'avantage de pouvoir se calculer. Pour expliciter $\mathrm{F}'(t)$, on fait le changement de variable classique $u = \tan\left(\frac{x}{2}\right)$:

$$\begin{aligned}\mathrm{F}'(t) &= \int_0^{+\infty} \frac{2\mathrm{d}u}{(1+u^2)\left(t-\dfrac{1-u^2}{1+u^2}\right)} = \int_0^{+\infty} \frac{2\mathrm{d}u}{(1+u^2)t-(1-u^2)} \\ &= \frac{2}{t+1}\int_0^{+\infty} \frac{\mathrm{d}u}{u^2+\dfrac{t-1}{t+1}} \\ &= \frac{2}{t+1}\left[\frac{1}{\sqrt{\dfrac{t-1}{t+1}}}\arctan\left(\frac{u}{\sqrt{\dfrac{t-1}{t+1}}}\right)\right]_0^{+\infty} = \frac{\pi}{\sqrt{t^2-1}}.\end{aligned}$$

Une primitive de $\dfrac{1}{\sqrt{t^2-1}}$ est $\operatorname{Argch} t = \ln(t+\sqrt{t^2-1})$. On en déduit que $\mathrm{F}(b) - \mathrm{F}(a) = \pi(\operatorname{Argch} b - \operatorname{Argch} a)$ ou encore

$$\boxed{\int_0^\pi \ln\left(\frac{b-\cos x}{a-\cos x}\right)\mathrm{d}x = \pi\ln\left(\frac{b+\sqrt{b^2-1}}{a+\sqrt{a^2-1}}\right)}. \quad \lhd$$

1.31. Mesure du domaine d'annulation d'une fonction

Soit f une fonction continue de $[0,1]$ dans $\mathbb{R}_+$ et g l'application de $\mathbb{R}$ dans $\mathbb{R}_+$ définie par $g(x) = \int_0^1 \sqrt{x^2+f(t)}\mathrm{d}t$.

1. Montrer que g est continue sur $\mathbb{R}$ et de classe $\mathcal{C}^1$ sur $\mathbb{R}^*$.

2. On suppose que $\{t \in [0,1], f(t) = 0\} = [a,b] \subset [0,1]$. Calculer $g'_d(0)$.

(**École normale supérieure**)

▷ **Solution.**

1. Considérons

$$\varphi : (x,t) \in \mathbb{R} \times [0,1] \longmapsto \sqrt{x^2 + f(t)}.$$

La fonction φ est continue sur $\mathbb{R} \times [0,1]$. On en déduit que g est continue sur $\mathbb{R}$. La fonction $x \longmapsto \sqrt{x}$ étant dérivable sur $\mathbb{R}^*$, la fonction φ admet sur $\mathbb{R}^* \times [0,1]$ une dérivée partielle $\dfrac{\partial \varphi}{\partial x}$. Pour $(x,t) \in \mathbb{R}^* \times [0,1]$,

$$\frac{\partial \varphi}{\partial x}(x,t) = \frac{x}{\sqrt{x^2 + f(t)}}.$$

$\dfrac{\partial \varphi}{\partial x}$ est continue sur $\mathbb{R}^* \times [0,1]$. On en déduit que g est $\mathcal{C}^1$ sur $\mathbb{R}^*$ et que pour $x \in \mathbb{R}^*$,

$$g'(x) = \int_0^1 \frac{x}{\sqrt{x^2 + f(t)}} \mathrm{d}t.$$

On peut noter de plus que la fonction g est convexe car, pour tout $a \geqslant 0$, la fonction $x \longmapsto \sqrt{x^2 + a}$ l'est.

2. Pour tout $t \in [0,1]$, la fonction $x \longmapsto \dfrac{x}{\sqrt{x^2 + f(t)}}$ est croissante. On en déduit que $g'_{|]0,+\infty[}$ est croissante et minorée par 0. Elle possède donc une limite finie en 0. Comme g est continue en 0, $g_{|[0,+\infty[}$ est $\mathcal{C}^1$ et $g'_d(0) = \lim\limits_{x \to 0^+} g'(x)$.

On écrit, pour $x > 0$,

$$g'(x) = \int_0^a \frac{x}{\sqrt{x^2 + f(t)}} \mathrm{d}t + (b-a) + \int_b^1 \frac{x}{\sqrt{x^2 + f(t)}} \mathrm{d}t.$$

Pour tout $t \in [0, a[\cup]b, 1[$, on a $\lim\limits_{x \to 0} \dfrac{x}{\sqrt{x^2 + f(t)}} = \dfrac{0}{\sqrt{f(t)}} = 0$. Comme de plus, pour $t \in [0,1]$ et $x \in \mathbb{R}$, on a $\left| \dfrac{x}{\sqrt{x^2 + f(t)}} \right| \leqslant 1$, on obtient par convergence dominée,

$$\lim_{x \to 0} \int_0^a \frac{x}{\sqrt{x^2 + f(t)}} \mathrm{d}t = \lim_{x \to 0} \int_b^1 \frac{x}{\sqrt{x^2 + f(t)}} \mathrm{d}t = 0.$$

Conclusion. $g'_d(0) = b - a$. ◁

Le lecteur connaissant l'intégrale de Lebesgue pourra prouver que si on prend f borélienne, $g'_d(0)$ est toujours égal à la mesure de Lebesgue de l'ensemble des zéros de f.

1.32. Limites de normes intégrales

Soit f continue de $[0,1]$ dans $\mathbb{R}_+^*$. On pose pour $\alpha > 0$,

$$\mathrm{I}(\alpha) = \left(\int_0^1 (f(t))^\alpha \, \mathrm{d}t \right)^{1/\alpha}.$$

Étudier les limites de $\mathrm{I}(\alpha)$ quand α tend vers 0^+ et quand α tend vers $+\infty$.

(École polytechnique)

▷ **Solution.**

• Soit F la fonction définie sur $\mathbb{R}$ par $\mathrm{F}(\alpha) = \int_0^1 (f(t))^\alpha \, \mathrm{d}t$. Si F est continue en 0, $\mathrm{F}(\alpha)$ tend vers $\mathrm{F}(0) = 1$ quand $\alpha \to 0^+$ et $\mathrm{I}(\alpha)$ est une forme indéterminée du type 1^∞. On va donc montrer que F est continue, et même dérivable, ce qui nous donnera un développement limité en 0 suffisant pour lever l'indétermination.

La fonction $\varphi : (t,\alpha) \in [0,1] \times \mathbb{R}^+ \longmapsto (f(t))^\alpha \in \mathbb{R}$ est continue sur $[0,1] \times \mathbb{R}$ et pour tout $(t,\alpha) \in [0,1] \times \mathbb{R}$, on a $\dfrac{\partial \varphi}{\partial \alpha}(t,\alpha) = \ln f(t) \, (f(t))^\alpha$. Donc $\dfrac{\partial \varphi}{\partial \alpha}$ est continue sur $[0,1] \times \mathbb{R}$. On en déduit que F est dérivable sur $\mathbb{R}$ et que, pour tout $\alpha \in \mathbb{R}$, on a

$$\mathrm{F}'(\alpha) = \int_0^1 \ln f(t) \, (f(t))^\alpha \, \mathrm{d}t.$$

Pour tout réel α, on a $\mathrm{I}(\alpha) = \exp\left(\dfrac{1}{\alpha} \ln \mathrm{F}(\alpha) \right)$. Sachant que $\lim\limits_{\alpha \to 0} \mathrm{F}(\alpha) = \mathrm{F}(0) = 1$, on obtient

$$\lim_{\alpha \to 0} \frac{1}{\alpha} \ln \mathrm{F}(\alpha) = \lim_{\alpha \to 0} \frac{\mathrm{F}(\alpha) - 1}{\alpha} = \mathrm{F}'(0) = \int_0^1 \ln f.$$

On en déduit que

$$\boxed{\lim_{\alpha \to 0} \mathrm{I}(\alpha) = \exp\left(\int_0^1 \ln f \right).}$$

• On a, pour tout $\alpha \in \mathbb{R}_+^*$, $\mathrm{F}(\alpha) \leqslant (\|f\|_\infty)^\alpha$ et donc $\mathrm{I}(\alpha) \leqslant \|f\|_\infty$. Comme f est continue sur le compact $[0,1]$, elle y atteint son maximum $\|f\|_\infty$ en un point t_0. Soit $\varepsilon > 0$. Par continuité de f en t_0, il existe un intervalle d'intérieur non vide $\mathrm{J} \subset [0,1]$, contenant t_0 tel que, pour tout $t \in \mathrm{J}$, on a $f(t) \geqslant \|f\|_\infty - \varepsilon$. Si on note ℓ la longueur de J, on obtient,

pour tout $\alpha \in \mathbb{R}_+^*$,

$$\mathrm{F}(\alpha) \geqslant \int_{\mathrm{J}} (\|f\|_\infty - \varepsilon)^\alpha \geqslant \ell\,(\|f\|_\infty - \varepsilon)^\alpha$$

et donc $\mathrm{I}(\alpha) \geqslant \ell^{\frac{1}{\alpha}}(\|f\|_\infty - \varepsilon)$. Comme $\lim\limits_{\alpha \to +\infty} \ell^{\frac{1}{\alpha}}(\|f\|_\infty - \varepsilon) = \|f\|_\infty - \varepsilon$, il existe $\mathrm{A} > 0$ tel que, pour $\alpha > \mathrm{A}$, on ait $\|f\|_\infty - 2\varepsilon \leqslant \mathrm{I}(\alpha) \leqslant \|f\|_\infty$. On conclut

$$\boxed{\lim_{\alpha \to +\infty} \mathrm{I}(\alpha) = \|f\|_\infty}. \quad \triangleleft$$

Pour $p \geqslant 1$ on sait que $f \longmapsto \|f\|_p = \left(\int_0^1 |f|^p\right)^{1/p}$ est une norme sur l'espace $C^0([0,1],\mathbb{R})$. On montre comme dans le second point que pour toute fonction $f \in C^0([0,1],\mathbb{R})$ la norme $\|f\|_p$ tend vers la norme infinie de f lorsque $p \to +\infty$.

Les exercices qui suivent sont de nature asymptotique et concernent des recherches de limites, d'équivalents, voire de développements asymptotiques.

1.33. Équivalent d'une primitive

Soit $f : x \longmapsto e^{-x^2} \displaystyle\int_0^x e^{t^2}\,\mathrm{d}t$. En donner un équivalent en $+\infty$.

(École polytechnique)

▷ **Solution.**

On ne peut pas calculer une primitive de la fonction $t \mapsto e^{t^2}$, c'est-à-dire l'exprimer à l'aide des fonctions usuelles (cela peut se formuler de manière précise et se démontrer mais ce n'est pas facile ; le lecteur intéressé pourra faire le problème d'écrit Lyon-Cachan de 1995). Si l'on essayait tout de même, une idée naturelle serait de la chercher sous la forme $g(t)e^{t^2}$. Comme on a

$$\frac{\mathrm{d}\big(g(t)e^{t^2}\big)}{\mathrm{d}t} = \Big(g'(t) + 2tg(t)\Big)e^{t^2},$$

on serait alors ramené à chercher les solutions de l'équation différentielle $g'(t) + 2tg(t) = 1$. Ce qui se passe, évidemment, c'est que la méthode de la variation de la constante conduit à l'intégrale que l'on cherche à calculer !

Mais ici on ne demande pas de calculer explicitement $\displaystyle\int_0^x e^{t^2}\mathrm{d}t$, on n'en veut qu'un équivalent. Il nous suffit pour cela d'utiliser le théorème

d'intégration des relations de comparaison. Plutôt que de chercher g telle que $g'(t)+2tg(t) = 1$, il nous suffit de trouver g telle que $g'(t)+2tg(t) \sim 1$ pour t tendant vers $+\infty$. Comme les intégrales divergent clairement, on a alors

$$g(x)e^{x^2} - g(0) = \int_0^x (g'(t) + 2tg(t))e^{t^2}\,\mathrm{d}t \sim \int_0^x e^{t^2}\,\mathrm{d}t$$

et on a obtenu notre équivalent. Trouver une fonction g de classe C^1 telle que $g'(t) + 2tg(t) \sim 1$ quand $t \to +\infty$ est facile. Il suffit de prendre $g(t) = \dfrac{1}{2t}$ ou plutôt $g(t) = \dfrac{1}{2(t+1)}$ pour éviter tout problème en 0. On en déduit donc que

$$\boxed{f(x) \underset{x\to+\infty}{\sim} \frac{1}{2x}}. \quad \lhd$$

1.34. Recherche de limite

Soient $f : \mathbb{R} \longrightarrow \mathbb{R}$ continue, g définie pour $x > 0$ par

$$g(x) = \frac{1}{x}\int_0^x \cos(x-y)f(y)\mathrm{d}y.$$

Déterminer la limite de g en 0. En supposant que f a une limite en $+\infty$, déterminer celle de g.

(École polytechnique)

Solution.

• Étudions la limite en 0 : pour tout $x > 0$, on a

$$g(x) = \frac{\cos x}{x}\int_0^x \cos y f(y)\mathrm{d}y + \frac{\sin x}{x}\int_0^x \sin y f(y)\mathrm{d}y.$$

Or si φ est une fonction continue de $\mathbb{R}$ dans $\mathbb{R}$, on a

$$\frac{1}{x}\int_0^x \varphi(y)\mathrm{d}y = \frac{\Phi(x)-\Phi(0)}{x} \xrightarrow[x\to 0]{} \Phi'(0) = \varphi(0).$$

où Φ désigne une primitive de φ. Ainsi, on obtient ici

$$\lim_{x\to 0} g(x) = \cos(0)\cos(0)f(0) + 0 = f(0).$$

• Passons à la limite en $+\infty$. Notons ℓ la limite de f en $+\infty$. Comme

$$\frac{1}{x}\int_0^x \cos(x-y)(f(y)-\ell)\mathrm{d}y = g(x) - \frac{\ell}{x}\int_0^x \cos(x-y)\mathrm{d}y,$$

et comme

$$\frac{1}{x}\int_0^x \cos(x-y)\mathrm{d}y = \frac{\sin x}{x} \xrightarrow[x\to+\infty]{} 0,$$

on se ramène à traiter le cas où la limite de f en l'infini est nulle. Soient ε et $A > 0$ tels que $|f(x)| \leqslant \varepsilon$ pour $x \geqslant A$. Pour $x \geqslant A$ on a

$$\begin{aligned}|g(x)| &\leqslant \frac{1}{x}\int_0^A |\cos(x-y)f(y)|\mathrm{d}y + \frac{1}{x}\int_A^x |\cos(x-y)f(y)|\mathrm{d}y\\ &\leqslant \frac{1}{x}\int_0^A |f(y)|\mathrm{d}y + \frac{1}{x}\int_A^x \varepsilon \mathrm{d}y \leqslant \frac{1}{x}\int_0^A |f(y)|\mathrm{d}y + \varepsilon.\end{aligned}$$

Comme $\lim\limits_{x\to+\infty} \dfrac{1}{x}\displaystyle\int_0^A |f(y)|\mathrm{d}y = 0$, pour x assez grand, $|g(x)| \leqslant 2\varepsilon$. On conclut que g tend vers 0. Dans le cas général on a donc

$$\boxed{\lim_{x\to+\infty} g(x) = \ell = \lim_{x\to+\infty} f(x)}. \quad \triangleleft$$

L'exercice suivant est un grand classique qu'il est bon de connaître.

1.35. Équivalent des intégrales de Wallis

Chercher un équivalent de I_n où $I_n = \displaystyle\int_0^{\frac{\pi}{2}} \cos^n t\,\mathrm{d}t$. En déduire la nature de la série $\sum (I_n)^\alpha$.

(École polytechnique)

▷ **Solution.**

On reconnaît les célèbres intégrales de Wallis. On effectue une intégration par parties afin d'obtenir une relation de récurrence. Il vient, pour $n \geqslant 2$,

$$I_n = \int_0^{\pi/2} \cos\cdot\cos^{n-1} = [\sin\cos^{n-1}]_0^{\pi/2} + (n-1)\int_0^{\pi/2} \sin^2\cos^{n-2}.$$

Le crochet est nul et, en remplaçant $\sin^2$ par $1-\cos^2$, on obtient la relation de récurrence $I_n = (n-1)(I_{n-2} - I_n)$ que peut aussi s'écrire $nI_n = (n-1)I_{n-2}$.

On a donc pour tout $n \geqslant 1$, $nI_{n-1}I_n = (n-1)I_{n-2}I_{n-1}$. Autrement dit, la suite de terme général $nI_{n-1}I_n$ est constante. Elle vaut $\dfrac{\pi}{2}$, puisque $I_0 = \dfrac{\pi}{2}$ et $I_1 = 1$. On en déduit que, pour $n \geqslant 1$, $nI_{n-1}I_n = \dfrac{\pi}{2}$.

Par ailleurs, on a, pour $n \in \mathbb{N}$ et $t \in \left[0, \dfrac{\pi}{2}\right]$, $0 \leqslant \cos^{n+1} t \leqslant \cos^n t$ et donc en intégrant cela $0 \leqslant I_{n+1} \leqslant I_n$: la suite (I_n) est décroissante.

On obtient, pour $n \geqslant 1$, $I_{n+1} \leqslant I_n \leqslant I_{n-1}$ et $I_nI_{n+1} \leqslant I_n^2 \leqslant I_{n-1}I_n$, c'est-à-dire, d'après ce qui précède,

$$\frac{\pi}{2(n+1)} \leqslant I_n^2 \leqslant \frac{\pi}{2n}.$$

On en déduit $I_n^2 \underset{n\to\infty}{\sim} \frac{\pi}{2n}$ et, I_n étant positif, $I_n \sim \sqrt{\frac{\pi}{2n}}$.

On a donc $I_n^\alpha \underset{n\to+\infty}{\sim} \left(\frac{\pi}{2n}\right)^{\frac{\alpha}{2}}$ et, par comparaison avec les séries de Riemann, la série $\sum I_n^\alpha$ converge si et seulement si $\alpha > 2$. ◁

1.36. Recherche d'un équivalent (1)

Soit $f : [0,1] \longrightarrow \mathbb{R}$ continue et telle que $f(0) \neq 0$.

1. Donner un équivalent en $+\infty$ de $g(t) = \int_0^1 \frac{f(x)}{1+tx}\,dx$.

2. Majorer la différence entre g et cet équivalent quand f est de classe C^1.

(École polytechnique)

◁ **Solution.**

1. Dans l'intégrale définissant g, effectuons le changement de variable $u = tx$ pour obtenir pour $t > 0$,

$$g(t) = \frac{1}{t}\int_0^t \frac{f\left(\frac{u}{t}\right)}{1+u}\,du.$$

Quand t tend vers $+\infty$, $f\left(\frac{u}{t}\right)$ est « proche » de $f(0)$ par continuité. Or, nous avons

$$\frac{1}{t}\int_0^t \frac{f(0)}{1+u}\,du = f(0)\frac{\ln(1+t)}{t} \underset{t\to+\infty}{\sim} f(0)\frac{\ln t}{t},$$

car $f(0) \neq 0$. On peut donc imaginer que $g(t) \underset{t\to+\infty}{\sim} f(0)\frac{\ln t}{t}$. Justifions cela.

Notons $g_0(t) = \int_0^1 \frac{f(0)}{1+tx}\,dx$ pour $t > 0$. Nous venons de montrer que $g_0(t) \underset{t\to+\infty}{\sim} f(0)\frac{\ln t}{t}$. Majorons la différence $|g(t) - g_0(t)|$.

Soit $\varepsilon > 0$ et $a \in]0,1[$ tel que $|f(x) - f(0)| \leqslant \varepsilon$ pour tout $x \in [0,a]$. En coupant l'intégrale en deux, on obtient

$$\begin{aligned}|g(t) - g_0(t)| &\leqslant \left|\int_0^a \frac{f(x)-f(0)}{1+tx}\,dx\right| + \left|\int_a^1 \frac{f(x)-f(0)}{1+tx}\,dx\right| \\ &\leqslant \varepsilon\int_0^a \frac{1}{1+tx}\,dx + 2\|f\|_\infty \int_a^1 \frac{dx}{tx}\end{aligned}$$

$$\leqslant \varepsilon \frac{\ln(1+t)}{t} + \frac{2\|f\|_\infty}{ta}$$
$$\leqslant \frac{\ln(1+t)}{t}\left(\varepsilon + \frac{2\|f\|_\infty}{a\ln(1+t)}\right).$$

Comme la quantité $\varepsilon + \dfrac{2\|f\|_\infty}{a\ln(1+t)}$ tend vers ε lorsque t tend vers l'infini, on en déduit que pour t assez grand, elle devient inférieure ou égale à 2ε. On a alors

$$|g(t) - g_0(t)| \leqslant 2\varepsilon \frac{\ln(1+t)}{t}$$

et par conséquent, $g(t) - g_0(t) = o(g_0(t))$, ou encore

$$\boxed{g(t) \underset{t\to+\infty}{\sim} g_0(t) \underset{t\to+\infty}{\sim} f(0)\frac{\ln t}{t}}.$$

2. Si on suppose f de classe $\mathcal{C}^1$, f' est en particulier continue sur le compact $[0,1]$ et, d'après l'inégalité des accroissements finis,

$$|g(t) - g_0(t)| \leqslant \int_0^1 \frac{|f(x) - f(0)|}{1+tx}\mathrm{d}x \leqslant \int_0^1 \frac{x\|f'\|_\infty}{1+tx}\mathrm{d}x$$
$$\leqslant \|f'\|_\infty \int_0^1 \frac{\mathrm{d}x}{t} = \frac{\|f'\|_\infty}{t}.$$

Par ailleurs,

$$\left|g_0(t) - f(0)\frac{\ln t}{t}\right| = |f(0)|\left|\int_0^1 \frac{\mathrm{d}x}{1+tx} - \frac{\ln t}{t}\right| \leqslant |f(0)|\frac{\ln(1+1/t)}{t}.$$

Ainsi, par inégalité triangulaire, on a

$$\boxed{\left|g(t) - f(0)\frac{\ln t}{t}\right| \leqslant \frac{1}{t}\left(\|f'\|_\infty + |f(0)|\ln 2\right)}. \quad \lhd$$

Dans l'exercice suivant on n'hésitera pas à utiliser le théorème de convergence dominée.

1.37. Recherche d'un équivalent (2)

Chercher un équivalent en $+\infty$ de

$$\Phi(t) = \int_0^1 \frac{\mathrm{d}x}{\left(1+x+x^2\right)^t}.$$

(École polytechnique)

▷ **Solution.**

Nous pouvons appliquer le théorème de convergence dominée pour prouver que Φ admet 0 comme limite lorsque t tend vers $+\infty$. En effet, la fonction $(t,x) \longmapsto \dfrac{1}{(1+x+x^2)^t}$ est continue sur $\mathbb{R}_+^2$; pour tout $x \in]0,1]$,

$$\frac{1}{(1+x+x^2)^t} \xrightarrow[t\to+\infty]{} 0$$

et on a la domination $0 \leqslant \dfrac{1}{(1+x+x^2)^t} \leqslant 1$ pour tout $t > 0$ avec la fonction constante 1 qui est intégrable sur $[0,1]$. On a donc

$$\boxed{\lim_{t\to+\infty} \Phi(t) = 0}.$$

Pour t grand, la principale contribution de la fonction $x \longmapsto \dfrac{1}{(1+x+x^2)^t}$ à l'intégrale se situe en 0. On a, pour x tendant vers 0,

$$\frac{1}{(1+x+x^2)^t} = \exp(-t\ln(1+x+x^2)) = \exp(-tx+o(x)),$$

si bien que l'on peut imaginer comparer $\Phi(t)$ à $\int_0^1 e^{-tx}\mathrm{d}x$ que l'on sait calculer :

$$\int_0^1 e^{-tx}\mathrm{d}x \underbrace{=}_{u=xt} \frac{1}{t}\int_0^t e^{-u}\mathrm{d}u \underset{t\to+\infty}{\sim} \frac{1}{t}.$$

Contrôlons pour $t > 0$ la différence entre $\Phi(t)$ et $\int_0^1 e^{-tx}\mathrm{d}x$:

$$\begin{aligned}\left|\Phi(t) - \int_0^1 e^{-xt}\mathrm{d}x\right| &\leqslant \int_0^1 \left|e^{-t\ln(1+x+x^2)} - e^{-tx}\right|\mathrm{d}x \\ &\leqslant \int_0^1 e^{-tx}\left|e^{-t(\ln(1+x+x^2)-x)} - 1\right|\mathrm{d}x \\ &\underbrace{\leqslant}_{u=xt} \frac{1}{t}\int_0^t e^{-u}\left|e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u} - 1\right|\mathrm{d}u.\end{aligned}$$

Une étude simple des variations de $x \longmapsto \ln(1+x+x^2) - x$ montre que cette fonction est toujours positive sur $[0,1]$ si bien que pour $u \in [0,t]$, on a $-t\ln\left(1+\dfrac{u}{t}+\dfrac{u^2}{t^2}\right)+u \leqslant 0$ et

$$\left|\Phi(t) - \int_0^1 e^{-xt}\mathrm{d}x\right| \leqslant \frac{1}{t}\int_0^t e^{-u}\left(1 - e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u}\right)\mathrm{d}u.$$

Dans ces conditions, on a pour tout $0 \leqslant u \leqslant t$ la domination

$$e^{-u}\left(1-e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u}\right) \leqslant e^{-u},$$

ou encore

$$\forall (u,t)\in\mathbb{R}_+^2,\quad e^{-u}\left(1-e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u}\right)1_{[0,t]}(u) \leqslant e^{-u}.$$

Or $u \longmapsto e^{-u}$ est intégrable sur $\mathbb{R}_+$. Par continuité de l'exponentielle, on a $e^{-u}\left(1-e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u}\right) = e^{-u}(1-e^{o(1/t)}) \xrightarrow[t\to+\infty]{} 0$ et il résulte du théorème de convergence dominée que

$$\int_{\mathbb{R}_+} e^{-u}\left(1-e^{-t\ln\left(1+\frac{u}{t}+\frac{u^2}{t^2}\right)+u}\right)1_{[0,t]}(u)\mathrm{d}u \xrightarrow[t\to+\infty]{} 0.$$

On en déduit que $\Phi(t) - \int_0^1 e^{-xt}\mathrm{d}x = o(1/t)$. On conclut donc que $\Phi(t) \underset{t\to+\infty}{\sim} \int_0^1 e^{-xt}\mathrm{d}x$ et finalement

$$\boxed{\Phi(t) \underset{t\to+\infty}{\sim} \frac{1}{t}}.$$

1.38. Recherche d'un équivalent (3)

On pose $u_n = \displaystyle\int_0^{1/2} \frac{\sin^2 \pi n x}{\tan \pi x}\mathrm{d}x$. Étudier la série $\sum \dfrac{u_n^\alpha}{n^\beta}$ pour $(\alpha,\beta)\in\mathbb{R}^2$.

(École polytechnique)

▷ **Solution.**

Le problème consiste essentiellement à déterminer un équivalent de la suite u_n.

La fonction continue $x\in\,]0,1/2[\longmapsto \dfrac{\sin^2 n\pi x}{\tan \pi x}$ se prolonge par continuité en 0 (avec la valeur 0) et en $1/2$ (avec également la valeur 0), si bien que u_n est bien défini pour tout $n\in\mathbb{N}$. La quantité $\tan \pi x$ tendant vers 0 en 0, tandis que, n tendant vers l'infini, $\sin^2 \pi n x$ vaut 1 en des points de plus en plus proches de 0. On peut donc imaginer que la majeure contribution se trouvera au voisinage de 0 et remplacer $\tan \pi x$ par l'équivalent simple πx pour x tendant vers 0. Posons donc

$$v_n = \int_0^{1/2} \frac{\sin^2(\pi n x)}{\pi x}\mathrm{d}x \quad \text{pour } n \text{ entier naturel.}$$

Opérons le changement de variables $z = n\pi x$ dans l'intégrale,

$$v_n = \frac{1}{\pi}\int_0^{\frac{n\pi}{2}} \frac{\sin^2 z}{z}\mathrm{d}z = \frac{1}{2\pi}\int_0^{\frac{n\pi}{2}} \frac{1-\cos 2z}{z}\mathrm{d}z = \frac{1}{2\pi}\int_0^{n\pi} \frac{1-\cos u}{u}\mathrm{d}u.$$

En notant $\mathrm{C} = \dfrac{1}{2\pi}\displaystyle\int_0^{\pi} \frac{1-\cos u}{u}\mathrm{d}u$, il vient

$$v_n = \mathrm{C} + \frac{1}{2\pi}\int_{\pi}^{n\pi} \frac{\mathrm{d}u}{u} - \frac{1}{2\pi}\int_{\pi}^{n\pi} \frac{\cos u}{u}\mathrm{d}u.$$

La dernière intégrale admet une limite finie lorsque n tend vers l'infini puisqu'en intégrant par parties,

$$\begin{aligned}\int_{\pi}^{n\pi} \frac{\cos u}{u}\mathrm{d}u &= \left[\frac{\sin u}{u}\right]_{\pi}^{n\pi} + \int_{\pi}^{n\pi} \frac{\sin u}{u^2}\mathrm{d}u \\ &= \int_{\pi}^{n\pi} \frac{\sin u}{u^2}\mathrm{d}u \xrightarrow[n\to+\infty]{} \int_{\pi}^{+\infty} \frac{\sin u}{u^2}\mathrm{d}u,\end{aligned}$$

car la fonction $u \longmapsto \dfrac{\sin u}{u^2}$, majorée en valeur absolue par $u \mapsto \dfrac{1}{u^2}$ est intégrable sur $[\pi, +\infty[$. Il reste donc au voisinage de $+\infty$,

$$v_n \underset{n\to+\infty}{\sim} \frac{\ln n}{2\pi} + O(1) \underset{n\to+\infty}{\sim} \frac{\ln n}{2\pi}.$$

Estimons maintenant la différence $u_n - v_n$ lorsque n tend vers $+\infty$. On peut écrire

$$u_n - v_n = \int_0^{1/2} \sin^2(n\pi x) f(x)\mathrm{d}x \quad \text{avec } f(x) = \frac{1}{\tan \pi x} - \frac{1}{\pi x}$$

pour x dans $]0, 1/2[$. Cette fonction f se prolonge par continuité en $1/2$ (avec la valeur $-2/\pi$) et également en 0 puisque,

$$f(x) = \frac{\pi x - \tan \pi x}{\pi x \tan \pi x} = \frac{o(x^2)}{\pi^2 x^2 + o(x^2)} = o(1) \text{ si } x \text{ tend vers } 0.$$

En posant $f(0) = 0$ et $f(1/2) = -2/\pi$, on obtient une fonction continue sur le compact $[0, 1/2]$ et donc nécessairement bornée. Il s'ensuit

$$|u_n - v_n| \leqslant \int_0^{1/2} |f(x)|\mathrm{d}x \leqslant \frac{\|f\|_\infty}{2}.$$

La différence $u_n - v_n$ est donc négligeable devant v_n, et on obtient donc l'équivalent

$$\boxed{u_n \underset{n\to+\infty}{\sim} \frac{\ln n}{2\pi}}.$$

Pour α et β réels, on a donc

$$\frac{u_n^\alpha}{n^\beta} \underset{n\to+\infty}{\sim} \frac{(\ln n)^\alpha}{(2\pi)^\alpha n^\beta}.$$

D'après le théorème de comparaison des séries à termes positifs, et les résultats sur les séries de Bertrand, il y a convergence si, et seulement si, $\beta > 1$ ou $\beta = 1$ et $\alpha < -1$. ◁

1.39. Développement asymptotique (1)

On pose $u_n = \displaystyle\int_0^1 \frac{x^n}{\sqrt{1+x}}\,\mathrm{d}x$. Déterminer $\lim\limits_{n\to+\infty} u_n$, puis un équivalent et un développement asymptotique à deux termes de u_n. (**École polytechnique**)

▷ **Solution.**

• On obtient le limite de la suite (u_n) en appliquant le théorème de convergence dominée. En effet, la suite de fonctions $x \longmapsto \dfrac{x^n}{\sqrt{1+x}}$ converge simplement sur $[0,1]$ vers une fonction qui est nulle sur $[0,1[$ et, pour tout $n \geqslant 1$ et tout $x \in [0,1]$, on a la majoration

$$0 \leqslant \frac{x^n}{\sqrt{1+x}} \leqslant \frac{1}{\sqrt{1+x}},$$

où $x \longmapsto \dfrac{1}{\sqrt{1+x}}$ est une fonction intégrable sur $[0,1]$. On peut donc affirmer que

$$\boxed{\lim_{n\to+\infty} u_n = 0}.$$

On peut remarquer ici qu'il n'y a pas convergence uniforme de la suite de fonctions vers la fonction nulle.

• Pour trouver un équivalent de u_n, il convient d'observer que la contribution à l'intégrale se concentre au voisinage de 1. Pour conjecturer l'équivalent, on remplace $x \longmapsto \dfrac{1}{\sqrt{1+x}}$ par sa valeur en 1. On trouve alors

$$\int_0^1 \frac{x^n}{\sqrt{2}}\mathrm{d}x = \frac{1}{(n+1)\sqrt{2}} \underset{n\to\infty}{\sim} \frac{1}{n\sqrt{2}}.$$

Démontrons que cela constitue un équivalent de u_n. On a $\left|u_n - \displaystyle\int_0^1 \frac{x^n}{\sqrt{2}}\,\mathrm{d}x\right| = \displaystyle\int_0^1 \frac{x^n(\sqrt{2}-\sqrt{1+x})}{\sqrt{2}\sqrt{1+x}}\,\mathrm{d}x$ et, pour tout $x \in [0,1]$,

$$\frac{\sqrt{2}-\sqrt{1+x}}{\sqrt{2}\sqrt{1+x}} = \frac{1-x}{\sqrt{2}\sqrt{1+x}(\sqrt{2}+\sqrt{1+x})} \leqslant \frac{1-x}{2}.$$

On en déduit que

$$\left|u_n - \int_0^1 \frac{x^n}{\sqrt{2}}\mathrm{d}x\right| \leqslant \frac{1}{2}\int_0^1 (x^n - x^{n+1})\,\mathrm{d}x = \frac{1}{2(n+1)(n+2)}.$$

On a donc $u_n = \dfrac{1}{\sqrt{2}(n+1)} + O\left(\dfrac{1}{n^2}\right)$, $u_n = \dfrac{1}{\sqrt{2}n} + O\left(\dfrac{1}{n^2}\right)$ et donc

$$\boxed{u_n \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2}n}}.$$

• Nous allons chercher un terme supplémentaire dans le développement asymptotique. Pour cela, écrivons le développement limité en 1 de la fonction $x \mapsto \dfrac{1}{\sqrt{1+x}}$:

$$\begin{aligned}\frac{1}{\sqrt{1+x}} &= \frac{1}{\sqrt{2}}\frac{1}{\sqrt{1+\dfrac{x-1}{2}}} = \frac{1}{\sqrt{2}}\left(1 - \frac{x-1}{4} + o(x-1)\right)\\ &= \frac{1}{\sqrt{2}} + \frac{1-x}{4\sqrt{2}} + \alpha(x)(1-x),\end{aligned}$$

où α est une fonction tendant vers 0 en 1. Cette fonction α se prolonge donc en une fonction continue sur tout $[0,1]$. On peut ainsi écrire

$$u_n = \frac{1}{\sqrt{2}(n+1)} + \int_0^1 \frac{x^n - x^{n+1}}{4\sqrt{2}}\mathrm{d}x + \int_0^1 (x^n - x^{n+1})\alpha(x)\mathrm{d}x.$$

Pour n tendant vers l'infini, on a $\dfrac{1}{\sqrt{2}(n+1)} = \dfrac{1}{\sqrt{2}n} - \dfrac{1}{\sqrt{2}n^2} + o\left(\dfrac{1}{n^2}\right)$ et

$$\int_0^1 \frac{x^n - x^{n+1}}{4\sqrt{2}}\mathrm{d}x = \frac{1}{4\sqrt{2}(n+1)(n+2)} = \frac{1}{4\sqrt{2}n^2} + o\left(\frac{1}{n^2}\right).$$

Quant au dernier terme, on va le majorer en coupant l'intégrale en deux. Soit $\varepsilon > 0$ et $a \in]0,1[$ tel que α soit en valeur absolue inférieure ou égale à ε sur $[a,1]$. Dans ces conditions, on a

$$\begin{aligned}\left|\int_0^1 (x^n - x^{n+1})\alpha(x)\mathrm{d}x\right| &\leqslant \|\alpha\|_\infty \int_0^a (x^n - x^{n+1})\mathrm{d}x + \varepsilon\int_a^1 (x^n - x^{n+1})\mathrm{d}x\\ &\leqslant \|\alpha\|_\infty a^n + \varepsilon\int_0^1 (x^n - x^{n+1})\mathrm{d}x\\ &\leqslant \|\alpha\|_\infty a^n + \frac{\varepsilon}{(n+1)(n+2)}\\ &\leqslant \frac{1}{n^2}\left(\|\alpha\|_\infty n^2 a^n + \varepsilon\right).\end{aligned}$$

Pour n assez grand, $\|\alpha\|_\infty n a^n + \varepsilon \leqslant 2\varepsilon$, et donc $\int_0^1 (x^n - x^{n+1})\alpha(x)\mathrm{d}x$ est négligeable devant $\frac{1}{n^2}$. On conclut ainsi que

$$\boxed{u_n = \frac{1}{\sqrt{2}n} - \frac{3}{4\sqrt{2}n^2} + o\left(\frac{1}{n^2}\right)}. \quad \lhd$$

La première partie de la solution montre que pour toute fonction f continue en 1 on a $\int_0^1 x^n f(x)\mathrm{d}x = \dfrac{f(1)}{n} + o\left(\dfrac{1}{n}\right)$.

1.40. Développement asymptotique (2)

1. Soit $0 < \varepsilon < 1$. On pose $\mathrm{I} = \displaystyle\int_0^{\pi/2} \frac{\cos x}{\sqrt{\sin^2 x + \varepsilon \cos^2 x}}\,\mathrm{d}x$. Calculer I. Donner un équivalent de I quand ε tend vers 0.

2. On pose $\mathrm{J} = \displaystyle\int_0^{\pi/2} \frac{1}{\sqrt{\sin^2 x + \varepsilon \cos^2 x}}\,\mathrm{d}x$. Déterminer les deux premiers termes du développement asymptotique de J quand ε tend vers 0.

(École polytechnique)

▷ **Solution.**

1. On va calculer l'intégrale I explicitement. Le changement de variable $u = \sqrt{\dfrac{1-\varepsilon}{\varepsilon}}\sin x$ conduit à

$$\mathrm{I} = \int_0^{\frac{\pi}{2}} \frac{\cos x}{\sqrt{(1-\varepsilon)\sin^2 x + \varepsilon}}\,\mathrm{d}x = \sqrt{\frac{\varepsilon}{1-\varepsilon}} \int_0^{\sqrt{\frac{1-\varepsilon}{\varepsilon}}} \frac{1}{\sqrt{\varepsilon u^2 + \varepsilon}}\,\mathrm{d}u$$

$$= \frac{1}{\sqrt{1-\varepsilon}} \int_0^{\sqrt{\frac{1-\varepsilon}{\varepsilon}}} \frac{1}{\sqrt{1+u^2}}\,\mathrm{d}u.$$

Une primitive de $x \longmapsto \dfrac{1}{\sqrt{1+x^2}}$ est $x \longmapsto \ln(x + \sqrt{1+x^2}) = \operatorname{Argsh} x$, donc on a

$$\mathrm{I} = \frac{1}{\sqrt{1-\varepsilon}} \ln\left(\sqrt{\frac{1-\varepsilon}{\varepsilon}} + \sqrt{\frac{1-\varepsilon}{\varepsilon}+1}\right)$$

$$= \frac{1}{\sqrt{1-\varepsilon}} \left(\ln\sqrt{\frac{1-\varepsilon}{\varepsilon}} + \ln\left(1 + \frac{1}{\sqrt{1-\varepsilon}}\right)\right).$$

Il est clair que

$$I \underset{\varepsilon\to 0}{\sim} -\frac{1}{2}\ln\varepsilon.$$

2. Dans l'intégrale J, la fonction qu'on intègre est maximale lorsque $x = 0$ et vaut alors $\frac{1}{\sqrt{\varepsilon}}$ qui tend vers l'infini lorsque ε tend vers 0. On peut donc penser que la contribution principale à l'intégrale provient des valeurs de x proches de 0. Mais dans ce cas $\cos x$ est proche de 1 et il est raisonnable de penser que $J \underset{\varepsilon\to 0}{\sim} I$. Pour le montrer, examinons $J - I$. On a

$$\begin{aligned} 0 \leqslant J - I &= \int_0^{\frac{\pi}{2}} \frac{1-\cos x}{\sqrt{\sin^2 x + \varepsilon\cos^2 x}}\,dx \\ &\leqslant \int_0^{\frac{\pi}{2}} \frac{2\sin^2\frac{x}{2}}{\sin x}\,dx = \int_0^{\frac{\pi}{2}} \tan\frac{x}{2}\,dx = \left[-2\ln\cos\frac{x}{2}\right]_0^{\frac{\pi}{2}} = \ln 2. \end{aligned}$$

Puisque I tend vers $+\infty$ quand ε tend vers 0, on en déduit que

$$J \underset{\varepsilon\to 0}{\sim} I \underset{\varepsilon\to 0}{\sim} -\frac{1}{2}\ln\varepsilon.$$

Pour déterminer le deuxième terme du développement asymptotique de J, déterminons, pour commencer le deuxième terme de celui de I. On a

$$\begin{aligned} I &= (1-\varepsilon)^{-\frac{1}{2}}\left(-\frac{1}{2}\ln\varepsilon + \frac{1}{2}\ln(1-\varepsilon) + \ln\left(1+(1-\varepsilon)^{-\frac{1}{2}}\right)\right) \\ &= (1+O(\varepsilon))\left(-\frac{1}{2}\ln\varepsilon + \ln 2 + O(\varepsilon)\right) = -\frac{1}{2}\ln\varepsilon + \ln 2 + o(1). \end{aligned}$$

Précisons maintenant le comportement de $J - I$, en montrant que $\lim_{\varepsilon\to 0}(J-I) = \ln 2$. Considérons la fonction $\varphi : [0, \frac{\pi}{2}] \times [0,1] \longrightarrow \mathbb{R}$ définie par

$$\varphi(x,\varepsilon) = \frac{1-\cos x}{\sqrt{\sin^2 x + \varepsilon\cos^2 x}} \quad \text{si} \quad (x,\varepsilon) \neq (0,0) \quad \text{et} \quad \varphi(0,0) = 0.$$

La fonction φ est continue sur $[0, \frac{\pi}{2}] \times [0,1]$. En effet, pour tout couple $(x,\varepsilon) \in]0, \frac{\pi}{2}] \times [0,1]$, on a $0 \leqslant \varphi(x,\varepsilon) \leqslant \frac{1-\cos x}{\sin x} = \tan\frac{x}{2}$, ce qui assure la continuité en $(0,0)$. Le théorème de continuité des intégrales à paramètre, montre que la fonction $\varepsilon \longmapsto \int_0^{\frac{\pi}{2}} \varphi(x,\varepsilon)\,dx = J - I$ est continue sur $[0,1]$. On a donc

$$\lim_{\varepsilon\to 0}(J-I) = \int_0^{\frac{\pi}{2}} \frac{1-\cos x}{\sin x}\,dx = \int_0^{\frac{\pi}{2}} \tan\frac{x}{2}\,dx = \ln 2.$$

On en déduit que

$$\boxed{\mathrm{J} = 1 + \ln 2 + o(1) = -\frac{1}{2}\ln\varepsilon + 2\ln 2 + o(1)}. \quad \triangleleft$$

Les deux énoncés qui suivent sont difficiles et concernent la méthode de Laplace.

1.41. Méthode de Laplace (1)

Soit $f : [0,1] \longrightarrow \mathbb{R}$ de classe $\mathcal{C}^\infty$. On suppose qu'il existe $k \in \mathbb{N}$ tel que $f^{(k)}(0) \neq 0$. Pour $\lambda > 0$, on pose $\mathrm{I}(\lambda) = \displaystyle\int_0^1 (1-t^2)^\lambda f(t)\mathrm{d}t$. Trouver un équivalent de $\mathrm{I}(\lambda)$ lorsque λ tend vers l'infini.

(École normale supérieure)

▷ **Solution.**

• On peut légitimement penser que la contribution à l'intégrale $\mathrm{I}(\lambda)$ se concentre en 0 car pour $t > 0$ la quantité $(1-t^2)^\lambda$ tend vers 0 lorsque λ tend vers $+\infty$. Par ailleurs, au voisinage de 0, on a

$$(1-t^2)^\lambda = e^{\lambda \ln(1-t^2)} = e^{-\lambda t^2 + o(t^3)}.$$

Si on remplace $(1-t^2)^\lambda$ par le terme $e^{-\lambda t^2}$, on obtient une intégrale plus facile à estimer. En effet, par le changement de variable $u = \sqrt{\lambda}t$,

$$\int_0^1 e^{-\lambda t^2}\mathrm{d}t = \int_0^{\sqrt{\lambda}} e^{-u^2}\frac{\mathrm{d}u}{\sqrt{\lambda}} \underset{\lambda\to+\infty}{\sim} \frac{1}{2}\sqrt{\frac{\pi}{\lambda}}$$

car $\displaystyle\int_{\mathbb{R}_+} e^{-u^2}\mathrm{d}u = \frac{\sqrt{\pi}}{2}$.

On va aussi remplacer f par un équivalent en 0. Si on note n le plus petit entier naturel tel que $f^{(n)}(0) \neq 0$, on a $f(t) \underset{t\to 0}{\sim} \dfrac{f^{(n)}(0)}{n!}t^n$ et nous allons donc nous intéresser aux intégrales

$$\mathrm{I}_n(\lambda) = \int_0^1 (1-t^2)^\lambda t^n \mathrm{d}t \quad \text{et} \quad \mathrm{J}_n(\lambda) = \int_0^1 e^{-\lambda t^2} t^n \mathrm{d}t,$$

puisque $(1-t^2)^\lambda$ est « proche » de $e^{-\lambda t^2}$ pour t voisin de 0. En faisant le changement de variables $u = \sqrt{\lambda}t$ dans $\mathrm{J}_n(\lambda)$, on arrive à

$$\mathrm{J}_n(\lambda) = \int_0^{\sqrt{\lambda}} e^{-u^2}\frac{u^n}{(\sqrt{\lambda})^{n+1}}\mathrm{d}u \underset{\lambda\to\infty}{\sim} \frac{1}{\lambda^{\frac{n+1}{2}}}\int_{\mathbb{R}_+} u^n e^{-u^2}\mathrm{d}u = \frac{\mathrm{A}_n}{\lambda^{\frac{n+1}{2}}},$$

en notant $A_n = \int_{\mathbb{R}_+} u^n e^{-u^2} du$. Ces constantes A_n sont des valeurs particulières de la fonction Γ d'Euler comme le montre le changement de variable $v = u^2$:

$$A_n = \frac{1}{2}\int_{\mathbb{R}_+} v^{\frac{n-1}{2}} e^{-v} dv = \frac{1}{2}\Gamma\left(\frac{n+1}{2}\right).$$

L'équation fonctionnelle $\Gamma(x+1) = x\Gamma(x)$ pour $x > 0$ conduit à la relation de récurrence $A_n = \frac{n-1}{2} A_{n-2}$. À partir de $A_0 = \frac{\sqrt{\pi}}{2}$ et $A_1 = \left[-\frac{e^{-u^2}}{2}\right]_0^{+\infty} = \frac{1}{2}$, on obtient donc

$$\begin{aligned} A_{2n} &= \frac{2n-1}{2} A_{2(n-1)} = \cdots = \frac{(2n-1)(2n-3)\ldots 1}{2^n} A_0 \\ &= \frac{(2n)!}{2^n 2.4\ldots(2n)} \frac{\sqrt{\pi}}{2} = \frac{(2n)!}{2^{2n+1} n!}\sqrt{\pi} \end{aligned}$$

et

$$A_{2n+1} = \frac{2n}{2} A_{2n-1} = n A_{2n-1} = \cdots = n! A_1 = \frac{n!}{2}.$$

• Nous allons maintenant prouver que $I_n(\lambda) \underset{\lambda\to\infty}{\sim} J_n(\lambda)$. Il suffit de montrer que la différences de ces deux quantités est négligeable devant $J_n(\lambda)$ *i.e.* devant $\frac{1}{\lambda^{\frac{n+1}{2}}}$. Comme pour $x > -1$, $\ln(1+x) \leqslant x$, on a pour tout $t \in [0,1[$, $e^{\lambda \ln(1-t^2)} \leqslant e^{-\lambda t^2}$, si bien que

$$\begin{aligned} 0 \leqslant J_n(\lambda) - I_n(\lambda) &= \int_0^1 \left(e^{-\lambda t^2} - (1-t^2)^\lambda\right) t^n dt \\ &= \int_0^1 e^{-\lambda t^2} t^n \left(1 - e^{\left(\lambda(\ln(1-t^2)+t^2)\right)}\right) dt. \end{aligned}$$

Dans cette dernière intégrale, on reconnaît la fonction intégrée dans $J_n(\lambda)$ à ceci près qu'elle est multipliée par $\left(1 - e^{\left(\lambda(\ln(1-t^2)+t^2)\right)}\right)$, fonction qui tend vers 0 en 0. On va donc essayer de majorer cette quantité au voisinage de 0.

Pour $x \in \left[0, \frac{1}{2}\right[$, la dérivée de $x \longmapsto \ln(1-x) + x$ est $\frac{-x}{1-x}$. La formule des accroissements finis assure l'existence de $c \in]0, x[$ tel que

$$\ln(1-x) + x - 0 = \frac{-c}{1-c} x \quad \text{et} \quad |\ln(1-x) + x| \leqslant 2x^2.$$

On va découper l'intégrale en question en $\alpha \in]0, \frac{1}{2}]$ car si $t \in [0, \alpha[$, $t^2 \leqslant \frac{1}{2}$ et on a donc

$$-\ln(1-t^2) - t^2 \leqslant 2t^4 \quad \text{soit} \quad \ln(1-t^2) + t^2 \geqslant -2t^4.$$

Ainsi, on a la majoration

$$\begin{aligned} 0 \leqslant \mathrm{J}_n(\lambda) - \mathrm{I}_n(\lambda) &\leqslant \int_0^\alpha e^{-\lambda t^2} t^n \left(1 - e^{-2\lambda t^4}\right) \mathrm{d}t + \int_\alpha^1 e^{-\lambda t^2} t^n \mathrm{d}t \\ &\leqslant \int_0^\alpha e^{-\lambda t^2} t^n \left(1 - e^{-2\lambda t^4}\right) \mathrm{d}t + e^{-\lambda \alpha^2} \\ &\leqslant \int_{\mathbb{R}_+} e^{-\lambda t^2} t^n \left(1 - e^{-2\lambda t^4}\right) \mathrm{d}t + e^{-\lambda \alpha^2} \\ &\leqslant \frac{1}{\lambda^{\frac{n+1}{2}}} \int_{\mathbb{R}_+} e^{-u^2} u^n \left(1 - e^{-\frac{2u^4}{\lambda}}\right) \mathrm{d}u + e^{-\lambda \alpha^2}, \end{aligned}$$

par le changement de variables $u = \sqrt{\lambda} t$. Démontrons à l'aide du théorème de convergence dominée que l'intégrale $\int_{\mathbb{R}_+} e^{-u^2} u^n \left(1 - e^{-2u^2/\lambda}\right) \mathrm{d}u$ tend vers 0 lorsque λ tend vers l'infini : pour tout $\lambda > 0$, on a la majoration

$$0 \leqslant e^{-u^2} u^n \left(1 - e^{-2u^4/\lambda}\right) \leqslant e^{-u^2} u^n,$$

et la fonction $u \longmapsto e^{-u^2} u^n$ est intégrable sur $\mathbb{R}_+$. D'autre part, lorsque λ tend vers l'infini, pour tout $u \geqslant 0$,

$$e^{-u^2} u^n \left(1 - e^{-2u^4/\lambda}\right) \xrightarrow[\lambda \to \infty]{} 0.$$

Le théorème de convergence dominée assure donc que

$$0 \leqslant \mathrm{J}_n(\lambda) - \mathrm{I}_n(\lambda) \leqslant o\left(\frac{1}{\lambda^{\frac{n+1}{2}}}\right) + e^{-\lambda \alpha^2} = o\left(\frac{1}{\lambda^{\frac{n+1}{2}}}\right).$$

On peut donc conclure que

$$\mathrm{I}_n(\lambda) \underset{\lambda \to \infty}{\sim} \mathrm{J}_n(\lambda) \underset{\lambda \to \infty}{\sim} \frac{\Lambda_n}{\lambda^{\frac{n+1}{2}}}.$$

• Entamons maintenant la dernière étape qui consiste à montrer que si n est le plus petit entier tel que $f^{(n)}(0) \neq 0$, alors $\mathrm{I}(\lambda)$ est équivalent à $\frac{f^{(n)}(0)}{n!} \mathrm{I}_n(\lambda)$. Notons $a = \frac{f^{(n)}(0)}{n!}$. La fonction $t \longmapsto \frac{f(t) - at^n}{t^{n+1}}$ admet une limite en 0 et se prolonge donc en une fonction continue sur $[0, 1]$. Elle est particulier bornée sur ce compact et il existe donc $\mathrm{M} \geqslant 0$ tel que $|f(t) - at^n| \leqslant \mathrm{M} t^{n+1}$ pour tout $0 \leqslant t \leqslant 1$. Par conséquent, on peut majorer en valeur absolue la différence de $\mathrm{I}(\lambda)$ et $a\mathrm{I}_n(\lambda)$:

$$\begin{aligned} |\mathrm{I}(\lambda) - a\mathrm{I}_n(\lambda)| &= \left| \int_0^1 (1 - t^2)^\lambda (f(t) - at^n) \mathrm{d}t \right| \\ &\leqslant \int_0^1 (1 - t^2)^\lambda |f(t) - at^n| \mathrm{d}t \leqslant \mathrm{M} \int_0^1 (1 - t^2)^\lambda t^{n+1} \mathrm{d}t \\ &\leqslant \mathrm{M} \mathrm{I}_{n+1}(\lambda) \end{aligned}$$

Or, $I_{n+1}(\lambda)$ est négligeable devant $I_n(\lambda)$ d'après les équivalents précédents. Par conséquent, $I(\lambda)$ est équivalent à $aI_n(\lambda)$.

Conclusion. Sachant que n désigne le plus petit entier naturel tel que $f^{(n)}(0) \neq 0$, on a

$$\boxed{I(\lambda) \underset{\lambda\to\infty}{\sim} \frac{A_n f^{(n)}(0)}{n!} \frac{1}{\lambda^{\frac{n+1}{2}}}}.$$

ce qui donne pour n pair $I(\lambda) \underset{\lambda\to\infty}{\sim} \frac{\sqrt{\pi} f^{(n)}(0)}{2^{n+1}(n/2)!} \frac{1}{\lambda^{\frac{n+1}{2}}}$ et pour n impair,

$I(\lambda) \underset{\lambda\to\infty}{\sim} \frac{\left(\frac{n-1}{2}\right)! f^{(n)}(0)}{2(n!)} \frac{1}{\lambda^{\frac{n+1}{2}}}$. ◁

Dans l'exercice suivant on utilise la méthode de Laplace pour obtenir un développement asymptotique de la suite $-x_n$ où x_n est l'unique racine du polynôme de Taylor en 0 de l'exponentielle à l'ordre n, pour n impair. L'équivalent de cette suite avait été obtenu par des méthodes plus élémentaires dans l'exercice 2.41 du tome 1 d'Analyse.

1.42. Méthode de Laplace (2)

1. Soit $\Phi :]a,b[\longrightarrow \mathbb{R}$ une fonction de classe C^2, croissante et $h : [a,b[\longrightarrow \mathbb{R}$ continue $(-\infty < a < b \leqslant +\infty)$. Soit $\xi \in]a,b[$ tel que $h(\xi) \neq 0$ et $\Phi'(\xi) \neq 0$. Soit enfin $(\alpha,\beta) \in \mathbb{R}^2$. On pose

$$I_n = \int_a^{\xi + \alpha\frac{\ln n}{n} + \frac{\beta}{n}} h(t) e^{n\Phi(t)} dt.$$

Trouver un équivalent de I_n.

2. Soit $P_n = \sum_{k=0}^{n} \frac{X^k}{k!}$. Pour n impair, on note $-x_n$ l'unique racine réelle de P_n. Trouver un développement asymptotique de x_n sous la forme $x_n = \xi n + \alpha \ln n + \beta + o(1)$.

(**École normale supérieure**)

▷ **Solution.**

1. On remarque que la fonction Φ étant croissante elle possède une limite finie ou égale à $-\infty$ en a. La fonction $\exp\circ\Phi$ possède donc une limite finie en a. Ceci assure l'existence de l'intégrale I_n.

Nous allons utiliser une variante de la méthode de Laplace. L'idée est que, la fonction Φ étant croissante, n tendant vers $+\infty$ et la fonction exp

croissant vivement, la contribution la plus importante à I_n sera fournie par les valeurs proches de la borne supérieure de l'intégrale, c'est-à-dire proche de ξ. On pose $\varepsilon_n = \alpha\dfrac{\ln n}{n} + \dfrac{\beta}{n}$, quantité qui tend vers 0 quand n tend vers $+\infty$. On multiplie I_n par $e^{-n\Phi(\xi)}$ et on considère l'intégrale

$$J_n = \int_a^{\xi+\varepsilon_n} h(x)e^{n(\Phi(x)-\Phi(\xi))}\,dx.$$

La fonction Φ est croissante donc $\Phi' \geqslant 0$. On a donc $\Phi'(\xi) > 0$ et $h(\xi) \neq 0$. Par continuité, il existe $\delta > 0$ tel que $I = [\xi-\delta, \xi+\delta] \subset]a, b[$ et, pour $x \in I$, $\Phi'(x) > 0$ et $h(x) \neq 0$. On pose $m = \inf_{x\in I} \Phi'(x) > 0$. La fonction Φ est strictement croissante sur I. On introduit enfin une suite (η_n) à termes positifs, à choisir, telle que $\varepsilon_n \underset{n\to\infty}{=} o(\eta_n)$. Il existe $n_0 \in \mathbb{N}^*$ tel que $|\varepsilon_n| < \eta_n < \delta$ pour $n \geqslant n_0$. On suppose désormais $n \geqslant n_0$.

On découpe l'intégrale J_n en trois morceaux :

$$K_1 = \int_a^{\xi-\delta} h(x)e^{n(\Phi(x)-\Phi(\xi))}\,dx,\ K_2 = \int_{\xi-\delta}^{\xi-\eta_n} h(x)e^{n(\Phi(x)-\Phi(\xi))}\,dx,$$

$$\text{et}\quad K_3 = \int_{\xi-\eta_n}^{\xi+\varepsilon_n} h(x)e^{n(\Phi(x)-\Phi(\xi))}\,dx.$$

On a $|K_1| \leqslant e^{n(\Phi(\xi-\delta)-\Phi(\xi))}\int_a^{\xi-\delta} |h(x)|\,dx$, par croissance de Φ. On en déduit que, $K_1 \underset{n\to+\infty}{=} O(e^{n(\Phi(\xi-\delta)-\Phi(\xi))})$, avec $\Phi(\xi-\delta)-\Phi(\xi) < 0$.

La fonction Φ' étant minorée par m, on obtient, par l'inégalité des accroissements finis, pour $x \in [\xi-\delta, \xi-\eta_n]$,

$$e^{n(\Phi(x)-\Phi(\xi))} \leqslant e^{nm(x-\xi)} \leqslant e^{-nm\eta_n}.$$

On en déduit que

$$|K_2| \leqslant e^{-nm\eta_n}\int_{\xi-\delta}^{\xi-\eta_n} |h(x)|\,dx \leqslant e^{-nm\eta_n}\int_{\xi-\delta}^{\xi} |h(x)|\,dx.$$

Si on prend $\eta_n = n^{-\lambda}$ avec $\lambda \in]0,1[$, qui a les propriétés voulues, on a $e^{-nm\eta_n} = e^{-mn^{1-\lambda}}$ et on obtient $K_2 \underset{n\to\infty}{=} O(e^{-mn^{1-\lambda}})$.

Intéressons-nous au dernier morceau. La fonction exp gardant un signe constant, il existe, d'après la formule de la moyenne, $\xi' \in [\xi-\eta_n, \xi+\varepsilon_n]$ tel que

$$K_3 = h(\xi')\int_{\xi-\eta_n}^{\xi+\varepsilon_n} e^{n(\Phi(x)-\Phi(\xi)}\,dx.$$

On applique la formule de Taylor à l'ordre 2. Pour tout $x \in [\xi - \eta_n, \xi + \varepsilon_n]$, il existe $c_x \in [\xi - \eta_n, \zeta + \varepsilon_n]$ tel que

$$\Phi(x) - \Phi(\xi) = (x - \xi)\Phi'(\xi) + \frac{1}{2}(x - \xi)^2 \Phi''(c_x).$$

On en déduit que

$$\mathrm{K}_3 = h(\xi') \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} e^{\frac{n}{2}(x-\xi)^2 \Phi''(c_x)} \, \mathrm{d}x.$$

Si M est un majorant de Φ'' sur I, on a pour $x \in [\xi - \eta_n, \xi + \varepsilon_n]$,

$$\frac{n}{2}(x - \xi)^2 |\Phi''(c_x)| \leqslant \frac{n}{2} \mathrm{M} \eta_n^2 \leqslant \frac{\mathrm{M}}{2} n^{1-2\lambda} \xrightarrow[n \to +\infty]{} 0$$

si $\lambda \in \left]\frac{1}{2}, 1\right[$ (on peut donc prendre par exemple $\eta_n = n^{-\frac{3}{4}}$). Pour ce choix de η_n, on obtient l'encadrement

$$\begin{aligned} e^{-\frac{\mathrm{M}}{2} n^{1-2\lambda}} \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} \, \mathrm{d}x &\leqslant \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(\Phi(x) - \Phi(\xi))} \, \mathrm{d}x \\ &\leqslant e^{\frac{\mathrm{M}}{2} n^{1-2\lambda}} \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} \, \mathrm{d}x \end{aligned}$$

d'où l'on tire les équivalences

$$\mathrm{K}_3 \underset{n \to \infty}{\sim} h(\xi') \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} \, \mathrm{d}x \underset{n \to \infty}{\sim} h(\xi) \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} \, \mathrm{d}x,$$

puisque η_n et ε_n tendent vers 0. Enfin, on obtient

$$\begin{aligned} \int_{\xi - \eta_n}^{\xi + \varepsilon_n} e^{n(x-\xi)\Phi'(\xi)} \, \mathrm{d}x &= \frac{e^{n\varepsilon_n \Phi'(\xi)} - e^{-n\eta_n \Phi'(\xi)}}{n\Phi'(\xi)} \\ &\underset{n \to \infty}{\sim} \frac{e^{\Phi'(\xi)(\alpha \ln n + \beta)} - e^{-\Phi'(\xi) n^{1-\lambda}}}{n\Phi'(\xi)} \\ &\underset{n \to \infty}{\sim} \frac{e^{\Phi'(\xi)(\alpha \ln n + \beta)}}{n\Phi'(\xi)}. \end{aligned}$$

puisque

$$-\Phi'(\xi) n^{1-\lambda} - \Phi'(\xi)(\alpha \ln n + \beta) \underset{n \to \infty}{\sim} -\Phi'(\xi) n^{1-\lambda} \xrightarrow[n \to +\infty]{} -\infty,$$

et donc

$$\mathrm{K}_3 \underset{n \to \infty}{\sim} \frac{h(\xi) e^{\Phi'(\xi)(\alpha \ln n + \beta)}}{n\Phi'(\xi)} \underset{n \to \infty}{\sim} \frac{h(\xi) e^{\beta \Phi'(\xi)}}{\Phi'(\xi)} n^{\alpha \Phi'(\xi) - 1}.$$

Il est clair que K_1 et K_2 sont négligeables devant K_3. Ainsi J_n est équivalent à K_3 et I_n à $e^{n\Phi(\xi)}K_3$. On obtient finalement

$$\boxed{I_n \underset{n\to\infty}{\sim} \frac{h(\xi)e^{\beta\Phi'(\xi)}}{\Phi'(\xi)} n^{\alpha\Phi'(\xi)-1} e^{n\Phi(\xi)}}.$$

2. On montre facilement que pour n pair P_n est toujours positif et que pour n impair P_n possède une unique racine négative notée $-x_n$. Le polynôme P_n est le polynôme de Taylor de exp à l'ordre n. La question 1 suggère d'écrire la formule de Taylor avec reste intégral en 0. On a, pour tout réel x, $e^x = P_n(x) + \frac{1}{n!}\int_0^x e^t(x-t)^n\,dt$. On suppose n impair. On obtient, pour $x = -x_n$, $e^{-x_n} = \frac{1}{n!}\int_0^{-x_n} e^t(-x_n - t)^n\,dt$, ce qui peut s'écrire

$$1 = \frac{1}{n!}\int_0^{-x_n} e^{t+x_n}(-x_n - t)^n\,dt = \frac{1}{n!}\int_0^{x_n} e^u u^n\,du,$$

en faisant le changement de variable $u = x_n + t$ et en tenant compte de l'imparité de n.

Le développement asymptotique cherché suggère que c'est $\frac{x_n}{n}$ qui joue le rôle tenu par $\xi + \varepsilon_n$ dans la question 1. Faisons donc le changement de variable $t = \frac{u}{n}$ dans l'intégrale précédente. On obtient

$$1 = \frac{n^{n+1}}{n!}\int_0^{\frac{x_n}{n}} e^{nt}t^n\,dt = \frac{n^{n+1}}{n!}\int_0^{\frac{x_n}{n}} e^{n(t+\ln t)}\,dt.$$

Déterminons ξ, α et β pour que $\frac{n^{n+1}}{n!}\int_0^{\xi+\alpha\frac{\ln n}{n}+\frac{\beta}{n}} e^{n(t+\ln t)}dt \underset{n\to\infty}{\sim} 1$. Il résulte de la question 1, en prenant $a = 0$, $b = +\infty$, $h(t) = 1$ et $\Phi(t) = t + \ln t$, dont on vérifie aisément qu'ils ont les propriétés voulues, que

$$I_n = \int_0^{\xi+\alpha\frac{\ln n}{n}+\frac{\beta}{n}} e^{n(t+\ln t)}dt \underset{n\to\infty}{\sim} \frac{1}{1+\frac{1}{\xi}} e^{\beta(1+\frac{1}{\xi})} n^{\alpha(1+\frac{1}{\xi})-1} e^{n(\xi+\ln\xi)}.$$

En utilisant la formule de Stirling, on obtient

$$\frac{n^{n+1}}{n!} \underset{n\to\infty}{\sim} \frac{n^{n+1}}{\sqrt{2\pi n}n^n e^{-n}} \underset{n\to\infty}{\sim} \frac{\sqrt{n}e^n}{\sqrt{2\pi}},$$

puis

$$\frac{n^{n+1}}{n!}\int_0^{\xi+\alpha\frac{\ln n}{n}+\frac{\beta}{n}} e^{n(t+\ln t)}dt \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi}(1+\frac{1}{\xi})} e^{\beta(1+\frac{1}{\xi})} n^{\alpha(1+\frac{1}{\xi})-\frac{1}{2}} e^{n(\xi+\ln\xi+1)}.$$

Ceci est équivalent à 1 si

$$\boxed{\alpha = \frac{\xi}{2(1+\xi)} \quad \beta = \frac{\xi}{\xi+1}\ln\left(\sqrt{2\pi}\frac{\xi+1}{\xi}\right) \text{ et } \xi e^{\xi+1} = 1}.$$

Toute solution de l'équation $xe^{x+1} = 1$ est strictement positive. Sur $\mathbb{R}_+$, la fonction $x \longmapsto xe^{x+1}$ croît strictement de 0 à $+\infty$. L'équation possède donc une unique solution ξ. On en déduit un unique triplet (α, β, ξ) ayant les propriétés voulues.

Montrons qu'avec ce triplet $\lim\limits_{n\to+\infty} x_n - (\xi n + \alpha \ln n + \beta) = 0$. Raisonnons par l'absurde. Pour simplifier, posons $y_n = \xi n + \alpha \ln n + \beta$ et, pour tout réel x, $\mathrm{F}_n(x) = \dfrac{n^{n+1}}{n!}\displaystyle\int_0^x e^{n(t+\ln t)}dt$. Ainsi on a $\mathrm{F}_n(\frac{x_n}{n}) = 1$ et $\mathrm{F}_n(\frac{y_n}{n}) \underset{n\to\infty}{\sim} 1$. La fonction F_n est croissante. Supposons qu'il existe $\varepsilon > 0$ tel que $x_n - y_n > \varepsilon$ pour une infinité de valeurs. On a alors pour une infinité de valeurs, $\mathrm{F}_n(\frac{x_n}{n}) \geqslant \mathrm{F}_n(\frac{y_n+\varepsilon}{n})$. Quand n tend vers l'infini, la méthode précédente fournit un équivalent de $\mathrm{F}_n(\frac{y_n+\varepsilon}{n})$: il suffit de remplacer β par $\beta + \varepsilon$. On obtient

$$\mathrm{F}_n(\frac{y_n+\varepsilon}{n}) \underset{n\to\infty}{\sim} \mathrm{F}_n(\frac{y_n}{n})e^{\varepsilon(1+\frac{1}{\xi})} \underset{n\to\infty}{\sim} e^{\varepsilon(1+\frac{1}{\xi})} > 1.$$

On obtiendra $\mathrm{F}_n(\frac{x_n}{n}) > 1$ pour une infinité de valeurs. C'est impossible. On démontre de même qu'il est impossible d'avoir $x_n - y_n < -\varepsilon$ pour une infinité de valeurs. Ainsi, quel que soit $\varepsilon > 0$, on obtient, pour n assez grand, $|x_n - y_n| \leqslant \varepsilon$. On a bien $\lim\limits_{n\to+\infty} x_n - (\xi n + \alpha \ln n + \beta) = 0$, pour les valeurs de α, β et ξ obtenues plus haut, c'est-à-dire

$$\boxed{x_n = \xi n + \alpha \ln n + \beta + o(1)}. \triangleleft$$

L'exercice suivant obtient le développement en série entière de la fonction $x \longmapsto (\arcsin x)^2$ à partir du calcul des intégrales de Wallis. Ce développement sera obtenu par une autre méthode dans l'exercice 3.19 du chapitre sur les séries entières.

1.43. Développement en série entière

1. Calculer $\mathrm{I}_n = \displaystyle\int_0^{\pi/2} (\cos x)^n \mathrm{d}x$ pour tout $n \in \mathbb{N}$.

2. Calculer $\displaystyle\int_0^{\pi/2} \frac{\mathrm{d}x}{1 - t\cos x}$ pour $|t| < 1$.

3. En déduire le développement en série entière de $(\arcsin t)^2$.
(**École polytechnique**)

▷ **Solution.**

1. On reconnaît dans I_n une intégrale de Wallis. On va utiliser la relation de récurrence $n\mathrm{I}_n = (n-1)\mathrm{I}_{n-2}$ qui découle d'une simple intégration par parties (cf. exercice 1.35). On a aisément $\mathrm{I}_0 = \dfrac{\pi}{2}$ et $\mathrm{I}_1 = 1$. On en déduit que

$$\begin{aligned}\mathrm{I}_{2p} &= \frac{2p-1}{2p}\mathrm{I}_{2(p-1)} = \frac{(2p-1)(2p-3)\cdots 1}{(2p)(2p-2)\cdots 2}\mathrm{I}_0 \\ &= \frac{(2p)!}{(2p)^2(2p-2)^2\cdots 2^2}\frac{\pi}{2} = \boxed{\frac{(2p)!}{2^{2p}(p!)^2}\frac{\pi}{2}} \quad \text{et}\end{aligned}$$

$$\mathrm{I}_{2p+1} = \frac{2p}{2p+1}\mathrm{I}_{2p-1} = \frac{(2p)(2p-2)\cdots 2}{(2p+1)(2p-1)\cdots 3}\mathrm{I}_1 = \boxed{\frac{2^{2p}(p!)^2}{(2p+1)!}}.$$

2. Nous allons calculer l'intégrale considérée, que nous noterons $\mathrm{J}(t)$, en faisant le changement de variable $u = \tan\dfrac{x}{2}$. On a $\cos x = \dfrac{1-u^2}{1+u^2}$ et $2\mathrm{d}u = (1+u^2)\mathrm{d}x$, d'où

$$\begin{aligned}\mathrm{J}(t) &= 2\int_0^1 \frac{\mathrm{d}u}{(1+u^2)\left(1-t\left(\dfrac{1-u^2}{1+u^2}\right)\right)} \\ &= 2\int_0^1 \frac{\mathrm{d}u}{1-t+(1+t)u^2} = \frac{2}{1-t}\int_0^1 \frac{\mathrm{d}u}{1+\left(\sqrt{\dfrac{1+t}{1-t}}\,u\right)^2} \\ &= \frac{2}{1-t}\sqrt{\frac{1-t}{1+t}}\arctan\left(\sqrt{\frac{1+t}{1-t}}\right) \\ &= \frac{2}{\sqrt{1-t^2}}\arctan\left(\sqrt{\frac{1+t}{1-t}}\right).\end{aligned}$$

On peut simplifier cette expression. En notant $\theta = \arccos t \in]0,\pi[$, on obtient

$$\begin{aligned}\arctan\left(\sqrt{\frac{1+t}{1-t}}\right) &= \arctan\sqrt{\frac{\cos^2\theta/2}{\sin^2\theta/2}} = \arctan\underbrace{(\operatorname{cotan}\theta/2)}_{\geqslant 0} \\ &= \arctan\left(\tan\left(\frac{\pi}{2}-\frac{\theta}{2}\right)\right)\end{aligned}$$

$$= \frac{\pi}{2} - \frac{\theta}{2} \quad \left(\operatorname{car} \frac{\pi}{2} - \frac{\theta}{2} \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[\right),$$
$$= \frac{\pi}{2} - \frac{\arccos t}{2} = \frac{\pi}{4} + \frac{\arcsin t}{2}.$$

On a finalement

$$\boxed{\int_0^{\pi/2} \frac{\mathrm{d}x}{1 - t\cos x} = \frac{\arcsin t}{\sqrt{1-t^2}} + \frac{\pi}{2\sqrt{1-t^2}}}$$

3. Fixons $t \in]-1, 1[$. Pour tout $x \in \mathbb{R}$, on a $|t\cos x| \leqslant |t|$, si bien que la série $\dfrac{1}{1 - t\cos x} = \sum_{n=0}^{+\infty} t^n \cos^n x$ est normalement convergente sur $\mathbb{R}$. Cela légitime l'interversion intégrale-série qui suit

$$\int_0^{\frac{\pi}{2}} \frac{\mathrm{d}x}{1 - t\cos x} = \int_0^{\frac{\pi}{2}} \sum_{n=0}^{+\infty} t^n \cos^n x \mathrm{d}x = \sum_{n=0}^{+\infty} \int_0^{\frac{\pi}{2}} t^n \cos^n x \mathrm{d}x = \sum_{n=0}^{+\infty} \mathrm{I}_n t^n.$$

Notons pour $t \in]1, 1[$, $f(t) = (\arcsin t)^2$. On a

$$f'(t) = \frac{2\arcsin t}{\sqrt{1-t^2}} = 2\sum_{n=0}^{+\infty} \mathrm{I}_n t^n - \frac{\pi}{\sqrt{1-t^2}}.$$

Or, le développement en série entière de $\dfrac{1}{\sqrt{1-t^2}}$ est connu. Il vaut

$$1 + \left(\frac{-1}{2}\right)(-t^2) + \frac{(-\frac{1}{2})(-\frac{3}{2})(-t^2)^2}{2} + \cdots + \frac{(-\frac{1}{2})\cdots(-\frac{(2p-1)}{2})(-t^2)^p}{p!} + \cdots$$

ce qui donne, après simplification et en faisant apparaître des factorielles comme dans la première question,

$$\frac{1}{\sqrt{1-t^2}} = \sum_{n=0}^{+\infty} \frac{(2p)!}{2^{2p}(p!)^2} t^{2p} = \frac{2}{\pi} \sum_{p=0}^{+\infty} \mathrm{I}_{2p} t^{2p}.$$

Ainsi, il reste $f'(t) = 2\sum_{p=0}^{+\infty} \mathrm{I}_{2p+1} t^{2p+1}$. Comme $f(0) = 0$, il vient par intégration

$$(\arcsin t)^2 = 2\sum_{p=0}^{+\infty} \frac{\mathrm{I}_{2p+1}}{2p+2} t^{2p+2} = 2\sum_{p=0}^{+\infty} \frac{2^{2p}(p!)^2}{(2p+2)!} t^{2p+2} = \sum_{p=0}^{+\infty} \frac{2^{2p+1} t^{2p+2}}{(p+1)^2 \mathrm{C}_{2p+2}^{p+1}},$$

ou encore

$$\boxed{(\arcsin t)^2 = \sum_{p=1}^{+\infty} \frac{2^{2p-1}}{p^2 C_{2p}^p} t^{2p}}. \quad \triangleleft$$

Ce développement en série entière sera obtenu par une autre méthode dans le chapitre série entière (cf. exercice 3.19).

Sont regroupés ci-après des exercices centrés sur les intégrales indéfinies, bien que celles-ci apparaissent déjà dans l'un ou l'autre des exercices précédents. Les deux premiers sont des exemples concrets du lemme de Gronwall qui consiste à obtenir une majoration de la fonction f à partir d'une inégalité de la forme $0 \leqslant f(x) \leqslant a + \int_0^x g(t)f(t)dt$. Ce lemme est notamment utilisé pour établir l'unicité dans le théorème de Cauchy-Lipschitz. Le lecteur trouvera la forme générale avec diverses généralisations dans le troisième tome d'analyse.

1.44. Lemme de Gronwall (1)

Soit $f : [0,1] \longrightarrow \mathbb{R}_+$ une fonction continue. On suppose qu'il existe un réel a tel que pour tout $t \in [0,1]$, $f(t) \leqslant a \int_0^t f(u)\mathrm{d}u$. Que dire de f ?

(École polytechnique)

▷ **Solution.**

Notons $\mathrm{F} : t \longmapsto \int_0^t f(u)\mathrm{d}u$ l'intégrale indéfinie de f. Il s'agit d'une fonction dérivable, qui vérifie par hypothèse $\mathrm{F}' \leqslant a\mathrm{F}$. On en déduit que la fonction $g(t) = \mathrm{F}(t)e^{-at}$ a une dérivée négative et par suite est décroissante sur $\mathbb{R}_+$. Or $g(0) = \mathrm{F}(0) = 0$ et g, tout comme F, est positive. Par suite g est nulle, donc F aussi. Il en résulte que f est la fonction nulle. $\triangleleft$

1.45. Lemme de Gronwall (2)

Soient $f : [1, +\infty[\longrightarrow \mathbb{R}$ positive continue, a et b dans $\mathbb{R}_+$. On suppose que pour $x \geqslant 1$,

$$f(x) \leqslant a \int_1^x \frac{f(t)}{t^2}\mathrm{d}t + b.$$

Trouver une fonction simple majorant f.

(École polytechnique)

▷ **Solution.**

Si $a = 0$, f est majorée par b. Supposons donc $a > 0$.

Posons, pour $x \geqslant 1$, $\mathrm{F}(x) = \displaystyle\int_1^x \frac{f(t)}{t^2}\mathrm{d}t$. En divisant la relation de l'énoncé par x^2, on obtient, pour $x \geqslant 1$, $\mathrm{F}'(x) \leqslant \dfrac{a}{x^2}\mathrm{F}(x) + \dfrac{b}{x^2}$. Comparativement à l'exercice précédent, il n'est pas forcément évident de penser ici à introduire la fonction $\mathrm{G}(x) = \mathrm{F}(x)e^{\frac{a}{x}}$ pour traduire l'inégalité ci-dessus en

$$\mathrm{G}'(x) = \left(\mathrm{F}'(x) - \frac{a}{x^2}\mathrm{F}(x)\right)e^{\frac{a}{x}} \leqslant \frac{b}{x^2}e^{\frac{a}{x}},$$

qui permet ensuite par intégration de majorer G, puis F et enfin f. Donnons donc une méthode générale et naturelle pour travailler avec une telle inéquation différentielle qui explique la provenance de cette fonction G. L'idée est simplement de se ramener à une équation différentielle en donnant un nom à la différence : posons $h(x) = \mathrm{F}'(x) - \dfrac{a}{x^2}\mathrm{F}(x) - \dfrac{b}{x^2}$. La fonction F est solution de cette équation différentielle avec par hypothèse h négative sur $[1, +\infty[$. Or, on sait très bien résoudre cette équation linéaire du premier ordre avec second membre : les solutions de l'équation homogène associée sont les fonctions $\mathrm{C}e^{-\frac{a}{x}}$, $\mathrm{C} \in \mathbb{R}$. On applique alors la méthode de variation de la constante en cherchant F sous la forme $\mathrm{F}(x) = \mathrm{C}(x)e^{-\frac{a}{x}}$ (la fonction C n'est autre que la fonction G parachutée ci-dessus !). En remplaçant dans l'équation on obtient

$$\mathrm{C}'(x) = e^{\frac{a}{x}}\left(\frac{b}{x^2} + h(x)\right) \leqslant e^{\frac{a}{x}}\frac{b}{x^2}.$$

En intégrant, on obtient, pour $x \geqslant 1$,

$$\mathrm{C}(x) = \int_1^x \mathrm{C}'(t)\mathrm{d}t \leqslant \int_1^x \frac{b}{t^2}e^{\frac{a}{t}}\mathrm{d}t = \left[-\frac{b}{a}e^{\frac{a}{t}}\mathrm{d}t\right]_1^x = \frac{b}{a}(e^a - e^{\frac{a}{x}}).$$

On en déduit que $\mathrm{F}(x) = \mathrm{C}(x)e^{-\frac{a}{x}} \leqslant \dfrac{b}{a}(e^{a-\frac{a}{x}} - 1)$, puis en reportant dans l'inégalité de l'énoncé

$$\boxed{f(x) \leqslant be^{a-\frac{a}{x}}}. \quad \triangleleft$$

1.46. Moyenne spatiale

Déterminer les fonctions continues f telles que pour tout $x \in \mathbb{R}$ et tout $a > 0$,

$$f(x) = \frac{1}{2a}\int_{x-a}^{x+a} f.$$

(École polytechnique)

▷ **Solution.**

La condition imposée à f signifie que la valeur moyenne de f sur tout intervalle centré en x est égale à $f(x)$. Visiblement les fonctions affines ont cette propriété. On va montrer que ce sont les seules.

Soit f une fonction vérifiant les conditions de l'énoncé et F une primitive de f sur $\mathbb{R}$. L'égalité $f(x) = \dfrac{1}{2a}\displaystyle\int_{x-a}^{x+a} f$ devient

$$f(x) = \frac{1}{2a}(\mathrm{F}(x+a) - \mathrm{F}(x-a)). \qquad *$$

On va en déduire une relation ne faisant intervenir que la fonction f. On a, pour $x \in \mathbb{R}$ et $a > 0$,

$$\begin{aligned} 4af(x) &= \mathrm{F}(x+2a) - \mathrm{F}(x-2a) \\ &= \mathrm{F}(x+2a) - \mathrm{F}(x) + \mathrm{F}(x) - \mathrm{F}(x-2a) \\ &= 2af(x+a) + 2af(x-a). \end{aligned}$$

On en déduit que, pour $x \in \mathbb{R}$ et $a > 0$, on a

$$2f(x) = f(x+a) + f(x-a).$$

D'autre part, on remarque que la relation $(*)$ implique que f est dérivable sur $\mathbb{R}$. En dérivant par rapport à a la dernière relation, on obtient, pour $x \in \mathbb{R}$ et $a > 0$,

$$0 = f'(x+a) - f'(x-a).$$

On en déduit que la fonction f' est constante, c'est-à-dire que f est affine. ◁

Les exercices qui suivent concernent des équations intégrales.

1.47. Une équation intégrale (1)

Soit $k > 0$. Trouver toutes les fonctions continues $f : \mathbb{R}_+^* \longrightarrow \mathbb{R}$ telles qu'il existe $c \in \mathbb{R}$ vérifiant

$$\forall x > 0, \quad \int_x^{kx} f(t)\mathrm{d}t = c.$$

(École polytechnique)

Solution.

Analyse. Soit f une solution. La fonction φ définie sur l'intervalle $\mathbb{R}_+^*$ par $\varphi(x) = \displaystyle\int_x^{kx} f(t)\mathrm{d}t$ est dérivable sur $\mathbb{R}_+^*$ et pour $x > 0$,

$$\varphi'(x) = kf(kx) - f(x).$$

Ainsi φ est constante si, pour tout $x > 0$.

$$kf(kx) = f(x).$$

En considérant la fonction g définie sur $\mathbb{R}_+^*$ par $g(x) = xf(x)$, on obtient, pour tout $x > 0$

$$g(kx) = g(x). \qquad *$$

Considérons la fonction $h = g \circ \exp$. De $(*)$ on tire, pour tout réel t,

$$h(t) = g(e^t) = g(ke^t) = h(t + \ln k).$$

Donc h doit être périodique de période $\ln k$. D'autre part, f étant continue, g et donc h le sont.

Synthèse. Réciproquement, si h est une fonction appartenant à $\mathcal{C}^0(\mathbb{R}, \mathbb{R})$ et $\ln k$-périodique, alors $g = h \circ \ln$ appartient à $\mathcal{C}^0(\mathbb{R}_+^*, \mathbb{R})$ et vérifie, pour tout $x > 0$

$$g(kx) = h(\ln(kx)) = h(\ln x + \ln k) = h(\ln x) = g(x).$$

La fonction $f : \mathbb{R}_+^* \longrightarrow \mathbb{R}$ définie par $f(x) = \dfrac{h(\ln x)}{x}$ a les propriétés voulues. ◁

1.48. Une équation intégrale (2)

Déterminer les fonctions continues $f : \mathbb{R} \longrightarrow \mathbb{R}$ telles que pour tout $x \in \mathbb{R}$,

$$f(x) = 2\int_0^x \sqrt{f(t)}\mathrm{d}t.$$

(École polytechnique)

Solution.

Soit f une solution. Il est clair que la fonction f est positive ou nulle. Mais pour $x \leqslant 0$, on a $\displaystyle\int_0^x \sqrt{f(t)}\mathrm{d}t \leqslant 0$ donc $f(x) \leqslant 0$. On en déduit que f est nulle sur $\mathbb{R}_-$.

La fonction est dérivable, de dérivée $2\sqrt{f} \geqslant 0$, donc f est croissante.

Sur un intervalle où f ne s'annule pas, on a $\dfrac{f'}{2\sqrt{f}} = 1$. Il existe une constante c telle que $\sqrt{f(x)} = x - c$ et $f(x) = (x-c)^2$. Si f n'est pas la fonction nulle. $\{x, f(x) = 0\}$ est un ensemble non vide (il contient $\mathbb{R}_-$) et majoré puisque f est croissante. Soit $a \geqslant 0$ sa borne supérieure. On a, par continuité de f, $f(a) = 0$ et, pour $x > a$, $f(x) = (x-c)^2$. On a donc $c = a$ et f est définie par

$$f(x) = \begin{cases} 0 & \text{si } x \leqslant a \\ (x-a)^2 & \text{si } x > a. \end{cases}$$

Il est clair réciproquement que toute fonction de ce type avec $a \geqslant 0$ est solution. Il faut signaler que la fonction nulle est aussi solution. ◁

1.49. Une équation intégrale (3)

Trouver les application continues f de $\mathbb{R}$ dans $\mathbb{R}$ telles que pour tout $(x, y) \in \mathbb{R}^2$,

$$f(x)f(y) = \int_{x-y}^{x+y} f(t)\mathrm{d}t.$$

(École polytechnique)

▷ **Solution.**

Si f est une solution qui n'est pas la fonction nulle, l'écriture

$$f(x) = \frac{1}{f(y)} \int_{x-y}^{x+y} f(t)\,\mathrm{d}t,$$

pour $f(y) \neq 0$, montre que f est dérivable sur $\mathbb{R}$. En dérivant par rapport à x, on obtient

$$f'(x) = \frac{1}{f(y)}(f(x+y) - f(x-y)),$$

ce qui montre que f est deux fois dérivable. En prolongeant le raisonnement on montrerait qu'elle est $\mathcal{C}^\infty$. En dérivant la formule initiale par rapport à x puis y, on obtient

$$f'(x)f(y) = f(x+y) - f(x-y) \quad \text{et} \quad f(x)f'(y) = f(x+y) + f(x-y),$$

puis en additionnant

$$f'(x)f(y) + f(x)f'(y) = 2f(x+y).$$

En dérivant de nouveau par rapport à x et y, on obtient

$$2f'(x+y) = f''(x)f(y) + f'(x)f'(y) = f'(x)f'(y) + f(x)f''(y)$$

et donc, pour tout $(x,y) \in \mathbb{R}^2$,

$$f''(x)f(y) = f(x)f''(y).$$

On considérant y tel que $f(y) \neq 0$, on montre l'existence de $\lambda \in \mathbb{R}$ tel que, pour tout réel x,

$$f''(x) = \lambda f(x).$$

D'autre part, en faisant $x = y = 0$ dans l'égalité initiale, on obtient $f(0) = 0$; si on fait $x = y = 0$ dans l'expression de $2f'(x+y)$ on a alors $2f'(0) = (f'(0))^2$. La valeur $f'(0) = 0$ conduisant à la fonction nulle, il faut $f'(0) = 2$.

- Si $\lambda = 0$, on obtient, pour tout x, $f'(x) = 2$ et $f(x) = 2x$.
- Si $\lambda > 0$, on écrit $\lambda = \omega^2$, et il existe $(a,b) \in \mathbb{R}^2$ tel que $f(x) = a\operatorname{ch}(\omega x) + b\operatorname{sh}(\omega x)$. Les conditions initiales imposent $a = 0$ et $b = \dfrac{2}{\omega}$ soit $f(x) = \dfrac{2}{\omega}\operatorname{sh}(\omega x)$.
- Si $\lambda < 0$, on écrit $\lambda = -\omega^2$, il existe $(a,b) \in \mathbb{R}^2$ tel que $f(x) = a\cos(\omega x) + b\sin(\omega x)$. Les conditions initiales imposent $a = 0$ et $b = \dfrac{2}{\omega}$ soit $f(x) = \dfrac{2}{\omega}\sin(\omega x)$.

On vérifie facilement que toutes les fonctions trouvées conviennent. ◁

1.50. Une équation intégrale (4)

1. Existe-t-il une fonction f continue et bornée de $[0,+\infty[$ dans $\mathbb{R}$ telle que, pour tout $x \geqslant 0$, on ait

$$f(x) = 1 + \int_0^x \frac{e^{-t^2}}{1+f(t)^2}\,\mathrm{d}t \ ?$$

2. Étudier la suite de fonctions $f_n : [0,+\infty[\longrightarrow \mathbb{R}$ définie par $f_0(x) = 0$ et, pour tout $(n,x) \in \mathbb{N} \times \mathbb{R}_+$,

$$f_{n+1}(x) = 1 + \int_0^x \frac{e^{-t^2}}{1+f_n(t)^2}\,\mathrm{d}t.$$

(École polytechnique)

▷ **Solution.**

1. Si f existe, elle est dérivable sur $[0,+\infty[$ et, pour tout $x \geqslant 0$, on a $f'(x) = \dfrac{e^{-x^2}}{1+(f(x))^2}$ et donc $f'(x) + f'(x)(f(x))^2 = e^{-x^2}$. On intègre, en notant que l'équation initiale implique $f(0) = 1$. On obtient, pour tout $x \geqslant 0$,

$$f(x) + \frac{1}{3}(f(x))^3 = \int_0^x e^{-t^2}\,dt + \frac{4}{3},$$

ce qui peut s'écrire $\Phi(f(x)) = \displaystyle\int_0^x e^{-t^2}\,dt + \frac{4}{3}$, où Φ est la fonction définie sur $\mathbb{R}$ par

$$\Phi(x) = \frac{1}{3}x^3 + x.$$

On a, pour tout réel x, $\Phi'(x) = x^2 + 1 > 0$. La fonction Φ réalise une bijection de $\mathbb{R}$ sur $\mathbb{R}$ et f est nécessairement définie par

$$f(x) = \Phi^{-1}\left(\int_0^x e^{-t^2}\,dt + \frac{4}{3}\right).$$

Montrons réciproquement que cette fonction a les propriétés voulues. On pose, pour tout $x \geqslant 0$, $g(x) = \displaystyle\int_0^x e^{-t^2}\,dt + \frac{4}{3}$ et donc $f = \Phi^{-1} \circ g$. La fonction Φ^{-1} est dérivable sur $\mathbb{R}$ (puisque Φ' ne s'annule pas) et $(\Phi^{-1})'(x) = \dfrac{1}{\Phi'(\Phi^{-1}(x))}$. La fonction f est donc dérivable sur $[0,+\infty[$ et, pour tout $x \geqslant 0$,

$$f'(x) = (\Phi^{-1})'(g(x))g'(x) = \frac{g'(x)}{\Phi'(\Phi^{-1}(g(x)))} = \frac{e^{-x^2}}{\Phi'(f(x))} = \frac{e^{-x^2}}{1+((f(x))^2}.$$

On a donc, pour tout $x \geqslant 0$,

$$\int_0^x \frac{e^{-t^2}}{1+(f(t))^2}\,dt = \int_0^x f'(t)\,dt = f(x) - f(0) = f(x) - \Phi^{-1}\left(\frac{4}{3}\right) = f(x) - 1.$$

Montrons enfin que f est bornée. La fonction g est croissante et

$$g(0) = \frac{1}{3} = \Phi(1) \quad \text{et} \quad \lim_{x\to+\infty} g(x) = \int_0^{\infty} e^{-t^2}\,dt + \frac{4}{3} = \frac{\sqrt{\pi}}{2} + \frac{4}{3} = \alpha,$$

(il est classique que $\displaystyle\int_0^{+\infty} e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2}$). Puisque Φ^{-1} est strictement croissante, on a, pour tout $x \in \mathbb{R}$, $f(x) \in [1, \Phi^{-1}(\alpha)[$: la fonction f est bornée.

2. Montrons que la suite (f_n) converge uniformément vers f sur $\mathbb{R}$. Notons pour commencer que, pour tout entier n, la fonction f_n est bornée. C'est vrai pour $n = 0$ et pour $n \in \mathbb{N}$, $x \geqslant 0$, on a

$$0 \leqslant \int_0^x \frac{e^{-t^2}}{1+(f_n(t))^2}\,dt \leqslant \int_0^x e^{-t^2}\,dt \leqslant \int_0^{+\infty} e^{-t^2}\,dt = \frac{\sqrt{\pi}}{2} < 1$$

et donc $0 < f_{n+1}(x) < 2$.

Pour $n \in \mathbb{N}$ et $x \geqslant 0$, on écrit

$$|f_{n+1}(x) - f(x)| = \left| \int_0^x \frac{e^{-t^2}(f^2(t) - f_n^2(t))}{(1+f^2(t))(1+f_n^2(t))}\,\mathrm{d}t \right|.$$

On en déduit que

$$\begin{aligned} |f_{n+1}(x) - f(x)| &\leqslant \int_0^x \frac{e^{-t^2}(|f(t) - f_n(t)|)(f(t)+f_n(t))}{(1+f^2(t))(1+f_n^2(t))}\,\mathrm{d}t \\ &\leqslant \|f_n - f\|_\infty \int_0^x \frac{e^{-t^2}((f(t)+f_n(t)}{(1+f^2(t))(1+f_n^2(t))}\,\mathrm{d}t. \end{aligned}$$

Puisque, pour tout $y \geqslant 0$, on a $2y \leqslant 1 + y^2$, on obtient, pour $t \geqslant 0$,

$$\frac{f(t)+f_n(t)}{(1+(f(t))^2)(1+(f_n(t))^2)} \leqslant \frac{f(t)}{1+(f(t))^2} + \frac{f_n(t)}{1+(f_n(t))^2} \leqslant \frac{1}{2} + \frac{1}{2} \leqslant 1$$

et donc, pour $x \geqslant 0$,

$$|f_{n+1}(x) - f(x)| \leqslant \|f_n - f\|_\infty \int_0^x e^{-t^2}\,\mathrm{d}t \leqslant \|f_n - f\|_\infty \int_0^{+\infty} e^{-t^2}\,\mathrm{d}t.$$

Enfin, ceci étant vrai pour tout $x \geqslant 0$, on a

$$\|f_{n+1} - f\|_\infty \leqslant k\|f_n - f\|_\infty,$$

où $k = \displaystyle\int_0^{+\infty} e^{-t^2}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}$. Puisque $k < 1$, on en déduit que la suite $(\|f_n - f\|_\infty)$ converge vers 0 et donc que (f_n) converge uniformément vers f sur $[0, +\infty[$. ◁

Les deux derniers exercices du chapitre concernent la notion d'indice d'un chemin fermé du plan complexe.

1.51. Indice d'une courbe fermée

Soit E l'ensemble des fonctions de classe $\mathcal{C}^1$ de $\mathbb{R}$ dans $\mathbb{C}^*$ qui sont 2π-périodiques. Pour tout $f \in \mathrm{E}$, on pose

$$d(f) = \frac{1}{2i\pi} \int_0^{2\pi} \frac{f'(\theta)}{f(\theta)}\mathrm{d}\theta.$$

1. d est-elle définie ?

2. En utilisant $\psi(t) = \exp\left(\displaystyle\int_0^t \frac{f'(\theta)}{f(\theta)}\mathrm{d}\theta\right)$, montrer que $d(f) \in \mathbb{Z}$.

3. Montrer que

$$\forall f \in \mathrm{E},\ \exists \varepsilon > 0,\ \forall g \in \mathrm{E},\ \|f-g\|_\infty \leqslant \varepsilon \Longrightarrow d(f) = d(g).$$

4. Soit $\mathrm{F} = \{f \in \mathcal{C}^0(\mathbb{R}, \mathbb{C}^*),\ f\ 2\pi\text{-périodique}\}$. Peut-on prolonger d en une fonction définie et continue sur F pour la norme $\|\ \|_\infty$?

(École normale supérieure)

▷ **Solution.**

1. La fonction f étant de classe $\mathcal{C}^1$ et ne s'annulant pas sur $\mathbb{R}$, la fonction $\dfrac{f'}{f}$ est définie et continue sur $\mathbb{R}$, ce qui justifie l'existence de d.

2. La fonction ψ est de classe $\mathcal{C}^1$ puisque $t \longmapsto \displaystyle\int_0^t \frac{f'}{f}$ est $\mathcal{C}^1$ et sa dérivée vaut

$$\psi'(t) = \frac{f'(t)}{f(t)} \exp\left(\int_0^t \frac{f'(\theta)}{f(\theta)} \mathrm{d}\theta\right) = \frac{f'(t)}{f(t)}\psi(t).$$

Autrement dit, tout comme f, ψ est solution de l'équation différentielle linéaire d'ordre 1, $y' - \dfrac{f'}{f}y$. Il existe donc un réel K tel que $\psi = \mathrm{K}f$. Or f est 2π-périodique, donc ψ l'est aussi. Ainsi, $\psi(2\pi) = \psi(0) = 1$ et il en résulte que $\displaystyle\int_0^{2\pi} \frac{f'}{f} \in 2i\pi\mathbb{Z}$ et $d(f) \in \mathbb{Z}$.

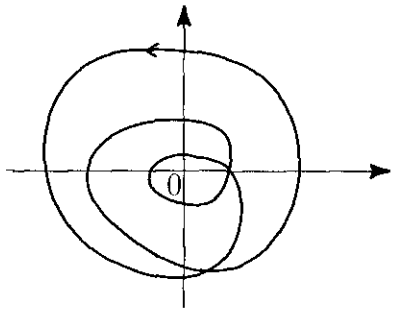
3. Comme f est à valeurs complexes non nulles, de classe $\mathcal{C}^1$, il existe, par le théorème du relèvement, $\rho : \mathbb{R} \longrightarrow \mathbb{R}_+^*$ et $\varphi : \mathbb{R} \longrightarrow \mathbb{R}$ de classe $\mathcal{C}^1$ telles que pour tout $t \in \mathbb{R}$, $f(t) = \rho(t)e^{i\varphi(t)}$. En fait, $\rho = |f|$ et $\varphi(t)$ représente l'argument de $f(t)$ suivi par continuité à partir d'un choix de l'argument de $f(0)$. Comme f est 2π-périodique, $\varphi(0) \equiv \varphi(2\pi) \pmod{2\pi}$. En fait,

$$\begin{aligned}\int_0^{2\pi} \frac{f'}{f} &= \int_0^{2\pi} \left(\frac{\rho'}{\rho} + i\varphi'\right) = [\ln\rho]_0^{2\pi} + [i\varphi]_0^{2\pi} \\ &= \ln\left(\frac{|f(2\pi)|}{|f(0)|}\right) + i(\varphi(2\pi) - \varphi(0)) \\ &= \ln(1) + i(\varphi(2\pi) - \varphi(0)) = i(\varphi(2\pi) - \varphi(0)) \in 2i\pi\mathbb{Z}.\end{aligned}$$

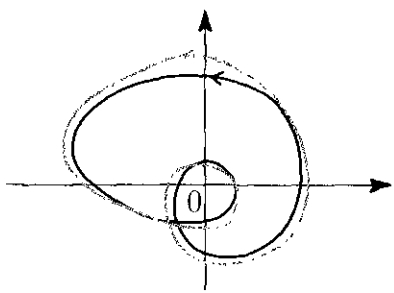
Ainsi, on obtient

$$\boxed{d(f) = \frac{\varphi(2\pi) - \varphi(0)}{2\pi}}$$

et $d(f)$ est le nombre de tours (comptés algébriquement) que fait la courbe fermée $f_{|[0,2\pi]} : [0, 2\pi] \longrightarrow \mathbb{C}^*$ autour de 0. Voici un exemple où l'indice vaut 3 :



On imagine bien que si une courbe g est suffisamment proche de celle de f, elle tournera autant de fois autour de zéro.



Formalisons cela : posons $\varepsilon = \min |f| > 0$ ($|f|$ est continue sur le compact $[0, 2\pi]$ donc est bornée et atteint ses bornes ; son minimum sur $[0, 2\pi]$ l'est aussi sur $\mathbb{R}$ puisque f est 2π-périodique) et prenons $g \in \mathrm{E}$ vérifiant $\|f - g\|_\infty \leqslant \dfrac{\varepsilon}{2}$. Il existe $\tilde{\rho} = |g|$ et $\tilde{\varphi}$ fonctions $\mathcal{C}^1$ de $\mathbb{R}$ dans $\mathbb{R}$ vérifiant $g = \tilde{\rho} e^{i\tilde{\varphi}}$.

Quitte à multiplier f et g par $e^{-i\varphi(0)}$, on peut supposer $\varphi(0) = 0$ (cela ne change ni $\|f - g\|$, ni l'indice de f et g). Quitte à retirer un multiple convenable de 2π à $\tilde{\varphi}$, on peut supposer $\tilde{\varphi}(0) \in [-\pi, \pi]$. On peut écrire la minoration suivante,

$$\frac{\varepsilon}{2} \geqslant |\rho e^{i\varphi} - \tilde{\rho} e^{i\tilde{\varphi}}| \geqslant \rho \left| 1 - \frac{\tilde{\rho}}{\rho} e^{i(\varphi - \tilde{\varphi})} \right| \geqslant \varepsilon \left| 1 - \frac{\tilde{\rho}}{\rho} e^{i(\tilde{\varphi} - \varphi)} \right|$$

et donc,

$$\left| 1 - \frac{\tilde{\rho}}{\rho} e^{i(\tilde{\varphi} - \varphi)} \right| \leqslant \frac{1}{2}.$$

Autrement dit, le complexe $\dfrac{\tilde{\rho}}{\rho} e^{i(\tilde{\varphi} - \varphi)}$ est dans le disque fermé de rayon $\dfrac{1}{2}$ et de centre 1. Cela nous garantit que la fonction $\tilde{\varphi} - \varphi$ prend exclusivement des valeurs comprises entre $\left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right]$ modulo 2π, autrement dit

dans la réunion des intervalles $\left[-\frac{\pi}{2}+2k\pi,\ \frac{\pi}{2}+2k\pi\right]$. Comme la fonction $\tilde{\varphi}-\varphi$ est continue, le théorème des valeurs intermédiaires nous assure que pour tout $x\in[0,2\pi]$, $(\tilde{\varphi}-\varphi)(x)$ reste dans $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$. En particulier, $\tilde{\varphi}(0)$ est dans l'intervalle $\left[-\frac{\pi}{2},\frac{\pi}{2}\right]$.

Or, si on imagine que $d(f)\neq d(g)$, cela signifie que la quantité

$$|\tilde{\varphi}(2\pi)-\tilde{\varphi}(0)-\Big(\varphi(2\pi)-\varphi(0)\Big)|=|\tilde{\varphi}(2\pi)-\tilde{\varphi}(0)-\varphi(2\pi)|\geqslant 2\pi,$$

et par inégalité triangulaire,

$$|\tilde{\varphi}(2\pi)-\varphi(2\pi)|\geqslant 2\pi-\frac{\pi}{2}\geqslant\pi,$$

ce qui constitue une contradiction.

Conclusion. Pour tout $f\in\mathrm{E}$ et tout $g\in\mathrm{E}$, si $\|f-g\|_\infty<\dfrac{\min|f|}{2}$, alors $d(f)=d(g)$.

4. Nous allons démontrer que le prolongement est possible.

• Soit f une fonction de F et posons $\varepsilon=\min|f|>0$. Le théorème de Weierstrass trigonométrique assure l'existence d'une suite de polynômes trigonométriques qui converge uniformément vers f. Si P est un élément de cette suite et si $\|\mathrm{P}-f\|<\varepsilon$, P ne peut s'annuler et $\mathrm{P}\in\mathrm{E}$. Soit $\alpha>0$. On va montrer qu'en choisissant bien α, pour tout g et h dans E vérifiant $\|f-g\|_\infty\leqslant\alpha$ et $\|f-h\|_\infty\leqslant\alpha$, on a l'égalité $d(g)=d(h)$. On a $\|g-h\|_\infty\leqslant 2\alpha$ d'une part, et d'autre part

$$|g|\geqslant|f|-|g-f|\geqslant\varepsilon-\alpha.$$

Or $\varepsilon-\alpha\geqslant 2\alpha$ dès que $\alpha\leqslant\dfrac{\varepsilon}{3}$. En faisant un tel choix, lorsque les conditions $\|f-g\|_\infty\leqslant\alpha$ et $\|f-h\|_\infty\leqslant\alpha$ sont toutes deux réalisées, on a $\|g-h\|_\infty\leqslant\min|g|$ et d'après la question précédente $d(g)=d(h)$.

On définira donc $d(f)$ comme la valeur commune des $d(g)$ où g est une fonction de E dans la boule fermée $\bar{\mathrm{B}}(f,\varepsilon/3)$.

• Montrons enfin que la fonction d ainsi prolongée sur F est continue pour la norme de la convergence uniforme. Pour cela, vérifions qu'elle est localement constante. Soit $f\in\mathrm{F}$ et $\varepsilon=\min|f|$. Prenons $g\in\mathrm{B}(f,\frac{\varepsilon}{3})$. L'intersection des boules ouvertes $\mathrm{B}(f,\frac{\varepsilon}{3})$ et $\mathrm{B}(g,\min|g|/3)$ est un ouvert non vide puisqu'il contient g. On peut trouver d'après le théorème de Weierstrass, une fonction $h:\mathbb{R}\longrightarrow\mathbb{C}$ de classe $\mathcal{C}^1$ dans cette intersection. Comme $\|f-h\|_\infty\leqslant\dfrac{\varepsilon}{3}$, h ne s'annule pas et $h\in\mathrm{E}$. Par définition du prolongement de d, $d(g)=d(h)=d(f)$.

Conclusion. d est localement constante donc continue sur $(\mathrm{F},\|\ \|_\infty)$. ◁

L'entier $d(f)$ représente l'indice de la courbe fermée associée à f.

Le célèbre théorème de d'Alembert-Gauss affirme que tout polynôme non constant à coefficients complexes admet au moins une racine dans $\mathbb{C}$. Par suite un tel polynôme est forcément scindé. L'exercice suivant fournit une preuve de ce théorème qui utilise la notion d'indice étudié dans l'exercice précédent.

1.52. Une preuve du théorème de d'Alembert-Gauss

Pour $f \in \mathcal{C}^1(\mathbb{R}, \mathbb{C})$ fonction 2π-périodique ne s'annulant jamais on pose $\mathrm{I}(f) = \dfrac{1}{2i\pi} \displaystyle\int_0^{2\pi} \frac{f'(t)}{f(t)}\,\mathrm{d}t$.

1. Montrer que $\mathrm{I}(f)$ est un entier.

Soit P un polynôme complexe. On pose $f_{\mathrm{P}}(t) = \mathrm{P}(e^{it})$. On admet le théorème de d'Alembert dans la question 2 mais pas dans la question 3.

2. Caractériser $\mathrm{I}(f_{\mathrm{P}})$ à l'aide des zéros de P.

3. En utilisant $\mathrm{P}(re^{it})$ pour r variable, donner une preuve du théorème de d'Alembert.

(École normale supérieure)

Solution.

1. On se reportera aux deux premières questions de l'exercice précédent.

2. On ne peut définir $\mathrm{I}(f_{\mathrm{P}})$ que si la fonction f_{P} ne s'annule pas sur le cercle unité, c'est-à-dire si P n'a pas de racine de module 1. Faisons cette hypothèse et notons $\alpha_1, \ldots, \alpha_n$ les racines de P où $n = \deg \mathrm{P}$ (on sait que ces racines existent par le théorème de d'Alembert qu'on admet dans cette question). On sait que

$$\frac{\mathrm{P}'}{\mathrm{P}} = \sum_{k=1}^{n} \frac{1}{\mathrm{X} - \alpha_k},$$

On en déduit que, pour tout $t \in [0, 2\pi]$,

$$\frac{f_{\mathrm{P}}'(t)}{f_{\mathrm{P}}(t)} = ie^{it}\frac{\mathrm{P}'(e^{it})}{\mathrm{P}(e^{it})} = \sum_{k=1}^{n} \frac{ie^{it}}{e^{it} - \alpha_k},$$

puis que

$$\mathrm{I}(f_{\mathrm{P}}) = \sum_{k=1}^{n} \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{it}}{e^{it} - \alpha_k}\,\mathrm{d}t.$$

Pour calculer ces intégrales, nous effectuons un développement en série entière de la fraction rationnelle $\dfrac{1}{\mathrm{X} - \alpha_k}$. Il faut distinguer deux cas selon que $|\alpha_k|$ est supérieur ou inférieur à 1.

Si $|\alpha_k| < 1$, on écrit, pour tout $t \in \mathbb{R}$,

$$\frac{e^{it}}{e^{it} - \alpha_k} = \frac{1}{1 - \alpha_k e^{-it}} = \sum_{p=0}^{+\infty} \alpha_k^p e^{-ipt}.$$

Cette série étant normalement convergente, on peut intervertir sommation et intégration. On obtient

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{it}}{e^{it} - \alpha_k} = \frac{1}{2\pi} \sum_{p=0}^{+\infty} \alpha_k^p \int_0^{2\pi} e^{-ipt}\, dt = 1,$$

car seul le terme qui correspond à $p = 0$ n'est pas nul et vaut 2π.

Si $|\alpha_k| > 1$, on écrit, pour tout $t \in \mathbb{R}$,

$$\frac{e^{it}}{e^{it} - \alpha_k} = -\frac{e^{it}}{\alpha_k} \frac{1}{1 - \dfrac{e^{it}}{\alpha_k}} = -\sum_{p=0}^{+\infty} \frac{e^{i(p+1)t}}{\alpha_k^{p+1}}.$$

Comme précédemment, on obtient

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{it}}{e^{it} - \alpha_k} = -\frac{1}{2\pi} \sum_{p=0}^{+\infty} \alpha_k^{-(p+1)} \int_0^{2\pi} e^{i(p+1)t}\, dt = 0,$$

car cette fois toutes les intégrales sont nulles.

On déduit de ces calculs que $I(f_P)$ est égal au nombre de racines du polynôme P qui sont à l'intérieur du disque unité.

3. Démontrons le théorème de d'Alembert en raisonnant par l'absurde. Soit P un polynôme non constant, de degré n. On suppose donc que P ne possède pas de racine. Pour tout $r \geqslant 0$, la fonction φ_r définie par $\varphi_r(t) = P(re^{it})$ ne s'annule pas sur $[0, 2\pi]$ et on peut définir $I(\varphi_r)$. On pose, pour tout $r \geqslant 0$,

$$F(r) = I(\varphi_r) = \frac{1}{2\pi} \int_0^{2\pi} \frac{P'(re^{it}) re^{it}}{P(re^{it})}\, dt.$$

La fonction $(r, t) \longmapsto \dfrac{P'(re^{it}) re^{it}}{P(re^{it})}$ étant continue sur $\mathbb{R}_+ \times [0, 2\pi]$, il résulte des propriétés des intégrales à paramètre que F est continue sur $\mathbb{R}_+$. Puisqu'elle est à valeurs dans $\mathbb{Z}$, on en déduit qu'elle est constante sur $\mathbb{R}_+$. Il est clair que $F(0) = 0$. On a donc $F(r) = 0$ pour tout $r \geqslant 0$. Calculons maintenant la limite de F en $+\infty$ en passant à la limite dans l'intégrale. Quand $|x|$ tend vers $+\infty$, on a $\dfrac{xP'(x)}{P(x)} \sim n$ (quotient des termes de plus haut degré). On en déduit que, pour tout $t \in [0, 2\pi]$, on a $\lim_{r \to +\infty} \dfrac{P'(re^{it}) re^{it}}{P(re^{it})} = n$. Justifions l'échange de la limite et de

l'intégrale en utilisant le théorème de convergence dominée. La fonction $x \longmapsto \dfrac{xP'(x)}{P(x)}$ est continue sur $\mathbb{C}$ et tend vers une limite finie quand $|x|$ tend vers $+\infty$. On en déduit qu'elle est bornée sur $\mathbb{C}$. Il existe donc $K > 0$ tel que, pour tout $(r,t) \in \mathbb{R}_+ \times [0,2\pi]$ on ait $\left|\dfrac{P'(re^{it})re^{it}}{P(re^{it})}\right| \leqslant K$. Le théorème de convergence dominée s'applique et on obtient

$$\lim_{r\to+\infty} F(r) = \frac{1}{2\pi}\int_0^{2\pi}\left(\lim_{r\to+\infty} \frac{P'(re^{it})re^{it}}{P(re^{it})}\right) dt = n.$$

Puisque $n \geqslant 1$ par hypothèse, cela est contradictoire avec le fait que F est nulle sur $\mathbb{R}_+$. L'hypothèse que P n'a pas de racine est donc absurde. ◁

On peut démontrer, comme dans la question précédente, que pour toutes les valeurs de r pour lesquelles P n'a pas de racine de module r, $I(r)$ est égal au nombre de racines de P qui sont à l'intérieur du cercle de centre 0 de rayon r.

Le lecteur trouvera une autre preuve du théorème de d'Alembert-Gauss dans l'exercice 3.41.

Chapitre 2

Suites et séries de fonctions

Jusqu'au XIX^{e} *siècle, la notion de convergence d'une suite de fonctions est utilisée sans soulever les questions fondamentales qu'elle entraîne. Cauchy, dans son* Cours d'Analyse *de l'École Polytechnique (1821), pense même pouvoir démontrer la continuité de la somme d'une série de fonctions continues. Abel relève l'erreur et donne l'exemple de la somme de la série trigonométrique* $\sum_{n=1}^{+\infty} \frac{(-1)^n \sin nx}{n}$ *qui n'est pas continue. Cauchy donnera un énoncé exact en 1853, en utilisant, mais sans la nommer, une hypothèse de convergence uniforme. C'est Seidel et Stokes qui donneront la notion générale de convergence uniforme vers 1848.*

Rappelons qu'une suite de fonctions $(f_n)_{n\geqslant 0}$ *définie sur un ensemble* X *à valeurs réelles ou complexes converge uniformément vers* f *sur* X *si pour tout* $\varepsilon > 0$ *il existe un rang* N *à partir duquel l'écart* $|f(x) - f_n(x)|$ *est majoré par* ε *pour tout point* x *de* X.

La convergence uniforme fournit un critère particulièrement simple pour obtenir la continuité de la limite f *lorsque les fonctions* f_n *sont toutes supposées continues. Elle n'est toutefois pas nécessaire comme on pourra s'en rendre compte en faisant l'exercice 2.10.*

On dit que la série de fonctions $\sum f_n(x)$ *converge uniformément sur* X *si la suite des sommes partielles converge uniformément sur* X. *Cela est notamment réalisé lorsque la série est normalement convergente sur* X, *c'est-à-dire lorsque la série à termes positifs* $\sum \|f_n\|_\infty$ *converge (où* $\|f_n\|_\infty = \sup_{x\in X} |f_n(x)|$*). La réciproque est fausse : il peut y avoir convergence uniforme et divergence de* $\sum \|f_n\|_\infty$. *Toutefois, dans les cas pratiques, comme on le constatera sur les premiers exercices ci-après, les séries de fonctions sont fréquemment normalement convergentes. On pensera ainsi aux séries entières qui convergent normalement sur tout compact inclus dans le disque ouvert de convergence.*

2.1. Majoration sur une demi-droite

Soit $(a_n)_{n\geqslant 1}$ une suite de nombres complexes. On suppose qu'il existe $\rho > 0$ et $M > 0$ tels que pour tout $n \in \mathbb{N}^*$, $\left|\frac{a_n}{n^\rho}\right| \leqslant M$. On pose $f(z) = \sum_{n=1}^{+\infty} a_n e^{2i\pi nz}$.

1. Montrer que f existe sur $P = \{z \in \mathbb{C}, \operatorname{Im} z > 0\}$.

2. Montrer que la série converge uniformément sur tout compact de P.

3. Montrer l'existence de $C > 0$ tel que pour tout $y > 0$ et $x \in \mathbb{R}$

$$|f(x+iy)| \leqslant \frac{C}{y^{1+\rho}}.$$

(École polytechnique)

▷ **Solution.**

1. Soit $z \in P$ et posons $y = \operatorname{Im} z > 0$. On a, pour tout $n \in \mathbb{N}$,

$$|a_n e^{2i\pi nz}| = |a_n| e^{-2\pi ny} \leqslant M n^\rho e^{-2\pi ny} = o\left(\frac{1}{n^2}\right).$$

La série $\sum a_n e^{2i\pi nz}$ converge donc absolument et f est définie sur P.

2. La fonction $z \longmapsto \operatorname{Im} z$ est continue sur $\mathbb{C}$. Sur un compact K de P, elle atteint son minimum $y_0 > 0$. On a, pour $z \in K$ et $n \in \mathbb{N}$, $|a_n e^{2i\pi nz}| \leqslant M n^\rho e^{-2\pi ny_0}$. La série $\sum M n^\rho e^{-2\pi ny_0}$ étant convergente, la série définissant f converge normalement donc uniformément sur K.

Il s'ensuit que f est continue sur P *tout entier.*

3. D'après la question 1, on a pour tout $z \in P$ la majoration $|f(z)| \leqslant M \sum_{n=1}^{+\infty} n^\rho e^{-2\pi ny}$ où $z = x+iy$ avec $y > 0$. Comme $t \longmapsto t^\rho e^{-2\pi ty}$ est décroissante pour t grand, nous allons utiliser une comparaison avec une intégrale pour majorer cette somme. On considère la fonction $\varphi : t \longmapsto t^\rho e^{-2\pi ty}$. La fonction φ est intégrable sur $\mathbb{R}_+$. On a, pour $t > 0$,

$$\varphi'(t) = e^{-2\pi ty}(\rho t^{\rho-1} - 2\pi y t^\rho) = e^{-2\pi ty} t^{\rho-1}(\rho - 2\pi yt).$$

On a donc $\varphi'(t) \leqslant 0$ si $t \geqslant \frac{\rho}{2\pi y}$ de sorte que φ décroît sur $\left[\frac{\rho}{2\pi y}, +\infty\right[$. On pose $\alpha = \frac{\rho}{2\pi y}$ et $n_0 = E(\alpha)$. On a, pour $n \geqslant n_0 + 2$, la majoration $\varphi(n) \leqslant \int_{n-1}^{n} \varphi(t)dt$ et donc $\sum_{n=n_0+2}^{+\infty} \varphi(t)dt \leqslant \int_{n_0+1}^{+\infty} \varphi(t)dt$.

Sur l'intervalle $[\alpha, n_0+1]$, la fonction φ est encore décroissante. On peut écrire

$$\int_{\alpha}^{n_0+1} \varphi(t)\mathrm{d}t \geqslant (n_0+1-\alpha)\varphi(n_0+1),$$

ce qui, additionné à l'inégalité précédente donne

$$\begin{aligned}\sum_{n=n_0+1}^{+\infty} \varphi(n) &\leqslant \int_{\alpha}^{+\infty} \varphi(t)\mathrm{d}t + (\alpha-n_0)\varphi(n_0+1)\\ &\leqslant \int_{\alpha}^{+\infty} \varphi(t)\mathrm{d}t + (\alpha-n_0)\varphi(\alpha),\end{aligned}$$

car le maximum de φ est atteint en α. Enfin, comme φ est croissante sur l'intervalle $[0, n_0]$,

$$\sum_{n=1}^{n_0} \varphi(n) \leqslant n_0\varphi(\alpha).$$

En additionnant, on obtient finalement

$$\sum_{n=1}^{+\infty} \varphi(n) \leqslant \alpha\varphi(\alpha) + \int_{\alpha}^{+\infty} \varphi(t)\mathrm{d}t \leqslant \alpha\varphi(\alpha) + \int_{0}^{+\infty} \varphi(t)\mathrm{d}t.$$

Le changement de variable $u = 2\pi ty$ dans l'intégrale donne

$$\int_0^{+\infty} \varphi(t)\mathrm{d}t = \int_0^{+\infty} t^{\rho}e^{-2\pi ty}\mathrm{d}t = \frac{\mathrm{I}}{(2\pi y)^{\rho+1}}\int_0^{+\infty} u^{\rho}e^{-u}\mathrm{d}u.$$

Par ailleurs, on a $\alpha\varphi(\alpha) = \dfrac{\rho^{\rho+1}e^{-\rho}}{(2\pi y)^{\rho+1}}$. On obtient donc

$$|f(z)| \leqslant \frac{\mathrm{C}}{y^{\rho+1}}, \text{ avec } \mathrm{C} = \frac{\mathrm{M}}{(2\pi)^{\rho+1}}\left(\int_0^{+\infty} u^{\rho}e^{-u}\mathrm{d}u + \rho^{\rho+1}e^{-\rho}\right).$$

On constate que C est bien indépendant de $z \in \mathrm{P}$. ◁

On notera que l'étude de la fonction f se ramène à celle d'une série entière si on pose $u = e^{2i\pi z}$ puisqu'on a alors $f(z) = \sum_{n=1}^{+\infty} a_n u^n$. L'application $\psi : z \mapsto e^{2i\pi z}$ envoie le demi-plan P *dans le disque unité ouvert* D *privé de* 0. *Le résultat de la question 2 s'éclaire : l'image d'un compact* K *de* P *par ψ est un compact* K′ *inclus dans le disque* D *et comme la série entière $\sum_{n=1}^{+\infty} a_n u^n$ a un rayon de convergence $\geqslant 1$ (par l'hypothèse faite sur la suite (a_n)) elle converge normalement sur* K′.

Avec cette traduction, la troisième question de l'exercice consiste à majorer la série entière sur un rayon de D. *Le lecteur pourra se reporter aux exercices 3.22 et 3.23 du chapitre sur les séries entières pour des questions du même type (estimation de la somme d'une série entière en des points du cercle de convergence). La technique de comparaison avec une intégrale utilisée dans la solution ci-dessus est très générale.*

2.2. Une décomposition en série de la valeur absolue

Soit $f : \mathbb{R} \to \mathbb{R}$ la fonction 1-périodique définie par $f(x) = x^2$ pour $|x| \leqslant \frac{1}{2}$. Montrer que pour tout $x \in \mathbb{R}$,

$$\sum_{n=-\infty}^{+\infty} 2^n f\left(\frac{x}{2^n}\right) = |x|.$$

(École polytechnique)

▷ **Solution.**

Posons pour tout x réel, $\varphi(x) = \sum_{n=-\infty}^{+\infty} 2^n f\left(\frac{x}{2^n}\right)$. On va commencer par prouver que φ est bien définie sur $\mathbb{R}$ et même que la série converge normalement sur tout segment $[-A, A]$ avec $A > 0$, ce qui permettra d'obtenir la continuité de φ. Soit $A > 0$ et n_0 un entier naturel tel que $\frac{A}{2^{n_0}} \leqslant \frac{1}{2}$. Pour $n \geqslant n_0$, on a

$$\left|2^n f\left(\frac{x}{2^n}\right)\right| = \frac{x^2}{2^n} \leqslant \frac{A^2}{2^n}.$$

Comme f est bornée par $\frac{1}{4}$, on a pour tout $n < n_0$,

$$\left|2^n f\left(\frac{x}{2^n}\right)\right| \leqslant \frac{1}{4.2^{-n}}.$$

Comme les séries $\sum_{n=n_0}^{+\infty} \frac{A^2}{2^n}$ et $\sum_{n=-\infty}^{n_0-1} \frac{1}{4.2^{-n}}$ convergent, il y a convergence normale sur $[-A, A]$ de la série définissant φ. Comme f est continue, il en va de même de φ sur $[-A, A]$, et finalement sur $\mathbb{R}$, la continuité étant une propriété locale.

La fonction φ étant clairement paire, il suffit de montrer que pour tout $x \geqslant 0$, $\varphi(x) = x$. Par un simple changement d'indices, on constate que

$$\varphi(2x) = \sum_{n=-\infty}^{+\infty} 2^n f\left(\frac{x}{2^{n-1}}\right) = 2 \sum_{n=-\infty}^{+\infty} 2^{n-1} f\left(\frac{x}{2^{n-1}}\right) = 2\varphi(x).$$

Calculons la différence $\varphi(x+1) - \varphi(x)$ pour x dans $[0, 1]$. On a

$$\begin{aligned}
\varphi(x+1)-\varphi(x) &= \sum_{n=-\infty}^{+\infty} 2^n \left(f\left(\frac{x+1}{2^n}\right) - f\left(\frac{x}{2^n}\right)\right) \\
&= \sum_{n=1}^{+\infty} 2^n \left(f\left(\frac{x+1}{2^n}\right) - f\left(\frac{x}{2^n}\right)\right), \text{ par 1-périodicité,} \\
&= \sum_{n=2}^{+\infty} 2^n \left(f\left(\frac{x+1}{2^n}\right) - f\left(\frac{x}{2^n}\right)\right) \\
&\quad + 2\left(\left(\frac{x-1}{2}\right)^2 - \left(\frac{x}{2}\right)^2 \right) \\
&= \sum_{n=2}^{+\infty} \frac{1}{2^n}(2x+1) + \frac{1}{2}(-2x+1) \\
&= x + \frac{1}{2} - x + \frac{1}{2} = 1.
\end{aligned}$$

On a clairement $\varphi(0) = 0$. La relation prouvée à l'instant montre que $\varphi(1) = 1$. Par l'identité $\varphi(2x) = 2\varphi(x)$ on obtient $\varphi\left(\frac{1}{2}\right) = \frac{1}{2}$, puis $\varphi\left(\frac{1}{4}\right) = \frac{1}{4}$. Comme $\varphi\left(\frac{3}{2}\right) = \varphi\left(1+\frac{1}{2}\right) = \varphi\left(\frac{1}{2}\right) + 1 = \frac{3}{2}$ on a aussi $\varphi\left(\frac{3}{4}\right) = \frac{3}{4}$. Montrons plus généralement par récurrence sur $k \geqslant 0$ que $\varphi\left(\frac{p}{2^k}\right) = \frac{p}{2^k}$ pour tout $p \in [\![0, 2^k]\!]$. On vient de voir à l'instant que cela est vrai pour $k = 0, 1, 2$. Supposons le résultat au rang k. Pour $p \in [\![0, 2^k]\!]$ on a directement

$$\varphi\left(\frac{p}{2^{k+1}}\right) = \frac{1}{2}\varphi\left(\frac{p}{2^k}\right) = \frac{1}{2}\cdot\frac{p}{2^k} = \frac{p}{2^{k+1}}.$$

Prenons maintenant p tel que $2^k < p \leqslant 2^{k+1}$. Alors $p' = p - 2^k$ est dans $[\![1, 2^k]\!]$ et

$$\varphi\left(\frac{p}{2^{k+1}}\right) = \frac{1}{2}\varphi\left(\frac{p}{2^k}\right) = \frac{1}{2}\varphi\left(\frac{p'}{2^k}+1\right) = \frac{1}{2}\left(\varphi\left(\frac{p'}{2^k}\right)+1\right) = \frac{p}{2^{k+1}}$$

en utilisant l'hypothèse de récurrence.

Ainsi, φ coïncide avec l'identité sur l'ensemble des nombres dyadiques du segment $[0,1]$. Cet ensemble étant dense dans $[0,1]$ et la fonction φ étant continue, on a $\varphi(x) = x$ pour tout $x \in [0,1]$. La propriété $\varphi(2x) = 2\varphi(x)$ permet alors de montrer facilement que φ est l'identité sur $\mathbb{R}_+$ et donc égale à la fonction valeur absolue sur $\mathbb{R}$ par parité.

Conclusion. Pour tout $x \in \mathbb{R}$, on a

$$\boxed{\sum_{n=-\infty}^{+\infty} 2^n f\left(\frac{x}{2^n}\right) = |x|}. \quad \lhd$$

Il est bien connu que les fonctions continues f de $\mathbb{R}$ dans $\mathbb{R}$ vérifiant $f(x+y)=f(x)+f(y)$ pour tout $(x,y)\in\mathbb{R}^2$ sont les fonctions linéaires $x\mapsto ax$. Dans l'exercice suivant on montre qu'une fonction f pour laquelle la fonction $(x,y)\mapsto f(x+y)-f(x)-f(y)$ est bornée diffère d'une fonction linéaire par une fonction bornée. C'est un grand classique des oraux.

2.3. Fonctions continues presque additives

Soit $f:\mathbb{R}\longrightarrow\mathbb{R}$ continue. On suppose qu'il existe $a>0$ tel que pour tout $(x,y)\in\mathbb{R}^2$

$$|f(x+y)-f(x)-f(y)|\leqslant a.$$

Montrer que pour tout $x\in\mathbb{R}$, $\left(\dfrac{f(2^n x)}{2^n}\right)_{n\in\mathbb{N}}$ a une limite $g(x)$.
Montrer que g est continue linéaire et que $f-g$ est bornée.

(École polytechnique, école normale supérieure)

▷ **Solution.**

Considérons, pour $n\in\mathbb{N}$, la fonction g_n définie par $g_n(x)=\dfrac{f(2^n x)}{2^n}$. Il résulte de l'inégalité vérifiée par f, dans laquelle on remplace x et y par $2^{n-1}x$ que, pour $n\in\mathbb{N}$ et $x\in\mathbb{R}$,

$$|g_n(x)-g_{n-1}(x)|=\frac{1}{2^n}|f(2^n x)-2f(2^{n-1}x)|\leqslant\frac{a}{2^n}.$$

Cela montre que la série de fonctions $\sum(g_n-g_{n-1})$ converge normalement et donc uniformément sur $\mathbb{R}$. La suite de fonction $(g_n)_{n\geqslant 0}$ converge donc uniformément sur $\mathbb{R}$. Appelons g sa limite. Celle-ci est continue sur $\mathbb{R}$ car les fonctions g_n le sont et la convergence est uniforme.

D'autre part, on a, d'après ce qui précède, pour $n\in\mathbb{N}$ et $x\in\mathbb{R}$,

$$|g_n(x)-f(x)|=|g_n(x)-g_0(x)|\leqslant\sum_{k=1}^{n}|g_k(x)-g_{k-1}(x)|\leqslant\sum_{k=1}^{n}\frac{a}{2^k}$$
$$\leqslant\frac{a}{2}\,\frac{1-\frac{1}{2^n}}{1-\frac{1}{2}}\leqslant a.$$

En faisant tendre n vers $+\infty$, on obtient $|g(x)-f(x)|\leqslant a$ pour tout réel x. Ainsi, la fonction $g-f$ est bornée.

Toujours d'après l'inégalité vérifiée par f, on a, pour $(x,y)\in\mathbb{R}^2$ et $n\in\mathbb{N}$,

$$|g_n(x+y) - g_n(x) - g_n(y)| = \frac{1}{2^n}|f(2^n x + 2^n y) - f(2^n x) - f(2^n y)| \leqslant \frac{a}{2^n}.$$

En faisant tendre n vers l'infini, il vient $g(x+y) = g(x) + g(y)$ pour tout couple $(x, y) \in \mathbb{R}^2$.

Comme nous l'avons dit plus haut, il est classique que les morphismes continus du groupe $(\mathbb{R}, +)$ dans lui-même sont les applications linéaires $x \mapsto ax$. Rappelons tout de même la preuve de ce résultat. Un morphisme de groupe g de $(\mathbb{R}, +)$ dans $(\mathbb{R}, +)$ vérifie, pour tout $x \in \mathbb{R}$ et $\lambda \in \mathbb{Q}$, $g(\lambda x) = \lambda g(x)$. En effet, si $\lambda \in \mathbb{Z}$, cela résulte directement du fait que g est un morphisme. Enfin si $\lambda = \dfrac{p}{q}$, avec $(p, q) \in \mathbb{Z}^2$, on écrit $qg(\lambda x) = g(q\lambda x) = g(px) = pg(x)$, d'où résulte $g(\lambda x) = \lambda g(x)$. Si $\lambda \in \mathbb{R}$, on considère une suite de rationnels (λ_n) qui converge vers λ et on utilise la continuité de g. On obtient

$$g(\lambda x) = \lim_{n \to +\infty} g(\lambda_n x) = \lim_{n \to +\infty} \lambda_n g(x) = \lambda g(x),$$

ce qui achève de démontrer que g est linéaire. ◁

Notons que la réciproque est triviale : toute fonction de la forme $x \mapsto ax + h(x)$ où h est continue et bornée vérifie la propriété de l'énoncé. Signalons aussi que ce résultat se généralise (avec la même démonstration) à une fonction $f : \mathrm{E} \to \mathrm{F}$ où E et F sont des espaces vectoriels normés réels, F étant supposé complet (pour que la convergence absolue de $\sum(g_n - g_{n-1})$ entraîne la convergence simple).

Il est plus délicat d'établir la convergence uniforme d'une série de fonctions lorsqu'il n'y a pas convergence normale. Dans ce cas, on peut être amené à utiliser le critère spécial des séries alternées ou à effectuer des transformations d'Abel.

2.4. Contrôle uniforme de séries alternées

On pose pour $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$

$$u_n(x) = (-1)^n \underbrace{\sin(\sin(\ldots(\sin(x))\ldots))}_{n \text{ fois}}.$$

Montrer que la série de fonction $\sum_{n=1}^{+\infty} u_n(x)$ converge simplement mais pas normalement sur $\mathbb{R}$. La convergence est-elle uniforme sur $\mathbb{R}$?

(École polytechnique)

▷ **Solution.**

• Posons pour tout $n \geqslant 1$ et tout $x \in \mathbb{R}$,

$$v_n(x) = \underbrace{\sin(\sin(\ldots(\sin(x))\ldots))}_{n \text{ fois}}.$$

On a $\sin(\mathbb{R}) = [-1,1]$ et $\sin([-1,1]) \subset [-1,1]$. On en déduit que pour tout $n \geqslant 1$, $v_n(x) \in [-1,1]$. Sur $[-1,1]$, $\sin x$ est du signe de x, donc $v_n(x)$ est du signe de $v_1(x)$. La série de terme général $u_n(x) = (-1)^n v_n(x)$ est donc alternée.

Montrons que $\sum u_n(x)$ vérifie pour tout $x \in \mathbb{R}$ le critère spécial des séries alternées. Pour x réel, $|\sin x| \leqslant |x|$, donc pour $n \in \mathbb{N}^*$, $|u_n(x)| \leqslant |u_{n-1}(x)|$. La suite $(|u_n(x)|)$ est décroissante et minorée par 0. Elle converge donc vers un réel ℓ. Pour $n \geqslant 2$, $|u_n(x)| = |\sin u_{n-1}(x)|$ et par passage à la limite, sin étant continue, $|\sin \ell| = |\ell|$. En étudiant les fonctions $x \longmapsto \sin x - x$ et $x \longmapsto \sin x + x$, on constate que nécessairement $\ell = 0$. La suite de terme général $|u_n(x)|$ converge vers 0 en décroissant et la série $\sum u_n(x)$ vérifie les hypothèses du théorème spécial des séries alternées. La série $\sum u_n(x)$ converge donc simplement.

• Montrons qu'elle ne converge pas normalement sur $\mathbb{R}$ *i.e.* que la série $\sum \|u_n\|_\infty$ diverge. Sur $[-1,1]$, la fonction sinus est strictement croissante ; $v_1(\mathbb{R}) = [-1,1]$, donc pour $n \geqslant 2$, $v_n(\mathbb{R}) = [v_{n-1}(-1), v_{n-1}(1)]$ et $\|u_n\|_\infty = v_{n-1}(1)$. Posons pour $n \geqslant 2$, $a_n = v_{n-1}(1)$. La suite $(a_n)_{n\geqslant 2}$ converge vers 0. Cherchons-en un équivalent par la méthode classique suivante (que l'on trouvera détaillée dans l'exercice 2.25 du premier tome d'analyse). Pour $n \geqslant 2$, on a $a_{n+1} = \sin a_n = a_n - \dfrac{a_n^3}{6} + o(a_n^3)$. Pour $\alpha \in \mathbb{R}$, on a

$$\begin{aligned} a_{n+1}^\alpha - a_n^\alpha &= a_n^\alpha \left[\left(1 - \frac{a_n^2}{6} + o(a_n^2)\right)^\alpha - 1 \right] \\ &= a_n^\alpha \left[1 - \frac{\alpha a_n^2}{6} + o(a_n^2) - 1 \right] \underset{n\to\infty}{\sim} -\frac{\alpha}{6} a_n^{\alpha+2}. \end{aligned}$$

Prenons $\alpha = -2$ de sorte que l'équivalent soit une constante. On a alors $a_{n+1}^{-2} - a_n^{-2} \underset{n\to\infty}{\sim} \dfrac{1}{3}$. Comme la série de terme général $\dfrac{1}{3}$ (à termes positifs) diverge, on obtient un équivalent sur les sommes partielles :

$$a_n^{-2} - a_2^{-2} = \sum_{k=2}^{n-2} a_{n+1}^{-2} - a_n^{-2} \underset{n\to\infty}{\sim} \frac{n-3}{3} \underset{n\to\infty}{\sim} \frac{n}{3}$$

d'où $a_n^{-2} \underset{n\to\infty}{\sim} \dfrac{n}{3}$ car a_n^{-2} tend vers $+\infty$ et $a_n \underset{n\to\infty}{\sim} \sqrt{\dfrac{3}{n}}$. On en déduit que $\sum a_n$ diverge et par conséquent $\sum \|u_n\|_\infty$ aussi. La série ne converge pas normalement sur $\mathbb{R}$.

• Elle converge cependant uniformément sur $\mathbb{R}$. En effet, pour tout $x \in \mathbb{R}$ le théorème spécial des séries alternées permet de majorer le reste d'indice n comme suit :

$$\left|\sum_{k=n}^{+\infty} u_k(x)\right| \leqslant |u_n(x)| \leqslant a_n \xrightarrow[n\to+\infty]{} 0. \quad \triangleleft$$

Il existe de nombreux procédés permettant de sommer des séries divergentes. Par exemple on dit qu'une série réelle ou complexe $\sum u_n$ converge au sens de Cesàro lorsque la suite de ses sommes partielles converge en moyenne. Ainsi la série $\sum(-1)^n$ diverge au sens usuel mais converge au sens de Cesàro avec une somme égale à $\frac{1}{2}$. Le théorème de Cesàro montre la cohérence du procédé : si une série converge, alors elle converge au sens de Cesàro avec la même somme.

Un autre procédé classique est la sommation au sens d'Abel : pour une suite u_n donnée on regarde si la série entière $\sum u_n x^n$ a un rayon de convergence supérieur ou égal à 1 et ensuite si sa somme possède une limite lorsque x tend vers 1 par valeurs inférieures. Cette limite est alors appelée somme au sens d'Abel de la série. Par exemple avec la suite $u_n = (-1)^n$ on a $\sum_{n=0}^{+\infty}(-1)^n x^n = \frac{1}{1+x}$ pour tout $x \in]-1,1[$ et la somme tend vers $\frac{1}{2}$ lorsque $x \to 1^-$. La série $\sum(-1)^n$ est donc aussi sommable au sens d'Abel avec une somme qui vaut $\frac{1}{2}$. Le théorème d'Abel-Dirichlet, que le lecteur trouvera dans l'exercice 3.6, montre que ce procédé de sommation est aussi cohérent.

La sommation au sens de Riemann qui fait l'objet de l'exercice suivant est de même nature. Elle intervient dans l'étude des séries trigonométriques. L'énoncé fait démontrer la cohérence du procédé. Il est assez difficile et on sera amené à utiliser une transformation d'Abel.

2.5. Sommation au sens de Riemann

On suppose que la série complexe $\sum a_n$ converge. On pose pour $h \neq 0$,

$$f(h) = a_0 + \sum_{n=1}^{+\infty} a_n \left(\frac{\sin nh}{nh}\right)^2.$$

Étudier l'existence et la continuité de f. La fonction f admet-elle une limite en 0 ?

(École polytechnique)

▷ **Solution.**

• La série $\sum a_n$ étant convergente, il existe $M \geqslant 0$ majorant les $|a_n|$. Dans ces conditions, si $\alpha > 0$, on peut écrire pour tout $|h| > \alpha$,

$$\left| a_n \left(\frac{\sin nh}{nh} \right)^2 \right| \leqslant \frac{M}{n^2 h^2} \leqslant \frac{M}{n^2 \alpha^2}.$$

Cette majoration assure la convergence normale de la série définissant f sur l'ouvert $]-\infty, -\alpha[\cup]\alpha, +\infty[$: la fonction f y est donc définie et continue. Comme α a été choisi quelconque, f est définie et continue sur $\mathbb{R}^*$ tout entier.

• Passons à l'étude de la limite en 0. Comme $\lim\limits_{x \to 0} \dfrac{\sin x}{x} = 1$, il semble raisonnable d'espérer que la limite de f en 0 est $\sum\limits_{n=0}^{+\infty} a_n$.

C'est le cas si la série de terme général a_n est absolument convergente : en effet, la convergence de la série définissant f est alors normale sur $\mathbb{R}^*$ (donc uniforme) puisque pour tout $h \neq 0$,

$$\left| a_n \left(\frac{\sin nh}{nh} \right)^2 \right| \leqslant |a_n|.$$

Le théorème d'interversion des limites justifie la limite attendue.

• La situation est plus délicate lorsque la série $\sum a_n$ est seulement semi-convergente. Nous allons effectuer une transformation d'Abel pour faire apparaître f comme limite uniforme de fonctions continues. Plus précisément, nous allons noter pour tout $h \in \mathbb{R}$, $\varepsilon_0(h) = 1$ et pour $n \geqslant 1$,

$$\varepsilon_n(h) = \begin{cases} \left(\dfrac{\sin nh}{nh} \right)^2 & \text{si } h \neq 0 \\ 1 & \text{si } h = 0. \end{cases}$$

D'autre part, on notera $R_n = \sum\limits_{k=n}^{+\infty} a_k$ pour tout $n \geqslant 0$. Pour $h \neq 0$, $f(h) = \sum\limits_{n=0}^{+\infty} a_n \varepsilon_n(h)$. Si on prolonge f en 0 en posant $f(0) = R_0 = \sum\limits_{n=0}^{+\infty} a_n$, il s'agit de montrer la continuité de f en 0.

Donnons une nouvelle expression de f en effectuant une transformation d'Abel : soit $N \geqslant 1$ et h réel,

$$\begin{aligned} \sum_{n=0}^{N} a_n \varepsilon_n(h) &= \sum_{n=0}^{N} (R_n - R_{n+1}) \varepsilon_n(h) = \sum_{n=0}^{N} R_n \varepsilon_n(h) - \sum_{n=1}^{N+1} R_n \varepsilon_{n-1}(h) \\ &= R_0 + \sum_{n=1}^{N} R_n (\varepsilon_n(h) - \varepsilon_{n-1}(h)) - R_{N+1} \varepsilon_N(h). \end{aligned}$$

Comme $|\varepsilon_N(h)| \leqslant 1$, la limite de $R_{N+1}\varepsilon_N(h)$ est nulle lorsque N tend vers l'infini. En faisant tendre N vers $+\infty$, il en résulte que la série $\sum R_n(\varepsilon_n(h) - \varepsilon_{n-1}(h))$ converge et que

$$\boxed{f(h) = R_0 + \sum_{n=1}^{+\infty} R_n(\varepsilon_n(h) - \varepsilon_{n-1}(h))}.$$

Nous allons prouver la convergence uniforme sur $\mathbb{R}$ de cette série de fonctions en majorant une tranche de Cauchy : prenons h non nul, $N \geqslant 1$ et $p \geqslant 0$. On peut écrire

$$\begin{aligned}
\left|\sum_{n=N}^{N+p} R_n(\varepsilon_n(h) - \varepsilon_{n-1}(h))\right| &\leqslant \sup_{n\geqslant N} |R_n| \sum_{n=N}^{N+p} |\varepsilon_n(h) - \varepsilon_{n-1}(h)| \\
&\leqslant \sup_{n\geqslant N} |R_n| \sum_{n=N}^{N+p} \left|\int_{(n-1)h}^{nh} \frac{d}{du}\left(\frac{\sin^2 u}{u^2}\right) du\right| \\
&\leqslant \sup_{n\geqslant N} |R_n| \sum_{n=N}^{N+p} \int_{(n-1)h}^{nh} \left|\frac{d}{du}\left(\frac{\sin^2 u}{u^2}\right)\right| du \\
&\leqslant \sup_{n\geqslant N} |R_n| \int_{(N-1)h}^{(N+p)h} \left|\frac{d}{du}\left(\frac{\sin^2 u}{u^2}\right)\right| du \\
&\leqslant \sup_{n\geqslant N} |R_n| \int_{\mathbb{R}_+} \left|\frac{d}{du}\left(\frac{\sin^2 u}{u^2}\right)\right| du.
\end{aligned}$$

La dernière inégalité est justifiée, car la fonction $D : u \longmapsto \dfrac{d}{du}\left(\dfrac{\sin^2 u}{u^2}\right)$ est intégrable sur $]0, +\infty[$. En effet, d'une part, D se prolonge par continuité en 0, puisque la fonction $u \longmapsto \dfrac{\sin u}{u}$, prolongée convenablement en 0, est de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$ (elle est développable en série entière sur $\mathbb{R}$). D'autre part,

$$D(u) = \frac{2\sin u(u\cos u - \sin u)}{u^3} = O\left(\frac{1}{u^2}\right) \text{ au voisinage de } +\infty.$$

Le théorème de comparaison des fonctions positives intégrables assure l'intégrabilité de D.

On remarquera que l'inégalité montrée ci-dessus pour h non nul reste encore valable pour $h = 0$. En faisant tendre p vers l'infini, le reste de la série se majore ainsi

$$\left|\sum_{n=N}^{+\infty} R_n(\varepsilon_n(h) - \varepsilon_{n-1}(h))\right| \leqslant \sup_{n\geqslant N} |R_n| \int_{\mathbb{R}_+} \left|\frac{d}{du}\left(\frac{\sin^2 u}{u^2}\right)\right| du \xrightarrow[N\to+\infty]{} 0.$$

La convergence de cette série de fonctions continues étant uniforme sur $\mathbb{R}$, f est continue en 0 et on a bien

$$\boxed{\lim_{h\to 0} f(h) = \sum_{n=0}^{+\infty} a_n}. \lhd$$

Les séries de Dirichlet de la forme $\sum \frac{a_n}{n^s}$ ont une grande importance en arithmétique. Si on prend par exemple la suite a_n constante égale à 1 on obtient la célèbre fonction zêta de Riemann dont les propriétés sont intimement liées à la répartition des nombres premiers. L'énoncé ci-après démontre un résultat sur la convergence uniforme des séries de Dirichlet qu'on peut voir comme l'analogue du théorème d'Abel sur les séries entières (cf. exercice 3.6).

2.6. Convergence uniforme des séries de Dirichlet

Soient $(a_n)_{n\in\mathbb{N}} \in \mathbb{C}^{\mathbb{N}}$ et $(\lambda_n)_{n\in\mathbb{N}}$ une suite strictement croissante de réels tendant vers $+\infty$. Montrer que si la série $\sum a_n e^{-\lambda_n z_0}$ converge, alors la série $\sum a_n e^{-\lambda_n z}$ converge uniformément sur l'ensemble $\{z \in \mathbb{C}, |\arg(z - z_0)| \leqslant \theta_0\}$, où θ_0 est donné dans $\left[0, \frac{\pi}{2}\right[$.

(École normale supérieure)

▷ **Solution.**

Notons qu'il suffit de démontrer la propriété dans le cas où $z_0 = 0$. Le cas général s'y ramène en remplaçant la suite a_n par la suite $a_n e^{-\lambda_n z_0}$. On pose, pour tout $n \in \mathbb{N}$, $A_n = \sum_{k=0}^{n} a_k$. Supposons que $\sum_{n=0}^{+\infty} a_n = 0$, ce qui revient à changer la valeur de a_0 et ne change rien à notre problème. Par ce biais, la suite (A_n) converge vers 0. Soit $z = \rho e^{i\theta}$, avec $|\theta| \leqslant \theta_0$ et $\rho > 0$. Si $0 \leqslant N \leqslant P$,

$$\begin{aligned}
\sum_{n=N}^{P} a_n e^{-\lambda_n z} &= \sum_{n=N}^{P} (A_n - A_{n-1}) e^{-\lambda_n z} \\
&= \sum_{n=N}^{P} A_n e^{-\lambda_n z} - \sum_{n=N-1}^{P-1} A_n e^{-\lambda_{n+1} z} \\
&= \sum_{n=N}^{P-1} A_n (e^{-\lambda_n z} - e^{-\lambda_{n+1} z}) + A_P e^{-\lambda_P z} - A_{N-1} e^{-\lambda_N z}.
\end{aligned}$$

On va majorer cette tranche de Cauchy.

$$\left|\sum_{n=\mathrm{N}}^{\mathrm{P}} a_n e^{-\lambda_n z}\right| \leqslant \sum_{n=\mathrm{N}}^{\mathrm{P}-1} |\mathrm{A}_n||e^{-\lambda_n z} - e^{-\lambda_{n+1} z}| + |\mathrm{A}_\mathrm{P}||e^{-\lambda_\mathrm{P} z}| + |\mathrm{A}_{\mathrm{N}-1}||e^{-\lambda_\mathrm{N} z}|$$

$$\leqslant \sup_{n \geqslant \mathrm{N}-1} |\mathrm{A}_n| \left(\sum_{n=\mathrm{N}}^{\mathrm{P}-1} |e^{-\lambda_n z} - e^{-\lambda_{n+1} z}| + |e^{-\lambda_\mathrm{P} z}| + |e^{-\lambda_\mathrm{N} z}| \right).$$

On a, pour tout $n \in \mathbb{N}$,

$$|e^{-\lambda_n z} - e^{-\lambda_{n+1} z}| = \left| \int_{\lambda_n}^{\lambda_{n+1}} z e^{-tz} \mathrm{d}t \right| \leqslant \rho \int_{\lambda_n}^{\lambda_{n+1}} |e^{-tz}| dt.$$

Par conséquent, on a

$$\sum_{n=\mathrm{N}}^{\mathrm{P}-1} |e^{-\lambda_n z} - e^{-\lambda_{n+1} z}| \leqslant \rho \int_{\lambda_\mathrm{N}}^{\lambda_\mathrm{P}} |e^{-tz}| \mathrm{d}t.$$

On a $|e^{-tz}| = e^{\mathrm{Re}(-tz)} = e^{-t\rho\cos\theta}$. Supposons N choisi tel que $\lambda_n > 0$ pour $n \geqslant \mathrm{N}$. Alors, pour $\lambda_\mathrm{N} \leqslant t \leqslant \lambda_\mathrm{P}$, on a $t > 0$ et donc la majoration $|e^{-tz}| \leqslant e^{-t\rho\cos\theta_0} \leqslant 1$, puisque $\cos\theta \geqslant \cos\theta_0$ si $|\theta| \leqslant \theta_0$; cela conduit à

$$\sum_{n=\mathrm{N}}^{\mathrm{P}-1} |e^{-\lambda_n z} - e^{-\lambda_{n+1} z}| \leqslant \rho \int_{\lambda_\mathrm{N}}^{\lambda_\mathrm{P}} e^{-t\rho\cos\theta_0} \mathrm{d}t$$

$$\leqslant \frac{1}{\cos\theta_0} (e^{-\lambda_\mathrm{N}\rho\cos\theta_0} - e^{-\lambda_\mathrm{P}\rho\cos\theta_0}).$$

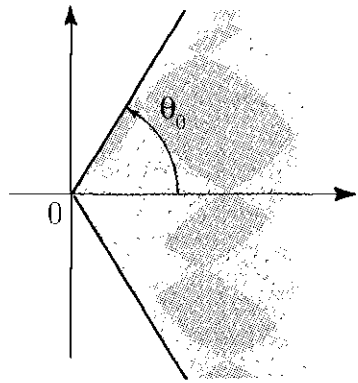
On reprend la majoration de $\left|\sum_{n=\mathrm{N}}^{\mathrm{P}} a_n e^{-\lambda_n z}\right|$:

$$\left|\sum_{n=\mathrm{N}}^{\mathrm{P}} a_n e^{-\lambda_n z}\right| \leqslant \sup_{n \geqslant \mathrm{N}-1} |\mathrm{A}_n| \left(\sum_{n=\mathrm{N}}^{\mathrm{P}-1} |(e^{-\lambda_n z} - e^{-\lambda_{n+1} z})| + |e^{-\lambda_\mathrm{P} z}| + |e^{-\lambda_\mathrm{N} z}| \right)$$

$$\leqslant \sup_{n \geqslant \mathrm{N}-1} |\mathrm{A}_n| \left(\frac{1}{\cos\theta_0} (e^{-\lambda_\mathrm{N}\rho\cos\theta_0} - e^{-\lambda_\mathrm{P}\rho\cos\theta_0}) + 2 \right)$$

$$\leqslant \sup_{n \geqslant \mathrm{N}-1} |\mathrm{A}_n| \left(\frac{1}{\cos\theta_0} + 2 \right).$$

La suite (A_n) convergeant vers 0, il existe, pour $\varepsilon > 0$ donné, un entier n_0 tel que $|\mathrm{A}_n| \leqslant \dfrac{\cos\theta_0}{2\cos\theta_0 + 1}\varepsilon$ pour $n \geqslant n_0$. On a alors, si $n_0 \leqslant \mathrm{N} \leqslant \mathrm{P}$ et $|\arg(z)| \leqslant \theta_0$, $\left|\sum_{n=\mathrm{N}}^{\mathrm{P}} \mathrm{A}_n e^{-\lambda_n z}\right| \leqslant \varepsilon$. La série $\sum a_n e^{-\lambda_n z}$ converge donc uniformément sur l'ensemble $\{z \in \mathbb{C}, |\arg(z)| \leqslant \theta_0\}$ représenté ci-après.



L'exercice suivant donne une condition nécessaire et suffisante pour avoir la convergence uniforme sur $\mathbb{R}$ d'une série trigonométrique en sinus dont les coefficients tendent vers 0 en décroissant. La seconde question est difficile.

2.7. Convergence d'une série trigonométrique

Soit $(b_n)_{n\geqslant 1}$ une suite réelle décroissante qui converge vers 0.

1. Montrer que la série $\sum_{n\geqslant 1} b_n \sin nt$ converge uniformément sur tout segment $[\alpha, 2\pi - \alpha]$ $(0 < \alpha < \pi)$.

2. Montrer l'équivalence des deux conditions :

(*i*) $b_n = o\left(\frac{1}{n}\right)$ lorsque n tend vers l'infini ;

(*ii*) la série $\sum_{n\geqslant 1} b_n \sin at$ converge uniformément sur $\mathbb{R}$.

(École polytechnique)

▷ **Solution.**

1. Posons, pour n entier naturel, $U_n(t) = \sum_{k=1}^{n} \sin kt$. Pour tout réel $t \notin 2\pi\mathbb{Z}$, $U_n(t)$ est la partie imaginaire de

$$1 + e^{it} + e^{2it} + \cdots + e^{int} = \frac{1 - e^{i(n+1)t}}{1 - e^{it}} = \frac{1 - e^{i(n+1)t}}{2ie^{it/2}\sin(t/2)}.$$

Ainsi, $|U_n(t)| \leqslant \dfrac{2}{2|\sin t/2|} = \dfrac{1}{|\sin t/2|}$. Soit $\alpha \in]0, \pi[$ et t un réel quelconque du segment $[\alpha, 2\pi - \alpha]$. On effectue une transformation d'Abel sur la tranche de Cauchy $S_{n,p} = \sum_{k=n}^{n+p} b_k \sin kt$. Il vient :

$$S_{n,p} = \sum_{k=n}^{n+p} b_k(U_k(t) - U_{k-1}(t))$$

$$= b_{n+p}U_{n+p}(t) - b_n U_{n-1}(t) + \sum_{k=n}^{n+p-1} U_k(t)(b_k - b_{k+1}).$$

On majore les termes U_k par l'inégalité ci-dessus. Comme la suite $(b_k)_{k\geqslant 1}$ est décroissante, il vient :

$$|S_{n,p}| \leqslant \frac{b_{n+p} + b_n}{\sin \alpha/2} + \sum_{k=n}^{n+p-1} \frac{b_k - b_{k+1}}{\sin \alpha/2} \leqslant \frac{2b_n}{\sin \alpha/2}.$$

Il en résulte que la série $\sum_{n\geqslant 1} b_n \sin nt$ converge et que cette convergence est uniforme sur le segment $[\alpha, 2\pi - \alpha]$. Évidemment la série converge aussi en un point de $2\pi\mathbb{Z}$, sa somme étant nulle. Elle converge donc pour tout réel t.

2. • Supposons (n) vérifiée et montrons (i). On utilise la tranche de Cauchy $u_n(t) = \sum_{k=n+1}^{2n} b_k \sin kt$. Par hypothèse, elle tend uniformément vers 0 sur $\mathbb{R}$. Pour $k \in [\![n+1, 2n]\!]$ on peut minorer chaque b_k par b_{2n}. Pour $t = \dfrac{\pi}{6n}$ et $k \in [\![n+1, 2n]\!]$ on a $\dfrac{\pi}{6} \leqslant kt \leqslant \dfrac{\pi}{3}$ de sorte que $\sin kt \geqslant \dfrac{1}{2}$. On peut donc écrire

$$u_n\left(\frac{\pi}{6n}\right) \geqslant \sum_{k=n+1}^{2n} b_{2n}\frac{1}{2} = \frac{1}{2}nb_{2n} \geqslant 0.$$

Il en résulte que la suite $2nb_{2n}$ tend vers 0. Or, comme $b_{2n+1} \leqslant b_{2n}$, $(2n+1)b_{2n+1}$ tend aussi vers 0. Il en résulte que $b_n = o\left(\dfrac{1}{n}\right)$.

• Réciproquement, supposons que la suite nb_n tend vers 0. On note $R_n(t) = \sum_{k=n+1}^{+\infty} b_k \sin kt$ le reste d'ordre n de la série. Il s'agit de montrer que R_n tend uniformément vers 0 sur $\mathbb{R}$. Par 2π-périodicité, il suffit de le faire sur $[0, 2\pi]$ et comme $|R_n(2\pi - t)| = |R_n(t)|$ on peut même se limiter à $[0, \pi]$. La première question montre que si $t \in]0, \pi]$,

$$|R_n(t)| \leqslant \frac{2b_n}{\sin t/2}.$$

Le problème est que cette majoration devient inutile au voisinage de 0. On va donc l'améliorer en utilisant le fait que la suite kb_k tend vers 0. Soit $t \in]0, \pi]$, $n \in \mathbb{N}^*$ et $p \in \mathbb{N}^*$. On a :

$$R_n(t) = \sum_{k=n+1}^{n+p} b_k \sin kt + R_{n+p}(t).$$

Dans la première somme on majore $|\sin kt|$ par kt ce qui fait apparaître un terme kb_k et pour le second reste on utilise la majoration ci-dessus. Il vient

$$|R_n(t)| \leqslant t \sum_{k=n+1}^{n+p} kb_k + \frac{2b_{n+p}}{|\sin t/2|} \leqslant tp\left(\sup_{k\geqslant n} kb_k\right) + \frac{2\pi b_{n+p}}{t}$$

puisque $\dfrac{1}{\sin t/2} \leqslant \dfrac{\pi}{t}$. Soit $\varepsilon > 0$. On peut trouver un rang N tel que pour tout $k \geqslant N$, $0 \leqslant kb_k \leqslant \varepsilon$. Si n est un entier quelconque supérieur à N on a donc, pour tout $p \in \mathbb{N}^*$ et tout $t \in]0, \pi]$,

$$|R_n(t)| \leqslant (pt)\varepsilon + \frac{2\pi(n+p)b_{n+p}}{(n+p)t} \leqslant (pt)\varepsilon + \frac{2\pi}{pt}\varepsilon.$$

Comme cela est vrai pour tout p et tout t, la subtilité consiste à appliquer cette inégalité avec un entier p judicieusement choisi comme fonction du réel t : on prend $p = E(1/t) + 1$. On a alors $1 \leqslant pt \leqslant t + 1 \leqslant \pi + 1$ de sorte que

$$\forall n \geqslant N, \quad \forall t \in]0, \pi], \quad |R_n(t)| \leqslant (3\pi + 1)\varepsilon.$$

Cette technique porte le nom de découpe taubérienne. Le lecteur trouvera d'autres exemples d'utilisation de cette technique dans l'exercice 3.26.

Comme $R_n(0) = 0$, l'inégalité demeure vraie pour $t = 0$ et cela signifie exactement que R_n converge uniformément vers 0. ◁

Le même problème pour une série trigonométrique en cosinus est plus facile. On montre, comme dans la première question, que si (a_n) est une suite réelle décroissante de limite nulle, la série $\sum a_n \cos nt$ converge uniformément sur tout segment $[\alpha, 2\pi - \alpha]$ avec $0 < \alpha < \pi$. Pour la convergence uniforme sur $\mathbb{R}$ c'est infiniment plus simple. En effet, si la série converge pour $x = 0$, la série $\sum a_n$ converge et dans ce cas la série $\sum a_n \cos nt$ converge normalement.

2.8. Étude de la convergence d'une série de fonctions

Soit $(f_n)_{n\geqslant 0}$ une suite de fonctions continues sur $[0, 1]$ à valeurs réelles. On suppose que $|f_n(x)f_m(x)| \leqslant 2^{-|n-m|}$ pour tout couple $(n, m) \in \mathbb{N}^2$ et tout $x \in [0, 1]$.

1. Montrer que la série $\sum f_n$ converge simplement sur $[0, 1]$.

2. Montrer que sa somme est bornée.

3. La convergence est-elle uniforme sur $[0, 1]$?

(École normale supérieure)

Solution.

1. Soit $x \in [0,1]$. Si $(f_n(x))_{n\geqslant 0}$ n'est pas la suite nulle, auquel cas la série $\sum f_n(x)$ converge, on considère un entier m tel que $f_m(x) \neq 0$. Pour tout $n \geqslant m$, on a

$$|f_n(x)| \leqslant \frac{2^{-|m-n|}}{|f_m(x)|} \leqslant \frac{2^m}{|f_m(x)|} 2^{-n}.$$

La série $\sum |f_n(x)|$ est majorée par une série géométrique convergente, donc elle converge. La série $\sum f_n(x)$ est donc absolument convergente pour tout $x \in [0,1]$ et la série $\sum f_n$ converge simplement sur $[0,1]$.

2. Montrons que la somme S est bornée. On a, pour tout $x \in [0,1]$,

$$|\mathrm{S}(x)| \leqslant \sum_{n=0}^{+\infty} |f_n(x)|.$$

Soit $x \in [0,1]$. Il existe $m \in \mathbb{N}$ tel que $|f_m(x)| = \max\limits_{n\in\mathbb{N}} |f_n(x)|$. En effet, la suite $(|f_n(x)|)_{n\geqslant 0}$ est bornée puisqu'elle converge vers 0 et on peut poser $\mathrm{M} = \sup\limits_{n\in\mathbb{N}} |f_n(x)|$. Si $\mathrm{M} = 0$, n'importe quel entier m convient, sinon il existe $\mathrm{N} \in \mathbb{N}$ tel que $|f_n(x)| < \mathrm{M}$ si $n > \mathrm{N}$. On alors $\mathrm{M} = \sup\limits_{0\leqslant n\leqslant \mathrm{N}} |f_n(x)| = \max\limits_{0\leqslant n\leqslant \mathrm{N}} |f_n(x)|$ et il existe bien $m \in [\![0, \mathrm{N}]\!]$ tel que $|f_m(x)| = \mathrm{M}$. On a alors, pour tout $n \in \mathbb{N}$,

$$|f_n(x)|^2 \leqslant |f_n(x)|\,|f_m(x)| \leqslant 2^{-|m-n|}$$

et $|f_n(x)| \leqslant (\sqrt{2})^{-|m-n|}$. On en déduit que

$$\begin{aligned}
|\mathrm{S}(x)| &\leqslant \sum_{n=0}^{+\infty} (\sqrt{2})^{-|m-n|} = \sum_{n=0}^{m} (\sqrt{2})^{-m+n} + \sum_{n=m+1}^{+\infty} (\sqrt{2})^{m-n} \\
&= \sum_{k=0}^{m} (\sqrt{2})^{-k} + \sum_{k=1}^{+\infty} (\sqrt{2})^{-k} = \frac{1-(\sqrt{2})^{-(m+1)}}{1-(\sqrt{2})^{-1}} + \frac{(\sqrt{2})^{-1}}{1-(\sqrt{2})^{-1}} \\
&\leqslant \frac{\sqrt{2}}{\sqrt{2}-1} + \frac{1}{\sqrt{2}-1} = (\sqrt{2}+1)^2.
\end{aligned}$$

Cela est valable pour tout $x \in [0,1]$ donc S est bornée sur $[0,1]$.

3. La méthode adoptée précédemment ne permet pas de conclure à la convergence uniforme, car l'entier m dépend de x. Montrons que l'on n'a pas nécessairement convergence uniforme. Pour tout $n \in \mathbb{N}$, considérons la fonction continue $f_n : [0,1] \longrightarrow \mathbb{R}$, nulle sur $\left[0, \frac{1}{2^{2n+2}}\right]$ et $\left[\frac{1}{2^{2n}}, 1\right]$, prenant la valeur 1 en $\frac{1}{2^{2n+1}}$ et affine sur $\left[\frac{1}{2^{2n+2}}, \frac{1}{2^{2n+1}}\right]$ et $\left[\frac{1}{2^{2n+1}}, \frac{1}{2^{2n}}\right]$. Pour $m \neq n$, on a pour tout $x \in [0,1]$, $f_m(x) f_n(x) = 0$ car

les intervalles $\left]\frac{1}{2^{2n+2}}, \frac{1}{2^{2n}}\right[$ et $\left]\frac{1}{2^{2m+2}}, \frac{1}{2^{2m}}\right[$ sont disjoints. Pour $m = n$, $f_m(x)f_n(x) = (f_n(x))^2 \leqslant 1$. L'hypothèse de l'énoncé est donc vérifiée.

Cependant la série $\sum f_n$ ne converge pas uniformément. En effet, en notant R_n le reste d'ordre n de la série, on obtient, pour tout $n \in \mathbb{N}$,

$$R_n\left(\frac{1}{2^{2n+3}}\right) = f_{n+1}\left(\frac{1}{2^{2n+3}}\right) = 1,$$

donc (R_n) ne peut pas converger uniformément vers 0. ◁

Dans les exemples où la convergence n'est pas uniforme c'est souvent à cause de problèmes en un point particulier et dès qu'on se place sur un compact qui évite ce point, la convergence est uniforme. Cette situation est toutefois loin d'être générale. L'exercice suivant donne ainsi un exemple d'une suite de fonctions continues qui converge simplement sur $\mathbb{R}$, mais pour laquelle la convergence n'est uniforme sur aucun intervalle ouvert.

2.9. Suite ne convergeant uniformément sur aucun ouvert

On considère $f_n : \mathbb{R} \longrightarrow \mathbb{R}$ définie par $f_n\left(\frac{1}{n}\right) = 1$, $f_n(x) = 0$ pour $x \in \mathbb{R}_- \cup \left[\frac{2}{n}, +\infty\right[$, et f_n affine sur $\left[0, \frac{1}{n}\right]$ et sur $\left[\frac{1}{n}, \frac{2}{n}\right]$. Soit $m \longmapsto r_m$ une bijection de $\mathbb{N}^*$ sur $\mathbb{Q}$. On pose, pour $t \in \mathbb{R}$,

$$g_n(t) = \sum_{m=1}^{+\infty} \frac{f_n(t - r_m)}{2^m}$$

1. Établir la continuité de g_n.

2. Étudier la convergence simple et uniforme de la suite $(g_n)_{n\geqslant 1}$.

(École polytechnique)

▷ **Solution.**

1. Fixons n dans $\mathbb{N}^*$. On remarque que, f_n étant continue, la fonction $t \longmapsto \frac{f_n(t - r_m)}{2^m}$ est continue pour tout $m \geqslant 1$. Comme $|f_n| \leqslant 1$, la majoration $\left|\frac{f_n(t - r_m)}{2^m}\right| \leqslant \frac{1}{2^m}$ valable pour tout $t \in \mathbb{R}$, garantit la convergence normale sur $\mathbb{R}$ de la série de fonction définissant g_n, qui est donc une fonction continue.

2. Soit $m \geqslant 1$. Pour tout t réel, $\frac{f_n(t - r_m)}{2^m} \xrightarrow[n \to +\infty]{} 0$, car la suite (f_n) converge simplement vers 0. Soit $N \geqslant 1$. On va couper en deux la

somme définissant g_n :

$$0 \leqslant g_n(t) = \sum_{m=1}^{N} \frac{f_n(t-r_m)}{2^m} + \sum_{m=N+1}^{+\infty} \frac{f_n(t-r_m)}{2^m}$$
$$\leqslant \sum_{m=1}^{N} \frac{f_n(t-r_m)}{2^m} + \sum_{m=N+1}^{+\infty} \frac{1}{2^m}.$$

Prenons $\varepsilon > 0$. On peut choisir N de telle manière que le reste de série convergente $\sum_{m=N+1}^{+\infty} \frac{1}{2^m}$ soit inférieur à ε. Dans ces conditions,

$$0 \leqslant g_n(t) \leqslant \sum_{m=1}^{N} \frac{f_n(t-r_m)}{2^m} + \varepsilon \leqslant 2\varepsilon,$$

pour n assez grand, puisque la somme finie tend vers 0 quand n tend vers l'infini. Nous venons donc de prouver que la suite (g_n) converge simplement vers 0.

La convergence n'est pas uniforme sur $\mathbb{R}$: en effet, pour tout $t \in \mathbb{R}$,

$$g_n(t) \geqslant \frac{f_n(t-r_1)}{2} \geqslant 0.$$

Donc $\|g_n\|_\infty \geqslant \sup_{t\in\mathbb{R}} \frac{f_n(t-r_1)}{2} = \frac{1}{2}$.

On a même mieux : la convergence de la suite (g_n) n'est uniforme sur aucun intervalle de longueur non nulle de $\mathbb{R}$. En effet, soit $a < b$. On peut trouver par densité de $\mathbb{Q}$ un rationnel r_m dans $]a,b[$. Dans ces conditions, pour n assez grand, $r_m + \frac{1}{n} \in]a,b[$ et

$$\sup_{t\in]a,b[} g_n(t) \geqslant \sup_{t\in]a,b[} \frac{f_n(t-r_m)}{2^m} = \frac{1}{2^m}.$$

La convergence de g_n n'est donc pas uniforme sur $]a,b[$.

Conclusion. La suite (g_n) converge simplement vers la fonction nulle, sans que la convergence soit uniforme sur aucun intervalle de $\mathbb{R}$ de longueur non nulle. ◁

Au cours des exercices qui précèdent nous avons utilisé à plusieurs reprises le fait qu'une limite uniforme de fonctions continues est continue. L'exercice suivant montre toutefois que la convergence uniforme n'est pas une condition nécessaire pour garantir la continuité d'une limite simple de fonctions continues.

2.10. Continuité et convergence de fonctions

Soit $(f_n)_{n\in\mathbb{N}}$ une suite d'applications de $\mathbb{R}$ dans $\mathbb{R}$. On suppose que cette suite converge simplement vers f sur $\mathbb{R}$. De plus, on suppose que pour tout $\varepsilon > 0$ et tout $n \in \mathbb{N}$

$$\exists n_1, \ldots, n_r \geqslant n,\ \forall x \in \mathbb{R},\ \exists k \in [\![1, r]\!],\ |f_{n_k}(x) - f(x)| \leqslant \varepsilon.$$

1. La suite $(f_n)_{n\in\mathbb{N}}$ converge-t-elle uniformément sur $\mathbb{R}$ vers la fonction f ?

2. On suppose que les fonctions f_n sont toutes continues en un réel x_0. Peut-on en déduire que f est continue en x_0 ?

(École polytechnique)

▷ **Solution.**

1. Montrons que ce n'est pas nécessairement le cas ; prenons $f = 0$ et la suite (f_n) définie de la manière suivante : pour $n \in \mathbb{N}$, on pose $f_{2n}(x) = 0$ pour $x < 0$ et $f_{2n}(x) = \dfrac{x}{2n+1}$ pour $x \geqslant 0$. Quant à f_{2n+1}, on le définit par $f_{2n+1}(x) = f_{2n}(-x)$. La suite (f_n) converge simplement vers la fonction nulle, sans converger uniformément, et comme pour tout $x \in \mathbb{R}$, on a $f_{2n}(x) = f(x)$ ou $f_{2n+1}(x) = f(x)$, la condition de l'énoncé est bien vérifiée.

2. Montrons la continuité de f en x_0 lorsqu'on suppose les f_n continues en ce point. Soit $\varepsilon > 0$. Il existe $\mathrm{N} \geqslant 0$ tel que pour tout $n \geqslant \mathrm{N}$, $|f_n(x_0) - f(x_0)| \leqslant \varepsilon$. Il existe $n_1, \ldots, n_r \geqslant \mathrm{N}$ tels que pour tout $x \in \mathbb{R}$, il existe $k \in [\![1, r]\!]$ avec

$$|f_{n_k}(x) - f(x)| \leqslant \varepsilon.$$

Chaque fonction f_{n_k} étant continue en x_0, il existe $\eta_k > 0$ tel que si $x \in [x_0 - \eta_k, x_0 + \eta_k]$, alors $|f_{n_k}(x) - f(x)| \leqslant \varepsilon$. Notons η le plus petit de ces réels η_k. Prenons $x \in [x_0 - \eta, x_0 + \eta]$ et considérons l'entier k tel que $|f_{n_k}(x) - f(x)| \leqslant \varepsilon$. Dans ces conditions,

$$\begin{aligned} |f(x) - f(x_0)| &\leqslant |f(x) - f_{n_k}(x)| + |f_{n_k}(x) - f_{n_k}(x_0)| + |f_{n_k}(x_0) - f(x_0)| \\ &\leqslant \varepsilon + \varepsilon + \varepsilon = 3\varepsilon, \end{aligned}$$

car $n_k \geqslant \mathrm{N}$ et $|x - x_0| \leqslant \eta_k$. La fonction f est donc continue en x_0. ◁

Lorsqu'on se place sur un compact X *la condition donnée dans l'énoncé équivaut à la continuité de la limite. En effet, le lecteur pourra prouver que si (f_n) est une suite de fonctions continues qui converge simplement sur le compact* X *vers f, alors f est continue si et seulement si pour tout $\varepsilon > 0$ et tout $n \in \mathbb{N}$*

$$\exists n_1, \ldots, n_r \geqslant n,\ \forall x \in \mathrm{X},\ \exists k \in [\![1, r]\!],\ |f_{n_k}(x) - f(x)| \leqslant \varepsilon.$$

Il s'agit toutefois d'une condition de continuité globale. En fait, la continuité en un point est une notion locale et on peut donner une condition nécessaire et suffisante générale pour avoir la continuité d'une limite simple en un point a. Soit (f_n) une suite de fonctions continues en a qui converge simplement vers une fonction f disons sur $\mathbb{R}$. Alors f est continue en a si et seulement si pour tout $\varepsilon > 0$ il existe un rang N tel que pour $n \geqslant \mathrm{N}$, on ait $|f(x) - f_n(x)| \leqslant \varepsilon$ sur un voisinage de a (voisinage qui dépend de n).

On voit bien que cette condition est réalisée lorsque la convergence est supposée uniforme, mais aussi lorsque la suite (f_n) est équicontinue en a (voir l'exercice 2.34 pour la définition de l'équicontinuité). Elle est toutefois beaucoup moins commode à utiliser que la notion de convergence uniforme.

Lorsqu'on étudie la conservation d'une régularité plus fine que la continuité par passage à la limite, les choses se compliquent. Ainsi la limite simple, ou même uniforme, d'une suite (ou une série) de fonctions dérivables n'est pas forcément dérivable, et lorsqu'elle l'est, sa dérivée n'est pas forcément la limite de la suite des fonctions dérivées. L'énoncé suivant en donne une illustration.

2.11. Série normalement convergente de somme non dérivable

Montrer que la fonction $f : x \longmapsto \sum_{n=1}^{+\infty} \dfrac{\sin(2^n x)}{2^n}$ n'est pas dérivable en 0.

(École normale supérieure)

▷ **Solution.**

On notera que la série qui définit f converge normalement sur $\mathbb{R}$ de sorte que f est définie et continue sur $\mathbb{R}$. Pour N entier naturel non nul, considérons $x_{\mathrm{N}} = \dfrac{\pi}{2^{\mathrm{N}+1}}$. Pour n entier $\geqslant \mathrm{N}+1$, $\sin(2^n x_{\mathrm{N}}) = 0$, si bien que

$$f(x_{\mathrm{N}}) = \sum_{n=1}^{\mathrm{N}} \frac{\sin(2^n x_{\mathrm{N}})}{2^n} \geqslant \sum_{n=1}^{\mathrm{N}} \frac{2}{\pi} \frac{2^n x_{\mathrm{N}}}{2^n} = \frac{2}{\pi} \mathrm{N} x_{\mathrm{N}}$$

en vertu de l'inégalité de convexité, $\sin x \geqslant \dfrac{2}{\pi} x$ valable pour x compris entre 0 et $\dfrac{\pi}{2}$. Ainsi, $\lim\limits_{\mathrm{N} \to +\infty} x_{\mathrm{N}} = 0$ et

$$\frac{f(x_{\mathrm{N}})}{x_{\mathrm{N}}} \geqslant \frac{2}{\pi} \mathrm{N} \xrightarrow[\mathrm{N} \to +\infty]{} +\infty.$$

Cela prouve que f n'est pas dérivable en 0. ◁

Signalons que l'on peut démontrer en fait que cette fonction f n'est dérivable nulle part. Le lecteur trouvera un exemple détaillé de fonction continue partout et dérivable nulle part dans l'exercice 2.15.

L'hypothèse essentielle du théorème de dérivation d'une suite de fonctions est la convergence uniforme de la suite des dérivées : on suppose que $f_n : I \longrightarrow \mathbb{R}$ est une suite de fonctions dérivables qui converge simplement vers une fonction f et dont la suite des dérivées (f'_n) converge uniformément sur I vers une fonction g. Alors f est dérivable sur I et sa dérivée est g.

2.12. Régularité d'une série trigonométrique

Soit $(a_n)_{n\in\mathbb{Z}} \in \mathbb{C}^{\mathbb{Z}}$. On pose, pour x réel, $u(x) = \sum_{n\in\mathbb{Z}} a_n e^{inx}$.

1. On suppose que la famille $(n^2 a_n^2)_{n\in\mathbb{Z}}$ est sommable. Montrer l'existence et la continuité de u sur $\mathbb{R}$.

2. Trouver une condition nécessaire et suffisante sur α pour que la sommabilité de la famille $(n^\alpha a_n^2)_{n\in\mathbb{Z}}$ garantisse l'existence et la continuité de u sur $\mathbb{R}$.

3. On suppose que la famille $(n^4 a_n^2)_{n\in\mathbb{Z}}$ est sommable. Que peut-on dire de plus de u ?

(École normale supérieure)

▷ **Solution.**

1. D'après l'inégalité de Cauchy-Schwarz on a pour tout $N \in \mathbb{N}$,

$$\sum_{\substack{|n|\leqslant N\\ n\neq 0}} |a_n| = \sum_{\substack{|n|\leqslant N\\ n\neq 0}} |na_n|\frac{1}{|n|} \leqslant \left(\sum_{\substack{|n|\leqslant N\\ n\neq 0}} n^2|a_n|^2 \sum_{\substack{|n|\leqslant N\\ n\neq 0}} \frac{1}{n^2}\right)^{1/2}$$

$$\leqslant \left(\sum_{n\in\mathbb{Z}} n^2|a_n|^2 \times 2\sum_{n=1}^{+\infty} \frac{1}{n^2}\right)^{1/2}.$$

On en déduit en particulier que les séries $\sum a_n e^{inx}$ et $\sum a_n e^{-inx}$ convergent normalement, donc uniformément sur $\mathbb{R}$ (puisque, pour tout réel x, $|a_n e^{inx}| = |a_n|$), ce qui montre que u est définie sur $\mathbb{R}$. Les fonctions $u_n : x \longmapsto a_n e^{inx}$ étant continues sur $\mathbb{R}$, la fonction u est également continue sur $\mathbb{R}$.

2. Pour $\alpha > 1$, on peut écrire de même,

$$\sum_{\substack{|n|\leqslant N\\ n\neq 0}} |a_n| = \sum_{\substack{|n|\leqslant N\\ n\neq 0}} |n|^{\frac{\alpha}{2}}|a_n|\frac{1}{|n|^{\frac{\alpha}{2}}} \leqslant \left(\sum_{\substack{|n|\leqslant N\\ n\neq 0}} |n|^{\alpha}|a_n|^2 \sum_{\substack{|n|\leqslant N\\ n\neq 0}} \frac{1}{|n|^{\alpha}}\right)^{1/2}$$
$$\leqslant \left(\sum_{n\in\mathbb{Z}} |n|^{\alpha}|a_n|^2 \times 2\sum_{n=1}^{+\infty}\frac{1}{n^{\alpha}}\right)^{1/2}$$

et la conclusion est la même.

Pour $\alpha = 1$, la conclusion n'est plus assurée. Considérons par exemple la suite $(a_n)_{n\in\mathbb{Z}}$ définie par $a_n = \dfrac{1}{n\ln n}$ si $n > 0$ et $a_n = 0$ sinon. On a $na_n^2 = \dfrac{1}{n(\ln n)^2}$. Il est classique la série $\sum \dfrac{1}{n(\ln n)^2}$ converge (c'est une série de Bertrand ; plus généralement la série $\sum \dfrac{1}{n(\ln n)^{\alpha}}$ converge pour $\alpha > 1$). Mais la série $\sum a_n$ diverge (c'est encore une série de Bertrand, avec cette fois-ci $\alpha = 1$). La fonction u n'est pas définie en 0. *A fortiori*, pour $\alpha < 1$, on ne peut pas conclure.

3. Pour tout $n \in \mathbb{Z}$, la fonction u_n est dérivable sur $\mathbb{R}$ et on a pour tout x, $u_n'(x) = ina_n e^{inx}$. Par hypothèse, la famille $(n^4a_n^2)_{n\in\mathbb{Z}}$ est sommable, c'est-à-dire que la famille $(n^2(ina_n)^2)_{n\in\mathbb{Z}}$ est sommable. Le résultat de la question 1 appliqué à $\sum_{n\in\mathbb{Z}} ina_n e^{inx}$ montre que les séries $\sum u_n'$ et $\sum u'_{n'}$ convergent uniformément sur $\mathbb{R}$. Le théorème de dérivation des séries de fonctions permet de conclure que la fonction u est dérivable sur $\mathbb{R}$ et que, pour $x \in \mathbb{R}$,

$$u'(x) = \sum_{n\in\mathbb{Z}} ina_n e^{inx}. \quad \triangleleft$$

2.13. Série de primitives successives

Soit $f_0 : [a,b] \longrightarrow \mathbb{R}$ une fonction continue avec $a < b$. On définit une suite de fonctions $f_n : [a,b] \longrightarrow \mathbb{R}$ par $f_{n+1}(x) = \displaystyle\int_a^x f_n$ pour $n \in \mathbb{N}$ et $x \in [a,b]$.

Étudier et évaluer la fonction $g : x \in [a,b] \longmapsto \displaystyle\sum_{n=1}^{+\infty} f_n(x)$.

(École polytechnique)

Solution.

Comme f_0 est continue, une récurrence immédiate sur $n \geqslant 1$, démontre que la fonction f_n est de classe $\mathcal{C}^n$ et que $f_n' = f_{n-1}$.

Si la série $\sum f_n$ converge et si on peut dériver terme à terme, on obtient

$$g'(x) = \sum_{n=1}^{+\infty} f_n' = \sum_{n=0}^{\infty} f_n = f_0 + g.$$

La fonction g vérifie une équation différentielle linéaire du premier ordre avec second membre dont la résolution avec la méthode de variation de la constante donne

$$\boxed{g(x) = e^x \left(\int_a^x f_0(t) e^{-t} \mathrm{d}t \right)},$$

pour tout $x \in [a, b]$, compte tenu de $g(a) = 0$.

Il nous reste à justifier l'existence de g et la dérivation terme à terme. Majorons $f_n(x)$ pour $x \in [a, b]$ et $n \geqslant 0$. Notons $\mathrm{M}_0 = \|f_0\|_\infty$ (f_0 est continue sur le segment $[a, b]$ donc bornée). On a pour tout $x \in [a, b]$,

$$|f_1(x)| \leqslant \int_a^x \mathrm{M}_0 \mathrm{d}t = \mathrm{M}_0(x - a)$$

et

$$|f_2(x)| = \int_a^x \mathrm{M}_0(t - a)\mathrm{d}t = \frac{\mathrm{M}_0(x - a)^2}{2}.$$

Établissons par récurrence sur $n \geqslant 0$ que $|f_n(x)| \leqslant \dfrac{\mathrm{M}_0(x - a)^n}{n!}$. On vient de le vérifier pour $n = 0$, 1 et 2. Pour $n \geqslant 3$, en supposant l'inégalité au rang $n - 1$, on a pour $x \in [a, b]$,

$$|f_n(x)| \leqslant \int_a^x |f_{n-1}(t)|\mathrm{d}t \leqslant \int_a^x \frac{\mathrm{M}_0(t - a)^{n-1}}{(n-1)!}\mathrm{d}t = \frac{\mathrm{M}_0(x - a)^n}{n!}.$$

En particulier, $\|f_n\| \leqslant \dfrac{\mathrm{M}_0(b - a)^n}{n!}$ et la série $\sum f_n$ converge normalement sur $[a, b]$, ainsi que la série des dérivées. Le théorème de dérivation des suites de fonctions justifie donc la dérivation terme à terme. ◁

2.14. Équivalent d'une série de fonctions

On note $u_n(x) = \dfrac{x^n}{1 - x^n}$ et $f(x) = \sum_{n=1}^{+\infty} u_n(x)$.

1. Étudier la définition, la continuité et la dérivabilité de f.

2. Donner un équivalent de f en 1^-.

3. Démontrer que pour $|x| < 1$, $f(x) = \sum_{n=1}^{+\infty} d(n)x^n$ où $d(n)$ désigne le nombre de diviseurs positifs de n.

(École polytechnique)

Solution.

1. Pour $|x| < 1$, le terme $|u_n(x)|$ est équivalent à $|x|^n$ lorsque n tend vers $+\infty$. Le théorème de comparaison des séries à termes positifs assure la convergence de la série $\sum u_n(x)$. En revanche, si $x = \pm 1$, $u_n(x)$ n'est pas défini pour tout n, donc f n'est pas définie pour ces valeurs là. Enfin, si $|x| > 1$, $|u_n(x)|$ tend vers 1 lorsque n tend vers l'infini et la série $\sum u_n(x)$ diverge. Le domaine de définition de f est donc $]-1, 1[$.

Soit a un réel de $]0, 1[$. La majoration,

$$\left|\frac{x^n}{1-x^n}\right| \leqslant \frac{a^n}{1-a^n}$$

pour $|x| \leqslant a$, assure la convergence normale de la série $\sum u_n(x)$ sur $[-a, a]$. Les u_n étant continues sur $]-1, 1[$, f est continue sur $[-a, a]$, et donc sur $]-1, 1[$ puisque la continuité est une propriété locale.

Chaque u_n est dérivable avec $u_n'(x) = \dfrac{nx^{n-1}}{(1-x^n)^2}$ et si $|x| \leqslant a$,

$$|u_n'(x)| \leqslant \frac{na^{n-1}}{(1-a^n)^2} \underset{n\to\infty}{\sim} na^{n-1} = o\left(\frac{1}{n^2}\right).$$

Le théorème de dérivation des suites de fonctions permet de conclure à la dérivabilité de f sur $[-a, a]$ pour tout $a < 1$, et donc sur $]-1, 1[$.

2. Cherchons un équivalent de f en 1^- par la méthode classique de comparaison série intégrale : pour $x \in]0, 1[$, la fonction $\varphi : t \longmapsto \dfrac{x^t}{1-x^t}$ (de dérivée $t \longmapsto \dfrac{\ln(x)x^t}{(1-x^t)^2} \leqslant 0$) est décroissante sur $\mathbb{R}_+^*$. Par conséquent, on a pour tout $n \geqslant 2$ l'encadrement

$$\int_n^{n+1} \varphi(t)\mathrm{d}t \leqslant \int_n^{n+1} \varphi(n)\mathrm{d}t = u_n(x) = \int_{n-1}^{n} \varphi(n)\mathrm{d}t \leqslant \int_{n-1}^{n} \varphi(t)\mathrm{d}t.$$

La fonction φ est continue et intégrable sur $[1, +\infty[$ ($\varphi(t)$ est équivalent à x^t négligeable devant $\dfrac{1}{t^2}$ lorsque t tend vers $+\infty$). En sommant de $n = 2$ à $n = \mathrm{N}$, on obtient

$$\int_2^{\mathrm{N}+1} \varphi(t)\mathrm{d}t \leqslant \sum_{n=2}^{\mathrm{N}} u_n(x) \leqslant \int_1^{\mathrm{N}} \varphi(t)\mathrm{d}t,$$

et en faisant tendre N vers l'infini, on obtient

$$\int_2^{+\infty} \varphi(t)\mathrm{d}t \leqslant \sum_{n=2}^{+\infty} u_n(x) \leqslant \int_1^{+\infty} \varphi(t)\mathrm{d}t. \qquad (*)$$

En effectuant le changement de variables $u = x^t$ (sur un segment $[a, b]$ de $\mathbb{R}_+^*$ puis en faisant tendre b vers l'infini), on trouve

$$\int_a^{+\infty} \varphi(t)\mathrm{d}t = -\frac{1}{\ln x}\int_0^{x^a} \frac{\mathrm{d}u}{1-u} = \frac{\ln(1-x^a)}{\ln x}.$$

Or, pour x tendant vers 1,

$$\begin{aligned}\ln(1-x^a) &= \ln(1-x) + \ln\left(\frac{1-x^a}{1-x}\right)\\ &= \ln(1-x) + \ln a + o(1) \underset{x\to 1}{\sim} \ln(1-x).\end{aligned}$$

Dans l'encadrement (∗), les membres de gauche et de droite sont équivalents à $\dfrac{\ln(1-x)}{\ln x} \underset{x\to 1^-}{\sim} \dfrac{\ln(1-x)}{x-1}$ lorsque x tend vers 1^-. Comme $u_1(x)$ est négligeable devant $\dfrac{\ln(1-x)}{x-1}$, on obtient

$$\boxed{f(x) \underset{x\to 1^-}{\sim} \frac{\ln(1-x)}{x-1}}. \lhd$$

3. On peut écrire pour $n \geqslant 1$ et $|x| < 1$

$$\frac{x^n}{1-x^n} = x^n \sum_{k=0}^{+\infty} (x^n)^k = \sum_{k=1}^{+\infty} x^{nk}.$$

Comme la série de terme général $\displaystyle\sum_{k=1}^{n} |x|^{nk} = \frac{|x|^n}{1-|x|^n}$ converge, la famille $(x^{nk})_{n,k\geqslant 1}$ est sommable en vertu des théorèmes sur les familles positives sommables (cf. tome 1 d'analyse). Le théorème d'associativité permet alors de rassembler les couples (n, k) pour lesquels le produit nk vaut p. Il y en a autant que de diviseurs de p, si bien que

$$f(x) = \sum_{(n,k)\in(\mathbb{N}^*)^2} x^{nk} = \sum_{p=1}^{+\infty} d(p)x^p.$$

Le lecteur pourra compléter cette étude en s'intéressant aux séries de Lambert $\displaystyle\sum_{n=1}^{+\infty} \frac{n^\alpha x^n}{1-x^n}$ *avec* $\alpha \in \mathbb{N}^*$ *(voir l'exercice 3.25).*

La théorie des suites et séries de fonctions permet souvent d'exprimer des fonctions solutions d'un problème donné tel qu'une équation différentielle (par exemple, sous la forme d'une série entière, ou d'une série de Fourier). Dans l'exercice suivant, nous allons utiliser une suite de fonctions affines par morceaux pour construire un exemple de fonction continue nulle part dérivable. C'est à Bolzano que l'on doit ce travail.

2.15. Courbe de Bolzano

On définit la suite de fonctions $(f_n)_{n\in\mathbb{N}}$ de $[0,1]$ dans $[0,1]$ par $f_0(x) = x$, f_{n+1} est affine sur chacun des intervalles $\left[\frac{k}{3^{n+1}}, \frac{k+1}{3^{n+1}}\right]$ et vérifie

$$f_{n+1}\left(\frac{k}{3^n}\right) = f_n\left(\frac{k}{3^n}\right), \quad f_{n+1}\left(\frac{k}{3^n} + \frac{1}{3^{n+1}}\right) = f_n\left(\frac{k}{3^n} + \frac{2}{3^{n+1}}\right)$$

$$\text{et} \quad f_{n+1}\left(\frac{k}{3^n} + \frac{2}{3^{n+1}}\right) = f_n\left(\frac{k}{3^n} + \frac{1}{3^{n+1}}\right).$$

1. Montrer que $|f_n(x) - f_n(y)| \leqslant 2^n|x-y|$ pour tout couple $(x,y) \in [0,1]^2$.

2. Montrer que pour tout $x \in [0,1]$, $|f_{n+1}(x) - f_n(x)| \leqslant \frac{2^n}{3^{n+1}}$.

3. Montrer que la suite $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers une fonction continue f.

4. Démontrer que pour $p \in [\![0, 3^n]\!]$, $f\left(\frac{p}{3^n}\right) = f_n\left(\frac{p}{3^n}\right)$.

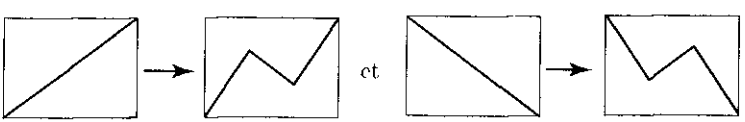
5. Montrer que f n'est dérivable en aucun point de $[0,1]$ de la forme $\frac{p}{3^n}$ où $p \in [\![0, 3^n]\!]$.

6. Montrer que f n'est dérivable en aucun point de $[0,1]$ (on considérera la fonction $(h,k) \longmapsto \frac{f(x+h) - f(x-k)}{h+k}$).

(École polytechnique)

Solution.

1. Chaque fonction f_n est affine par morceaux. Le dessin suivant montre comment on passe de f_n à f_{n+1} sur un intervalle du type $\left[\frac{k}{3^n}, \frac{k+1}{3^n}\right]$:

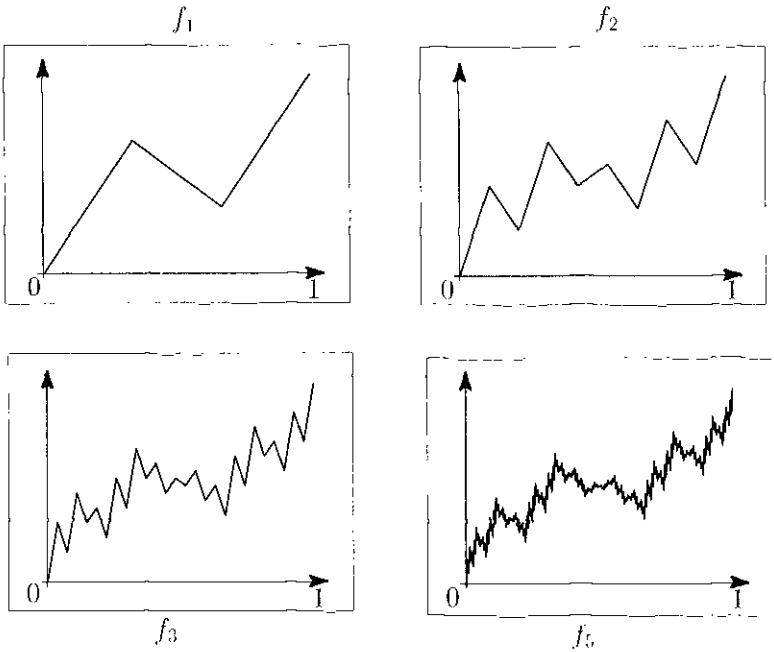


Les graphes des fonctions f_1, f_2, f_3 et f_5 représentés plus loin permettent de visualiser ce processus.

Si on note α la pente de f_n sur $\left[\frac{k}{3^n}, \frac{k+1}{3^n}\right]$, f_{n+1} admet comme pente 2α, (resp. $-\alpha$ et 2α) sur les intervalles $\left[\frac{k}{3^n}, \frac{k}{3^n} + \frac{1}{3^{n+1}}\right]$, (resp. $\left[\frac{k}{3^n} + \frac{1}{3^{n+1}}, \frac{k}{3^n} + \frac{2}{3^{n+1}}\right]$ et $\left[\frac{k}{3^n} + \frac{2}{3^{n+1}}, \frac{k+1}{3^n}\right]$). La fonction f_0 est

affine de pente $1 = 2^0$. Par récurrence, on établit donc que les pentes de f_n sur les intervalles où elle est affine sont en valeurs absolues comprises entre 1 et 2^n. Si $x < y$ sont dans $[0,1]$, il existe une subdivision du segment $[x,y]$ $\{x = a_0 < a_1 < \cdots < a_p = y\}$ pour laquelle f_n est affine sur chaque $]a_i, a_{i+1}[$. Dans ces conditions,

$$|f_n(x) - f_n(y)| \leqslant \sum_{i=0}^{p-1} |f_n(a_i) - f_n(a_{i+1})| \leqslant 2^n \sum_{i=0}^{p-1} |a_i - a_{i-1}|$$
$$\leqslant 2^n (y - x) = 2^n |x - y|.$$



2. Si on reprend le dessin, on constate que sur $\left[\frac{k}{3^n}, \frac{k+1}{3^n}\right]$, la différence $|f_{n+1}(x) - f_n(x)|$ est maximale en $x_1 = \frac{k}{3^n} + \frac{1}{3^{n+1}}$ et $x_2 = \frac{k}{3^n} + \frac{2}{3^{n+1}}$, ce que confirmerait une petite étude analytique laissée au lecteur. D'autre part, sur les intervalles $\left[\frac{k}{3^n}, x_1\right]$ et $\left[x_2, \frac{k+1}{3^n}\right]$, $f_{n+1} - f_n$ est affine de pente égale à $\pm(2^{n+1} - 2^n) = \pm 2^n$. Comme f_n et f_{n+1} coïncident en $\frac{k}{3^n}$ et $\frac{k+1}{3^n}$, on obtient

$$|f_{n+1}(x_1) - f_n(x_1)| = |f_{n+1}(x_2) - f_n(x_2)| \leqslant 2^n \left|x_1 - \frac{k}{3^n}\right| = \frac{2^n}{3^{n+1}}.$$

Ainsi, pour tout $x \in [0,1]$, $|f_{n+1}(x) - f_n(x)| \leqslant \dfrac{2^n}{3^{n+1}}$.

3. Pour tout entier naturel n, $f_n = f_0 + \sum_{k=1}^{n} (f_k - f_{k-1})$. La convergence uniforme de la suite f_n équivaut donc à celle de la série de fonctions $\sum (f_n - f_{n-1})$. Or cette dernière converge normalement sur $[0,1]$ puisque $\|f_{n+1} - f_n\|_\infty \leqslant \dfrac{1}{3}\left(\dfrac{2}{3}\right)^n$. En particulier, la limite f de la suite $(f_n)_{n\in\mathbb{N}}$ est continue puisque les f_n sont continues.

4. Montrons par récurrence sur $k \geqslant n$, que $f_k\left(\dfrac{p}{3^n}\right) = f_n\left(\dfrac{p}{3^n}\right)$. C'est évident pour $k = n$ et par construction pour $k = n+1$. Si $k \geqslant n+1$,

$$f_{k+1}\left(\frac{p}{3^n}\right) = f_{k+1}\left(\frac{3^{k-n}p}{3^k}\right) = f_k\left(\frac{3^{k-n}p}{3^k}\right) = f_k\left(\frac{p}{3^n}\right) = f_n\left(\frac{p}{3^n}\right),$$

par hypothèse de récurrence.

En faisant tendre k vers l'infini, il vient $f\left(\dfrac{p}{3^n}\right) = f_n\left(\dfrac{p}{3^n}\right)$.

5. On considère $a = \dfrac{p}{3^n}$ avec $p \in [\![0, 3^n - 1]\!]$. En reprenant la remarque du départ sur les pentes, la pente de f_n à droite de a est un réel α avec $|\alpha| \geqslant 1$. Pour $k \geqslant n$, lorsqu'on passe de f_k en f_{k+1}, la pente à droite de $a = \dfrac{p3^{k-n}}{3^k}$ est multipliée par 2. Autrement dit, on a

$$\left|f(a) - f\left(a + \frac{1}{3^k}\right)\right| = \left|f_k(a) - f_k\left(a + \frac{1}{3^k}\right)\right| = 2^{k-n}|\alpha|\frac{1}{3^k}.$$

Dans ces conditions, le rapport $\dfrac{\left|f(a) - f\left(a + \dfrac{1}{3^k}\right)\right|}{1/3^k}$ tend vers $+\infty$, et f n'est donc pas dérivable en a.

Pour $a = 1$, on raisonne de même en considérant $f\left(1 - \dfrac{1}{3^k}\right)$.

6. Il reste à vérifier que f n'est pas dérivable en un point a de $[0,1]$ qui n'est pas de forme $\dfrac{p}{3^n}$ avec p et n entiers naturels. Soit $n \geqslant 0$ et $p \in [\![0, 3^n - 1]\!]$ tel que $a \in \left]\dfrac{p}{3^n}, \dfrac{p+1}{3^n}\right[$. La pente de f_n sur $\left]\dfrac{p}{3^n}, \dfrac{p+1}{3^n}\right[$ sera notée α, $|\alpha| \geqslant 1$. On pose $a_0 = \dfrac{p}{3^n}$, $a_1 = \dfrac{3p+1}{3^{n+1}}$, $a_2 = \dfrac{3p+2}{3^{n+1}}$ et $a_3 = \dfrac{p+1}{3^n}$. Pour $i < j$, $\dfrac{f_{n+1}(a_j) - f_n(a_i)}{a_j - a_i}$ vaut 2α, $\pm\alpha$ ou $\dfrac{\alpha}{2}$. On peut choisir a_{i_1} et a_{i_2} dans $[a_0, a[$, puis a_{j_1}, a_{i_2} dans $]a, a_3]$ de telle manière que les pentes

$$\frac{f_{n+1}(a_{j_1}) - f_{n+1}(a_{i_1})}{a_{j_1} - a_{i_1}} \text{ et } \frac{f_{n+1}(a_{j_2}) - f_{n+1}(a_{i_2})}{a_{j_2} - a_{i_2}}$$

soient différentes. On a donc

$$\left| \frac{f(a_{j_1}) - f(a_{i_1})}{a_{j_1} - a_{i_1}} - \frac{f(a_{j_2}) - f(a_{i_2})}{a_{j_2} - a_{i_2}} \right| \geqslant \frac{|\alpha|}{2} \geqslant \frac{1}{2},$$

puisque f et f_{n+1} coïncident en les points a_{i_1}, a_{i_2}, a_{j_1} et a_{j_2}.

Comme $|a_i - a| \leqslant \dfrac{1}{3^n}$ $(0 \leqslant i \leqslant 3)$ et n pouvant être choisi arbitrairement grand, il s'ensuit que la fonction

$$\Phi : (h,k) \longmapsto \frac{f(a+h) - f(a-k)}{h+k} \text{ pour } h > 0 \text{ et } k > 0,$$

ne vérifie pas le critère de Cauchy au voisinage de $(0,0)$. En particulier, cette fonction n'a pas de limite pour $(h,k) \in \mathbb{R}_+^* \times \mathbb{R}_+^*$ tendant vers $(0,0)$.

Raisonnons par l'absurde et supposons f dérivable en a. Soit $\varepsilon > 0$. Il existe $\eta > 0$ tel que pour $|h| \leqslant \eta$,

$$|f(a+h) - f(a) - hf'(a)| \leqslant \varepsilon |h|.$$

Si $(h,k) \in]0,\eta[^2$, on a

$$\begin{aligned} |\Phi(h,k) - f'(a)| &= \frac{|f(a+h) - f(a) - hf'(a) + f(a) - f(a-k) - kf'(a)|}{h+k} \\ &\leqslant \frac{|f(a+h) - f(a) - hf'(a)| + |f(a) - f(a-k) - kf'(a)|}{h+k} \\ &\leqslant \varepsilon \frac{h+k}{h+k} = \varepsilon. \end{aligned}$$

Par conséquent, Φ admet $f'(a)$ comme limite : c'est contradictoire. Donc f n'est pas dérivable en a.

Conclusion. La fonction f est un exemple de fonction continue, nulle part dérivable sur $[0,1]$. ◁

D'autres exemples fameux furent donnés par Van der Waerden et Weierstrass. Pour ce dernier, il s'agit de la série de fonctions $\sum \beta^n \cos \pi \alpha^n x$ *avec* $\alpha\beta > 1 + \dfrac{3\pi}{2}$.

Voici maintenant deux exercices qui concernent l'intégration d'une série de fonctions. Rappelons que la convergence uniforme sur un segment offre une situation simple où il est légitime de passer à la limite sous le signe intégral pour les suites de fonctions et d'intervertir intégration et sommation pour les séries.

2.16. Interversion série-intégrale

Soit $T > 0$ et $f : [0, T] \longrightarrow \mathbb{C}$, continue.

1. Pour $n \in \mathbb{N}$ et $t \in [0, T]$, on pose

$$g_n(t) = \sum_{k=0}^{+\infty} \frac{(-1)^k}{k!} \int_0^T f(u) e^{-kn(t-u)}\, du.$$

Étudier la limite simple de la suite $(g_n)_{n \geqslant 0}$.

2. On suppose que la suite $\left(\int_0^T f(t) e^{nt}\, dt \right)_{n \geqslant 0}$ est bornée. Que peut-on en déduire ?

(École polytechnique)

▷ **Solution.**

1. Avant de calculer la limite de $g_n(t)$ on va transformer l'expression de $g_n(t)$ en échangeant intégration et sommation. Ayant fixé n et t, considérons la suite de fonctions définie, pour $u \in [0, T]$ et $k \in \mathbb{N}$, par

$$v_k(u) = \frac{(-1)^k}{k!} f(u) e^{-kn(t-u)} = \frac{1}{k!} f(u) \left(-e^{-n(t-u)} \right)^k.$$

La série $\sum v_n(u)$ converge et

$$\sum_{k=0}^{+\infty} v_k(u) = f(u) \exp\left(-e^{-n(t-u)} \right).$$

D'autre part, on a pour tout $u \in [0, T]$, $|v_k(u)| \leqslant \frac{\|f\|_\infty}{k!} \left(e^{n(T-t)} \right)^k$ donc $\sum v_k$ converge uniformément sur $[0, T]$.

On en déduit qu'on peut échanger intégration et sommation et écrire

$$\begin{aligned} \int_0^T f(u) \exp\left(-e^{-n(t-u)} \right) du &= \int_0^T \left(\sum_{k=0}^{+\infty} v_k(u) \right) du = \sum_{k=0}^{+\infty} \int_0^T v_k(u)\, du \\ &= g_n(t). \end{aligned}$$

Posons, pour tout $n \in \mathbb{N}$ et $u \in [0, T]$, $h_n(u) = f(u) \exp\left(-e^{-n(t-u)} \right)$. Il est immédiat de voir que la suite (h_n) converge simplement vers la fonction h définie par :

$$\begin{cases} h(u) = f(u) & \text{si } u \in [0, t[\\ h(u) = f(u) e^{-1} & \text{si } u = t \\ h(u) = 0 & \text{si } u \in]t, T]. \end{cases}$$

La fonction h est continue par morceaux donc intégrable sur $[0,\mathrm{T}]$. D'autre part, on a, pour $n \in \mathbb{N}$ et $u \in [0,\mathrm{T}]$, $|h_n(u)| \leqslant \|f\|_\infty$. Du théorème de convergence dominée, on peut déduire que, pour tout $t \in [0,\mathrm{T}]$,

$$\boxed{\lim_{n\to+\infty} g_n(t) = \int_0^{\mathrm{T}} h(u)\,\mathrm{d}u = \int_0^t f(u)\,\mathrm{d}u}.$$

2. Montrons que si la suite $\left(\int_0^{\mathrm{T}} f(t)e^{nt}\,\mathrm{d}t\right)_{n\geqslant 0}$ est bornée, la fonction f est nulle. Ce résultat est facile dans le cas où f est positive. Pour le cas général on va utiliser la première question.

On note M un majorant de $\left(\left|\int_0^{\mathrm{T}} f(t)e^{nt}\,\mathrm{d}t\right|\right)_{n\geqslant 0}$. On a, pour tout couple $(k,u) \in \mathbb{N}^2$ et tout $t \in [0,\mathrm{T}]$,

$$\left|\frac{(-1)^k}{k!}\int_0^{\mathrm{T}} f(u)e^{-kn(t-u)}\,\mathrm{d}u\right| = \frac{1}{k!}e^{-knt}\left|\int_0^{\mathrm{T}} f(u)e^{knu}\,\mathrm{d}u\right| \leqslant \mathrm{M}\frac{e^{-knt}}{k!}.$$

On en déduit que

$$\left|\sum_{k=1}^{+\infty}\frac{(-1)^k}{k!}\int_0^{\mathrm{T}} f(u)e^{-kn(t-u)}\,\mathrm{d}u\right| \leqslant \mathrm{M}\sum_{k=1}^{+\infty}\frac{e^{-knt}}{k!} \leqslant \mathrm{M}(\exp(e^{-nt})-1).$$

Pour $t > 0$, on a $\lim\limits_{n\to\infty}(\exp(e^{-nt})-1) = 0$ et donc

$$\lim_{n\to\infty}\sum_{k=1}^{+\infty}\frac{(-1)^k}{k!}\int_0^{\mathrm{T}} f(u)e^{-kn(t-u)}\,\mathrm{d}u = 0.$$

On en déduit que, pour $t \in]0,\mathrm{T}]$, on a

$$\lim_{n\to+\infty} g_n(t) = \int_0^{\mathrm{T}} f(u)$$

(c'est le terme qui correspond à $k=0$; il est indépendant de n). On a donc, d'après la première question, pour tout $t \in]0,\mathrm{T}]$,

$$\int_0^t f(u)\,\mathrm{d}u = \int_0^{\mathrm{T}} f(u)\,\mathrm{d}u.$$

La fonction $u \longmapsto \int_0^t f(u)\,\mathrm{d}u$ est constante sur $]0,\mathrm{T}]$. Sa dérivée f est nulle sur $]0,\mathrm{T}]$ et par continuité, f est nulle sur $[0,\mathrm{T}]$. ◁

L'exercice suivant étudie ce que l'on peut dire de la convergence d'une série de fonctions $\sum|f_n(t)|$ sous l'hypothèse que la série des intégrales $\sum\int_a^b|f_n|$ converge.

2.17. Sur le théorème d'intégration d'une série de fonctions

Soit $f_n : [a,b] \longrightarrow \mathbb{R}$ $(n \in \mathbb{N})$ une suite de fonctions continues. On suppose que la série $\sum \int_a^b |f_n|$ converge.

1. Peut-on dire que pour tout $t \in [a,b]$, la série $\sum |f_n(t)|$ converge ?

2. Montrer que l'ensemble des $x \in [a,b]$ tels que la série $\sum |f_n(x)|$ converge est dense dans $[a,b]$.

3. Construire une suite $(f_n)_{n\in\mathbb{N}}$ vérifiant les hypothèses de l'énoncé telle que $\left\{x \in [a,b], \sum |f_n(x)| \text{ diverge}\right\}$ est dense.

(École normale supérieure)

Solution.

Remarquons que la présence des valeurs absolues nous permet de supposer simplement que chaque fonction f_n est positive, ce que nous ferons dans la suite.

1. Il est facile de trouver des fonctions d'intégrales petites mais qui ne sont pas uniformément petites sur tout l'intervalle. Prenons par exemple l'intervalle $[0,1]$ et posons pour $n \geqslant 0$, $f_n : t \in [0,1] \longmapsto t^{(n+1)^2-1}$. Dans ces conditions, comme $\int_0^1 f_n = \dfrac{1}{(n+1)^2}$, la série $\sum \int_0^1 f_n$ converge. Mais $\sum f_n(1)$ ne converge pas puisque $f_n(1) = 1$ pour tout n. La réponse est donc négative.

2. Raisonnons par l'absurde et supposons qu'il existe $\alpha < \beta$ dans $[a,b]$ tels que pour tout x dans $[\alpha,\beta]$, $\sum f_n(x)$ diverge. Soit $\ell > 0$. Introduisons pour $n \in \mathbb{N}$

$$\Omega_n = \left\{ x \in [\alpha,\beta], \sum_{k=0}^{n} f_k(x) > \ell \right\}.$$

Comme pour tout $n \in \mathbb{N}$, $f_n \geqslant 0$, $\Omega_n \subset \Omega_{n+1}$. De plus, les f_k étant continues, Ω_n est un ouvert de $[\alpha,\beta]$ (en tant qu'image réciproque d'un ouvert par une fonction continue). Comme $[\alpha,\beta]$ est compact et comme $[\alpha,\beta] = \bigcup_{n\in\mathbb{N}} \Omega_n$ (puisque pour tout $x \in [\alpha,\beta]$, $\lim\limits_{N\to+\infty} \sum\limits_{n=0}^{N} f_n(x) = +\infty$), il existe $N_0 \in \mathbb{N}$ tel que $[\alpha,\beta] = \Omega_{N_0}$. Autrement dit, pour tout $x \in [\alpha,\beta]$, on a

$$\sum_{n=0}^{N_0} f_n(x) > \ell.$$

On a alors,

$$\sum_{n=0}^{+\infty} \int_a^b f_n \geqslant \sum_{n=0}^{N_0} \int_a^b f_n \geqslant \sum_{n=0}^{N_0} \int_\alpha^\beta f_n = \int_\alpha^\beta \sum_{n=0}^{N_0} f_n \geqslant \int_\alpha^\beta \ell \geqslant \ell(\beta-\alpha).$$

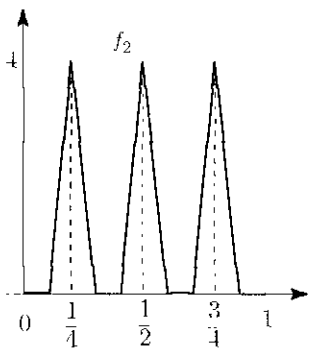
Or, cette inégalité est valable pour tout $\ell > 0$, ce qui est impossible.

Conclusion. $\{x \in [a,b], \sum f_n(x) \text{ converge}\}$ est dense dans $[a,b]$.

3. On va se placer par exemple sur l'intervalle $[0,1]$. On va construire des fonctions affines par morceaux qui présentent de plus en plus de pics. Soit $n \geqslant 0$. On définit ainsi f_n : si $1 \leqslant k \leqslant 2^n - 1$, f_n est affine sur $\left[\frac{k}{2^n} - \frac{1}{8^n}, \frac{k}{2^n}\right]$ et sur $\left[\frac{k}{2^n}, \frac{k}{2^n} + \frac{1}{8^n}\right]$ avec

$$f_n\left(\frac{k}{2^n} - \frac{1}{8^n}\right) = f_n\left(\frac{k}{2^n} + \frac{1}{8^n}\right) = 0 \quad \text{et} \quad f_n\left(\frac{k}{2^n}\right) = 2^n.$$

En dehors de ces intervalles $\left[\frac{k}{2^n} - \frac{1}{8^n}, \frac{k}{2^n} + \frac{1}{8^n}\right]$ (k décrivant les entiers de 1 à $2^n - 1$, f_n est nulle. Voici par exemple le graphe de f_2 :



Pour $n \in \mathbb{N}$, $0 \leqslant \int_0^1 f_n = (2^n - 1) \times \frac{1}{8^n} \times 2^n \leqslant \frac{1}{2^n}$. Par conséquent, $\sum \int_0^1 f_n$ converge d'après le théorème de comparaison des séries à termes positifs.

Maintenant, pour tout réel dyadique $x = \frac{k}{2^{n_0}}$ appartenant à $[0,1]$ la série $\sum f_n(x)$ diverge, puisque $f_n(x) = 2^n$ pour $n \geqslant n_0$. La série diverge donc sur un ensemble dense de $[0,1]$. ◁

Après avoir étudié les propriétés des limites de suites de fonctions, nous allons d'une certaine manière inverser notre point de vue et nous intéresser à l'approximation de fonctions. Le problème est le suivant. On dispose d'un ensemble F *de fonctions à valeurs réelles (ou complexes) définies sur un ensemble* X. *Peut-on décrire les fonctions de* X *dans* $\mathbb{R}$ *(ou* $\mathbb{C}$*) qui sont limites uniformes d'une suite de fonctions de* F *? La plupart du temps on travaille avec des fonctions bornées. Si on munit l'espace vectoriel des fonctions bornées sur* X *de la norme de la convergence uniforme* $\| \ \|_\infty$*, le problème précédent, formulé topologiquement, consiste à trouver l'adhérence de l'ensemble* F.

Un exemple important de cette situation, étudié en cours, est le théorème de Weierstrass qui affirme que l'espace des fonctions polynomiales est dense dans l'espace des fonctions continues de I *dans* $\mathbb{R}$ *lorsque* I *est un segment. L'exercice suivant donne une preuve de ce résultat, due à Bernstein, qui, pour* f *donnée, exhibe une suite explicite de polynômes qui converge vers* f.

2.18. Polynômes de Bernstein

1. Simplifier, pour tout $(x, y) \in \mathbb{R}^2$ et tout $n \in \mathbb{N}^*$,

$$\sum_{k=0}^{n} k\mathrm{C}_n^k x^k y^{n-k} \quad \text{et} \quad \sum_{k=0}^{n} k(k-1)\mathrm{C}_n^k x^k y^{n-k}.$$

2. On pose $r_k(x) = \mathrm{C}_n^k x^k (1-x)^{n-k}$ pour tout $x \in \mathbb{R}$ et tout $k \in \{0, 1, \ldots, n\}$. Démontrer l'égalité

$$\sum_{k=0}^{n} (k - nx)^2 r_k(x) = nx(1-x).$$

3. Soit $f : [0,1] \longrightarrow \mathbb{R}$ continue. Pour tout $n \in \mathbb{N}^*$, on pose

$$\mathrm{B}_n = \sum_{k=0}^{n} \mathrm{C}_n^k f\left(\frac{k}{n}\right) \mathrm{X}^k (1-\mathrm{X})^{n-k} \in \mathbb{R}[\mathrm{X}].$$

Soit $\varepsilon > 0$. Justifier l'existence de $\eta > 0$ tel que pour tout $x \in [0,1]$, $|f(x) - \mathrm{B}_n(x)| \leqslant \varepsilon + \dfrac{\|f\|_\infty}{2n\eta^2}$ (on pourra considérer les indices k tels que $|k - nx| \leqslant n\eta$ et ceux tels que $|k - nx| > n\eta$). En déduire que la suite $(\mathrm{B}_n)_{n\in\mathbb{N}}$ converge uniformément vers f sur $[a, b]$.

4. Soit $\mathrm{F} : [a,b] \longrightarrow \mathbb{R}$ continue. Montrer que F est limite uniforme sur $[a,b]$ d'une suite de polynômes.

(École polytechnique)

Solution.

1. Notons pour $(x, y) \in \mathbb{R}^2$, $\mathrm{P}(x, y) = (x+y)^n = \sum_{k=0}^{n} \mathrm{C}_n^k x^k y^{n-k}$. En dérivant par rapport à x, il vient

$$\frac{\partial \mathrm{P}}{\partial x}(x, y) = n(x+y)^{n-1} = \sum_{k=1}^{n} k\mathrm{C}_n^k x^{k-1} y^{n-k}.$$

En multipliant par x et en rajoutant le terme nul d'indice $k = 0$, on obtient

$$\sum_{k=0}^{n} k\mathrm{C}_n^k x^k y^{n-k} = nx(x+y)^{n-1}.$$

En calculant la dérivée seconde, on a

$$\frac{\partial^2 \mathrm{P}}{\partial x^2}(x,y) = n(n-1)(x+y)^{n-2} = \sum_{k=2}^{n} k(k-1)\mathrm{C}_n^k x^{k-2} y^{n-k}.$$

d'où

$$\sum_{k=0}^{n} k(k-1)\mathrm{C}_n^k x^k y^{n-k} = n(n-1)x^2(x+y)^{n-2}.$$

2. On peut écrire, en faisant $y = 1-x$ dans les égalités de la question précédente,

$$\begin{aligned}\sum_{k=0}^{n}(k-nx)^2 r_k(x) &= \sum_{k=0}^{n}(k^2 - 2knx + n^2x^2)r_k(x)\\ &= \sum_{k=0}^{n}(k(k-1) + k(1-2nx) + n^2x^2)r_k(x)\\ &= n(n-1)x^2.1^{n-2} + (1-2nx)nx.1^{n-1} + n^2x^2.1^n\\ &= nx(1-x).\end{aligned}$$

3. En remarquant que $f(x) = \sum_{k=0}^{n} f(x)r_k(x)$, pour $x \in [0,1]$ et $n \geqslant 1$, on peut écrire

$$\begin{aligned}|f(x) - \mathrm{B}_n(x)| &= \left|\sum_{k=0}^{n}\left(f(x) - f\left(\frac{k}{n}\right)\right) r_k(x)\right|\\ &\leqslant \sum_{k=0}^{n}\left|f(x) - f\left(\frac{k}{n}\right)\right| r_k(x),\end{aligned}$$

les $r_k(x)$ étant positifs. Pour $\dfrac{k}{n}$ proche de x, on peut contrôler la différence $f(x) - f\left(\dfrac{k}{n}\right)$. Pour obtenir une majoration uniforme, nous allons exploiter le fait que f étant continue sur le segment $[0,1]$, elle est uniformément continue. Prenons $\eta > 0$ un module d'uniforme continuité de f pour $\varepsilon > 0$. Dans ces conditions, en posant $\delta_n(x) = |f(x) - \mathrm{B}_n(x)|$, on a

$$\begin{aligned}\delta_n(x) &\leqslant \sum_{|\frac{k}{n}-x|\leqslant\eta}\left|f(x) - f\left(\frac{k}{n}\right)\right| r_k(x) + \sum_{|\frac{k}{n}-x|>\eta}\left|f(x) - f\left(\frac{k}{n}\right)\right| r_k(x)\\ &\leqslant \varepsilon \sum_{|\frac{k}{n}-x|\leqslant\eta} r_k(x) + \sum_{|\frac{k}{n}-x|>\eta}\left|f(x) - f\left(\frac{k}{n}\right)\right| r_k(x)\end{aligned}$$

$$\leqslant \varepsilon + 2\|f\|_\infty \sum_{|\frac{k}{n}-x|>\eta} r_k(x),$$

la fonction étant bornée sur le compact $[0,1]$. Pour contrôler le second terme du majorant, on va utiliser l'identité prouvée à la question précédente. On peut écrire :

$$\begin{aligned}\sum_{|\frac{k}{n}-x|>\eta} r_k(x) &\leqslant \sum_{|\frac{k}{n}-x|>\eta} \frac{(k-nx)^2}{\eta^2 n^2} r_k(x)\\ &= \frac{1}{\eta^2 n^2} \sum_{|\frac{k}{n}-x|>\eta} (k-nx)^2 r_k(x)\\ &\leqslant \frac{1}{\eta^2 n^2} \sum_{k=0}^{n} (k-nx)^2 r_k(x) \leqslant \frac{nx(1-x)}{\eta^2 n^2} \leqslant \frac{1}{4\eta^2 n},\end{aligned}$$

puisque le maximum de la fonction $x \longmapsto x(1-x)$ sur $[0,1]$ est $\frac{1}{4}$. On obtient ainsi la majoration voulue

$$|f(x) - \mathrm{B}_n(x)| \leqslant \varepsilon + \frac{\|f\|_\infty}{2n\eta^2}.$$

En prenant $n \geqslant \dfrac{\|f\|_\infty}{2\varepsilon\eta^2}$, on obtient donc $\|f - \mathrm{B}_n\|_\infty \leqslant 2\varepsilon$. La suite $(\mathrm{B}_n)_{n\in\mathbb{N}}$ converge donc uniformément vers f sur $[0,1]$.

4. Si on pose $f(x) = \mathrm{F}(a + (b-a) - x)$ pour $x \in [0.1]$, on définit une fonction continue et nous pouvons lui associer la suite $(\mathrm{B}_n)_{n\in\mathbb{N}}$ construite précédemment. On vérifie alors que la suite de fonctions $x \longmapsto \mathrm{B}_n\left(\dfrac{x-a}{b-a}\right)$ converge uniformément vers F sur $[a,b]$. ◁

Ce théorème a donné lieu à des études quantitatives dont l'exercice 5.41 du tome 1 d'algèbre peuvent donner une petite idée. Le lecteur trouvera la version pour les polynômes trigonométriques dans le chapitre sur les séries de Fourier (cf. exercices 4.26 et 4.27).

Lorsqu'on se place sur un intervalle ouvert borné $]a,b[$, le lecteur pourra montrer que les limites uniformes de fonctions polynomiales sont exactement les fonctions uniformément continues sur $]a,b[$, ou si l'on préfère, les fonctions continues qui admettent une limite finie en a et b. Enfin lorsqu'on se place sur un intervalle I *non borné de* $\mathbb{R}$, *on montre qu'une limite uniforme de fonctions polynomiales est encore polynomiale. En effet, si* (P_n) *converge uniformément vers* f *sur* I, *on peut trouver un rang* N *tel que* $\|\mathrm{P}_n - f\|_\infty \leqslant 1$ *pour* $n \geqslant \mathrm{N}$. *Mais, si* $n \geqslant \mathrm{N}$, *on a alors* $\|\mathrm{P}_n - \mathrm{P}_{n+1}\|_\infty \leqslant 2$ *de sorte que le polynôme* $\mathrm{P}_n - \mathrm{P}_{n+1}$ *est borné sur* I *et par suite constant. Ainsi, pour* $n \geqslant \mathrm{N}$, $\mathrm{P}_n = \mathrm{P}_\mathrm{N} + \mathrm{P}_n(\alpha) - \mathrm{P}_\mathrm{N}(\alpha)$ *(où* $\alpha \in \mathrm{I}$ *est fixé) et la suite* (P_n) *converge vers* $f = \mathrm{P}_\mathrm{N} - \mathrm{P}_\mathrm{N}(\alpha) + f(\alpha)$

qui est donc polynomiale. Notons que cette question est souvent proposée aux oraux.

L'exercice suivant étudie les limites uniformes de polynômes strictement croissants.

2.19. Limites uniformes de polynômes strictement croissants

Quelles sont les fonctions continues de $[a, b]$ dans $\mathbb{R}$ qui sont limite uniforme sur $[a, b]$ d'une suite de polynômes strictement croissants ?
(**École normale supérieure**)

▷ **Solution.**

Il est clair que la limite simple d'une suite de polynômes strictement croissants est une fonction croissante. Montrons réciproquement que toute fonction continue croissante sur $[a, b]$ est limite uniforme d'une suite de polynômes strictement croissants. Il suffit bien entendu de le faire pour le segment $[0, 1]$ et dans la suite la norme infinie fait référence à ce segment. Nous proposons deux solutions.

• La première solution repose sur le résultat classique de l'exercice 2.18 : si f est continue sur $[0, 1]$, la suite $(\mathrm{B}_n(f))_{n\geqslant 1}$ des polynômes de Bernstein définis par

$$\mathrm{B}_n(f) = \sum_{k=0}^{n} \mathrm{C}_n^k f\left(\frac{k}{n}\right) \mathrm{X}^k (1-\mathrm{X})^{n-k}$$

converge vers f uniformément sur $[0, 1]$.

Considérons une fonction f continue et croissante sur $[0, 1]$ et montrons qu'alors le polynôme $\mathrm{B}_n(f)$ est, pour tout $n \in \mathbb{N}^*$, croissant sur $[0, 1]$. Pour tout $x \in [0, 1[$, $\mathrm{B}_n(f)'(x)$ vaut

$$\sum_{k=1}^{n} \mathrm{C}_n^k f\left(\frac{k}{n}\right) k x^{k-1}(1-x)^{n-k} - \sum_{k=0}^{n-1} \mathrm{C}_n^k f\left(\frac{k}{n}\right)(n-k) x^k (1-x)^{n-k-1}.$$

À l'aide des formules $k\,\mathrm{C}_n^k = n\,\mathrm{C}_{n-1}^{k-1}$ et $(n-k)\mathrm{C}_n^k = n\,\mathrm{C}_{n-1}^k$, on obtient

$$n \sum_{k=1}^{n} \mathrm{C}_{n-1}^{k-1} f\left(\frac{k}{n}\right) x^{k-1}(1-x)^{n-k} - n \sum_{k=0}^{n-1} \mathrm{C}_{n-1}^k f\left(\frac{k}{n}\right) x^k (1-x)^{n-k-1}$$

ce qui donne, après un décalage d'indice dans la première somme,

$$\mathrm{B}_n(f)'(x) = n \sum_{k=0}^{n-1} \mathrm{C}_{n-1}^k \left(f\left(\frac{k+1}{n}\right) - f\left(\frac{k}{n}\right)\right) x^k (1-x)^{n-k-1}.$$

Comme f est croissante, on a $\mathrm{B}'_n(f)(x) \geqslant 0$ pour tout $x \in [0,1]$ et la fonction $\mathrm{B}_n(f)$ est croissante sur $[0,1]$. On n'est toutefois pas sûr qu'elle soit strictement croissante. Posons $\mathrm{P}_n(f) = \dfrac{\mathrm{X}}{n} + \mathrm{B}_n(f)$. On a alors, pour tout $x \in [0,1]$, $\mathrm{P}'_n(f)(x) = \dfrac{1}{n} + \mathrm{B}'_n(f)(x) > 0$. La fonction polynomiale $\mathrm{P}_n(f)$ est strictement croissante sur $[0,1]$ et comme on a l'inégalité $\|f - \mathrm{P}_n(f)\|_\infty \leqslant \|f - \mathrm{B}_n(f)\|_\infty + \dfrac{1}{n}$, la suite $(\mathrm{P}_n)_{n\in\mathbb{N}^*}$ converge encore uniformément vers f sur $[0,1]$.

• Voici une seconde solution, plus susceptible d'être découverte à l'oral par un candidat ne connaissant pas les polynômes de Bernstein. On se donne toujours f continue et croissante sur $[0,1]$. Supposons d'abord que f est de classe $\mathcal{C}^1$. On a donc $f' \geqslant 0$. Soit $\varepsilon > 0$. On commence par choisir un polynôme Q strictement positif sur $[0,1]$ tel que $\|f' - \mathrm{Q}\|_\infty \leqslant \dfrac{\varepsilon}{2}$. Cela est possible grâce au théorème de Weierstrass : on choisit un polynôme R tel que $\|f' - \mathrm{R}\|_\infty < \dfrac{\varepsilon}{2}$ et on pose $\mathrm{Q} = \mathrm{R} + \dfrac{\varepsilon}{2}$. Soit alors P le polynôme défini par $\mathrm{P}(x) = f(0) + \displaystyle\int_0^x \mathrm{Q}(t)\mathrm{d}t$ pour tout $x \in [0,1]$. Le polynôme P est strictement croissant sur $[0,1]$ puisque $\mathrm{P}' = \mathrm{Q}$ et pour tout $x \in [0,1]$ on a

$$|\mathrm{P}(x) - f(x)| = \left| \int_0^x \mathrm{Q}(t)\mathrm{d}t - \int_0^x f'(t)\mathrm{d}t \right| \leqslant \int_0^x |\mathrm{Q}(t) - f'(t)|\,\mathrm{d}t \leqslant \varepsilon$$

et donc $\|\mathrm{P} - f\|_\infty \leqslant \varepsilon$. On peut donc trouver, pour tout $\varepsilon > 0$, un polynôme strictement croissant qui approche uniformément f sur $[0,1]$ à ε près ce qui prouve le résultat.

Pour passer au cas général, il suffit de prouver qu'une fonction continue croissante peut être approchée uniformément par des fonctions croissantes de classe $\mathcal{C}^1$. On utilise un procédé de régularisation classique. Prolongeons f sur $[1,2]$ par la valeur $f(1)$ et posons pour $n \geqslant 1$ et $x \in [0,1]$, $f_n(x) = n\displaystyle\int_x^{x+\frac{1}{n}} f(t)\mathrm{d}t$. Il est clair que f_n est de classe $\mathcal{C}^1$ pour tout n et croissante puisque $f'_n(x) = n\left(f\left(x+\dfrac{1}{n}\right) - f(x)\right) \geqslant 0$. Montrons que la suite (f_n) converge uniformément vers f sur $[0,1]$. Soit $\varepsilon > 0$ et η un module d'uniforme continuité de f associé à ε (le théorème de Heine garantit que f est uniformément continue sur le segment $[0,2]$). Pour $n \geqslant \dfrac{1}{\eta}$ et $x \in [0,1]$ on a

$$\begin{aligned} |f_n(x) - f(x)| &= \left| n\int_x^{x+\frac{1}{n}} (f(t) - f(x))\mathrm{d}t \right| \\ &\leqslant n\int_x^{x+\frac{1}{n}} |f(t) - f(x)|\mathrm{d}t \leqslant n\int_x^{x+\frac{1}{n}} \varepsilon = \varepsilon. \end{aligned}$$

Donc pour n assez grand on a $\|f_n - f\|_\infty \leqslant \varepsilon$ et cela prouve le résultat annoncé. $\lhd$

L'exercice suivant concerne l'étude simultanée de l'interpolation et de l'approximation uniforme.

2.20. Un théorème de Walsh

Soit $x_1, \ldots, x_n$ des points deux à deux distincts d'un segment $[a,b]$ et $f : [a,b] \to \mathbb{R}$ continue. Montrer qu'il existe une suite de polynômes coïncidant avec f en chaque point x_k, qui converge uniformément vers f sur $[a,b]$.

(École normale supérieure)

$\rhd$ **Solution.**

Notons $L_1, \ldots, L_n$ les polynômes interpolateurs de Lagrange pour le système de points $(x_1, \ldots, x_n)$. Rappelons que $L_k(X) = \dfrac{\prod_{i \neq k} (X - x_i)}{\prod_{i \neq k} (x_k - x_i)}$ est de degré $n-1$ et vérifie $L_k(x_i) = \delta_{ik}$. On note $M = \max_{1 \leqslant k \leqslant n} \|L_k\|_\infty$. Pour prouver le résultat il suffit de montrer que pour tout $\varepsilon > 0$ on peut trouver un polynôme Q tel que $Q(x_k) = f(x_k)$ pour tout k et vérifiant $\|f - Q\|_\infty \leqslant \varepsilon$.

Soit $\varepsilon > 0$. On commence par choisir un polynôme P uniformément proche de f à ε près. L'existence d'un tel polynôme découle du théorème de Weierstrass. Soit L l'interpolateur de Lagrange de la fonction $f - P$ en les x_k :

$$L(X) = \sum_{k=1}^{n} (f(x_k) - P(x_k)) L_k(X).$$

On observe que $\|L\|_\infty \leqslant nM\varepsilon$. Posons alors $Q = P + L$. On a bien $Q(x_k) = f(x_k)$ pour tout k et de plus

$$\|f - Q\|_\infty \leqslant \|f - P\|_\infty + \|L\|_\infty \leqslant (1 + nM)\varepsilon.$$

Le coefficient $1 + nM$ est une constante et le résultat est donc prouvé. $\lhd$

On obtient de nouveaux problèmes très intéressants si on remplace les polynômes à coefficients réels par les polynômes à coefficients entiers. L'exercice suivant montre ainsi que si I est un segment inclus dans $]0,1[$, l'ensemble $\mathbb{Z}[X]$ reste dense dans l'espace des fonctions continues sur I.

2.21. Théorème de Chudnovsky

Soit I un segment contenu dans $]0,1[$.

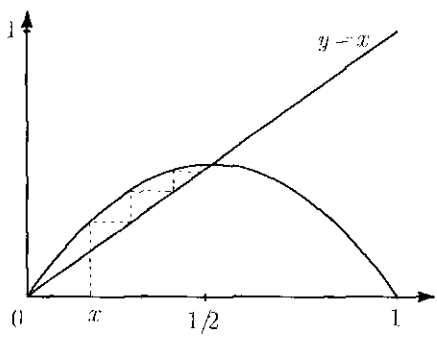
1. Soit φ l'application $x \longmapsto 2x(1-x)$. Étudier la convergence sur I de la suite de fonctions (φ_n) (où $\varphi_n = \varphi \circ \cdots \circ \varphi$).

2. Soit $f : \mathrm{I} \to \mathbb{R}$ une fonction continue. Montrer qu'il existe une suite de fonctions polynômes à coefficients dans $\mathbb{Z}$ convergeant uniformément vers f sur I.

(École polytechnique)

▷ **Solution.**

1. Une étude rapide de la fonction φ montre que $\varphi(]0,1[) = \left]0, \frac{1}{2}\right]$, que φ est croissante sur $\left[0, \frac{1}{2}\right]$ et que, pour tout $x \in \left]0, \frac{1}{2}\right]$, on a $\varphi(x) \geqslant x$. Enfin les points fixes de φ sont 0 et $\frac{1}{2}$. Voici le graphe de φ :



Soit $x \in]0,1]$. Ce qui précède montre que la suite $(\varphi_n(x))_{n\in\mathbb{N}}$ prend, pour $n \geqslant 1$, ses valeurs dans $\left]0, \frac{1}{2}\right]$ et est croissante. Elle converge donc. Sa limite ne peut être 0 puisqu'elle croît ; c'est donc $\frac{1}{2}$. Ceci montre la convergence simple sur $]0,1[$ de la suite $(\varphi_n)_{n\in\mathbb{N}}$ vers la fonction constante $x \longmapsto \frac{1}{2}$.

Montrons que la convergence est uniforme sur I, où I est un segment contenu dans $]0,1[$. L'image $\varphi(\mathrm{I})$ est un segment inclus dans $\left]0, \frac{1}{2}\right]$. Soit a la borne inférieure de ce segment. Sachant que φ est croissante sur $\left]0, \frac{1}{2}\right]$, on a, pour tout $x \in \mathrm{I}$ et $n \in \mathbb{N}^*$, $\varphi_{n-1}(a) \leqslant \varphi_n(x) \leqslant \frac{1}{2}$, puisque $\varphi(x) \in \left[a, \frac{1}{2}\right]$. La suite $(\varphi_n(a))$ convergeant vers $\frac{1}{2}$, cela montre que (φ_n) converge uniformément sur I.

2. Notons $\mathcal{C}$ l'algèbre des fonctions continues sur I, munie de la norme de la convergence uniforme. Dans la suite on identifie tout polynôme

formel de $\mathbb{R}[X]$ avec la fonction de $\mathcal{C}$ qui lui est canoniquement associée. Avec cette identification, la question posée consiste à trouver l'adhérence de $\mathbb{Z}[X]$ dans $\mathcal{C}$. Comme φ_n est dans $\mathbb{Z}[X]$ pour tout n, la première question montre que la fonction constante $\frac{1}{2}$ est dans cette adhérence.

Notons que cette adhérence est stable pour la somme et pour le produit. En effet, si (P_n) et (Q_n) sont des suites de $\mathbb{Z}[X]$ qui convergent uniformément vers f et g sur I, alors $(P_n + Q_n)$ converge uniformément vers $f+g$ car $\|(f_n+g_n)-(f+g)\|_\infty \leqslant \|f_n - f\|_\infty + \|g_n - g\|_\infty$. De même en écrivant

$$\|f_n g_n - fg\|_\infty \leqslant \|f_n\|_\infty \|g_n - g\|_\infty + \|g\|_\infty \|f_n - f\|_\infty,$$

on voit que $(P_n Q_n)$ converge uniformément vers fg car la suite $\|f_n\|_\infty$ est bornée (puisqu'elle converge vers $\|f\|_\infty$).

Il en résulte que la fonction constante $\frac{1}{2^m}$ est dans $\overline{\mathbb{Z}[X]}$ pour tout $m \in \mathbb{N}$. Comme on a aussi toutes les fonction constantes à valeurs dans $\mathbb{Z}$, on a dans $\overline{\mathbb{Z}[X]}$ toutes les fonctions constantes $x \longmapsto \frac{p}{2^m}$ avec $p \in \mathbb{Z}$ et $m \in \mathbb{N}$. Comme l'ensemble $D = \left\{ \frac{p}{2^m}, (p,m) \in \mathbb{Z} \times \mathbb{N} \right\}$ des rationnels dyadiques est dense dans $\mathbb{R}$, $\overline{\mathbb{Z}[X]}$ contient finalement toutes les fonctions constantes. Il en résulte que $\mathbb{R}[X] \subset \overline{\mathbb{Z}[X]}$ et donc que $\overline{\mathbb{R}[X]} \subset \overline{\mathbb{Z}[X]}$. Or par théorème de Weierstrass $\overline{\mathbb{R}[X]} = \mathcal{C}$. On a donc démontré que $\mathbb{Z}[X]$ est dense dans $\mathcal{C}$. ◁

Si on se place sur un segment I *qui contient des entiers, les fonctions dans l'adhérence de* $\mathbb{Z}[X]$ *doivent bien entendu prendre des valeurs entières en chaque point entier de* I. *Mais cela ne suffit pas. Le lecteur pourra ainsi montrer que si* $I = [a,b]$ *avec* $b - a > 4$ *alors* $\mathbb{Z}[X]$ *est fermé dans* $\mathcal{C}^0(I,\mathbb{R})$ *(utiliser le résultat de l'exercice 5.36 du tome 1 d'algèbre). La description générale de cette adhérence est assez délicate et dépend de la longueur de* I. *Le lecteur intéressé par cette question se reportera à l'article de Hervé Pépin et Nicolas Tosel publié dans la RMS de janvier 2004.*

Dans sa forme trigonométrique le théorème de Weierstrass affirme qu'une fonction continue et 2π*-périodique est limite uniforme sur* $\mathbb{R}$ *d'une suite de polynômes trigonométriques c'est-à-dire de fonctions de la forme* $\sum_{k=0}^{n} a_k \cos kx + b_k \sin kx$. *On en déduit que toute fonction continue* T*-périodique (avec* $T > 0$*) est limite uniforme sur* $\mathbb{R}$ *d'une suite de polynômes trigonométriques de la forme* $\sum_{k=0}^{n} a_k \cos \frac{2k\pi}{T} x + b_k \sin \frac{2k\pi}{T} x$. *Il est alors assez naturel de s'intéresser à l'adhérence, dans l'espace des fonctions continues bornées sur* $\mathbb{R}$ *muni de la norme de la convergence*

uniforme, du sous-espace engendré par toutes les fonctions $x \mapsto \cos \omega x$, $x \mapsto \sin \omega x$ lorsque ω parcourt $\mathbb{R}$. Cette adhérence contient toutes les fonctions continues périodiques d'après ce qu'on vient de dire. Mais ce n'est pas tout ! En effet, cette adhérence est un sous-espace vectoriel (et même une algèbre) et l'ensemble des fonctions périodiques n'est pas stable pour la somme : par exemple la fonction $x \mapsto \cos x + \cos \sqrt{2}x$ n'est pas périodique (il est facile de voir qu'elle ne prend la valeur 2 qu'en $x = 0$). Par définition cette adhérence est l'algèbre des fonctions presque périodiques.

2.22. Algèbre des fonctions presque périodiques

Soit E l'espace vectoriel réel engendré par les fonctions $t \mapsto \sin \omega t$ et $t \mapsto \cos \omega t$ lorsque ω décrit $\mathbb{R}$, et F l'ensemble des limites uniformes des suites à valeurs dans E.

1. Montrer que F est une sous-algèbre de l'algèbre des fonctions continues bornées de $\mathbb{R}$ dans $\mathbb{R}$.

2. Soit $f \in \mathcal{C}^0(\mathbb{R}, \mathbb{R})$ et $g \in \mathrm{F}$. Montrer que $f \circ g \in \mathrm{F}$.

3. Pour $\omega \in \mathbb{R}$ et $g \in \mathrm{F}$, on note

$$\widetilde{g}(\omega) = \lim_{\mathrm{T} \to \infty} \frac{1}{2\mathrm{T}} \int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega x} g(x) \mathrm{d}x.$$

Montrer que $\widetilde{g}(\omega)$ est bien défini, et qu'il est nul sauf pour un ensemble dénombrable de valeurs de ω.

(École normale supérieure)

▷ **Solution.**

1. Notons $\mathcal{C}$ l'algèbre des fonctions continues bornées de $\mathbb{R}$ dans $\mathbb{R}$ munie de la norme de la convergence uniforme. L'ensemble E est par définition un sous-espace vectoriel de $\mathcal{C}$ (tout élément de E est une fonction bornée sur $\mathbb{R}$ car les fonctions $t \longmapsto \sin \omega t$ et $t \longmapsto \cos \omega t$ le sont et une combinaison linéaire de fonctions bornées est bornée). C'est même une sous-algèbre. Pour le montrer, il faut vérifier que le produit de deux éléments de E est dans E ; par linéarité, il suffit de le vérifier pour les fonctions $t \longmapsto \sin \omega t$ et $t \longmapsto \cos \omega t$, où ω est un réel quelconque. Cela résulte de formules de trigonométrie bien connues : pour $(\omega, \omega', t) \in \mathbb{R}^3$,

$$\begin{cases} \cos \omega t \cos \omega' t = \dfrac{1}{2} \cos(\omega + \omega')t + \dfrac{1}{2} \cos(\omega - \omega')t \\ \sin \omega t \cos \omega' t = \dfrac{1}{2} \sin(\omega + \omega')t + \dfrac{1}{2} \sin(\omega - \omega')t \\ \sin \omega t \sin \omega' t = \dfrac{1}{2} \cos(\omega - \omega')t - \dfrac{1}{2} \cos(\omega + \omega')t. \end{cases}$$

Par définition, F est l'adhérence de E et il s'agit donc d'une sous-

algèbre de $\mathcal{C}$. En effet, il suffit de vérifier que F est stable par combinaison linéaire et pour le produit. Soient f, g dans F et $(f_n)_{n\in\mathbb{N}}$, $(g_n)_{n\in\mathbb{N}}$ deux suites à valeurs dans F qui convergent uniformément vers f et g. Alors pour λ et μ dans $\mathbb{R}$, $\lambda f_n + \mu g_n$ converge uniformément sur $\mathbb{R}$ vers $\lambda f + \mu g$ qui est donc dans F et $(f_n g_n)$ converge uniformément sur $\mathbb{R}$ vers fg. Pour voir ce dernier point il suffit d'écrire que

$$\|f_n g_n - fg\|_\infty \leqslant \|f_n\|_\infty \|g_n - g\|_\infty + \|g\|_\infty \|f_n - f\|_\infty$$

et de constater que le majorant tend vers 0 (la suite $\|f_n\|_\infty$ converge vers $\|f\|_\infty$).

2. Soit $g \in \mathrm{F}$ et f une fonction continue sur $\mathbb{R}$ (pas forcément bornée). Si f est un polynôme, il est clair que $f \circ g \in \mathrm{F}$ car F est une sous-algèbre de $\mathcal{C}$. On va passer au cas général en utilisant le théorème de Weierstrass. La fonction g est bornée : posons $\mathrm{M} = \|g\|_\infty$. La restriction de f au segment $[-\mathrm{M}, \mathrm{M}]$ est, d'après le théorème de Weierstrass, limite uniforme sur ce segment d'une suite de fonctions polynômes $(\mathrm{P}_n)_{n\geqslant 0}$. Comme

$$\|\mathrm{P}_n \circ g - f \circ g\|_\infty \leqslant \sup_{t\in[-\mathrm{M},\mathrm{M}]} |\mathrm{P}_n(t) - f(t)|,$$

la suite $(\mathrm{P}_n \circ g)_{n\geqslant 0}$ de F converge uniformément vers $f \circ g$ sur $\mathbb{R}$. Comme F est fermé, on a donc $f \circ g \in \mathrm{F}$.

3. Soit $\omega \in \mathbb{R}$. Montrons, pour commencer que $\widetilde{g}(\omega)$ est bien défini si g est dans E. Étant données la linéarité de l'intégrale et de la limite, il suffit de le démontrer pour les fonctions $c_\alpha : t \mapsto \cos \alpha t$ et $s_\alpha : t \mapsto \sin \alpha t$. On a

$$\frac{1}{2\mathrm{T}} \int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega t} e^{i\alpha t} \mathrm{d}t = \begin{cases} 1 \text{ si } \alpha = \omega \\ \dfrac{e^{i(\alpha-\omega)\mathrm{T}} - e^{-i(\alpha-\omega)\mathrm{T}}}{i(\alpha-\omega)2\mathrm{T}} \text{ si } \alpha \neq \omega. \end{cases}$$

On en déduit que, si $\alpha \neq \omega$, $\left| \dfrac{1}{2\mathrm{T}} \displaystyle\int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega t} e^{i\alpha t} \mathrm{d}t \right| \leqslant \dfrac{1}{|\mathrm{T}(\alpha - \omega)|}$. On a donc $\displaystyle\lim_{\mathrm{T}\to+\infty} \frac{1}{2\mathrm{T}} \int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega t} e^{i\alpha t} \mathrm{d}t = 0$ si $\alpha \neq \omega$ et 1 si $\alpha = \omega$.

Sachant que $c_\alpha(t) = \dfrac{e^{i\alpha t} + e^{-i\alpha t}}{2}$ et $s_\alpha(t) = \dfrac{e^{i\alpha t} - e^{-i\alpha t}}{2i}$, on en déduit l'existence de $\widetilde{c_\alpha}(\omega)$ et $\widetilde{s_\alpha}(\omega)$:

$$\widetilde{c_\alpha}(\omega) = \begin{cases} 0 & \text{si } |\alpha| \neq |\omega| \\ \dfrac{1}{2} & \text{si } |\alpha| = |\omega| \neq 0 \\ 1 & \text{si } \alpha = \omega = 0 \end{cases} \quad \text{et} \quad \widetilde{s_\alpha}(\omega) = \begin{cases} 0 & \text{si } |\alpha| \neq |\omega| \\ \dfrac{1}{2i} & \text{si } \alpha = \omega \neq 0 \\ \dfrac{-1}{2i} & \text{si } \alpha = -\omega \neq 0 \\ 0 & \text{si } \alpha = \omega = 0. \end{cases}$$

Établissons maintenant l'existence de $\widetilde{g}(\omega)$ pour g dans F et ω dans $\mathbb{R}$. Soit $(g_n)_{n\in\mathbb{N}}$ une suite à valeurs dans E convergeant uniformément vers g. Notons φ_n la fonction $\mathrm{T} \mapsto \dfrac{1}{2\mathrm{T}}\displaystyle\int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega x} g_n(x)\mathrm{d}x$ pour tout n et soit $\varphi : \mathrm{T} \mapsto \dfrac{1}{2\mathrm{T}}\displaystyle\int_{-\mathrm{T}}^{\mathrm{T}} e^{-i\omega x} g(x)\mathrm{d}x$. Ces fonctions sont définies sur $\mathbb{R}_+^*$. D'après ce qui précède chaque fonction φ_n admet une limite lorsque $\mathrm{T} \to +\infty$. Or, pour tout $\mathrm{T} > 0$ le théorème de majoration montre que

$$|\varphi_n(\mathrm{T}) - \varphi(\mathrm{T})| \leqslant \frac{1}{2\mathrm{T}} \int_{-\mathrm{T}}^{\mathrm{T}} |g_n(x) - g(x)|\mathrm{d}x \leqslant \|g_n - g\|_\infty.$$

On en déduit que la suite $(\varphi_n)_{n\geqslant 0}$ converge uniformément vers φ sur $\mathbb{R}_+^*$. Le théorème d'interversion des limites permet alors de dire que φ admet une limite lorsque T tend vers $+\infty$. Cela prouve l'existence de $\widetilde{g}(\omega)$ et on peut même noter que $\widetilde{g}(\omega) = \lim\limits_{n\to+\infty} \widetilde{g_n}(\omega)$.

Il reste à démontrer que $\widetilde{g}(\omega)$ est nul sauf pour un nombre dénombrable de valeurs de ω. Pour tout n de $\mathbb{N}$, g_n est combinaison d'un nombre fini de fonctions c_ω et s_ω. Il existe donc un couple $(k,l) \in \mathbb{N}^2$ et $(\omega_1, \omega_2, \ldots, \omega_k, \alpha_1, \alpha_2, \ldots, \alpha_l) \in \mathbb{R}^{k+l}$ tel que $g_n \in \mathrm{Vect}(c_{\omega_1}, c_{\omega_2}, \ldots, c_{\omega_k}, s_{\alpha_1}, s_{\alpha_2}, \ldots, s_{\alpha_l})$. Soit A_n la réunion des ensembles $\{\omega_1, \omega_2, \ldots, \omega_k\}$ et $\{\alpha_1, \alpha_2, \ldots, \alpha_l\}$ et C_n la réunion de A_n et de $\mathrm{B}_n = \{-\omega, \omega \in \mathrm{A}_n\}$. De l'étude qui a été faite de $\widetilde{c_\alpha}(\omega)$ et $\widetilde{s_\alpha}(\omega)$, on déduit que si $\omega \notin \mathrm{C}_n$, alors $\widetilde{g_n}(\omega) = 0$.

Soit $\mathrm{C} = \bigcup_{n=0}^{+\infty} \mathrm{C}_n$. En tant que réunion dénombrable d'ensembles finis, C est dénombrable et si $\gamma \notin \mathrm{C}$ on a $\widetilde{g_n}(\omega) = 0$ pour tout $n \in \mathbb{N}$. On en déduit que si $\gamma \notin \mathrm{C}$, alors $\widetilde{g}(\omega) = \lim\limits_{n\to+\infty} \widetilde{g_n}(\omega) = 0$. Hors de l'ensemble dénombrable C, $\widetilde{g}$ est nulle. ◁

L'ensemble des valeurs de ω tel que $\widetilde{g}(\omega)$ soit non nul est appelé le spectre de f. L'étude des fonctions presque périodiques a fait l'objet de plusieurs problèmes de concours très intéressants : un sujet posé à l'ENS Lyon en 1992 et, plus récemment, le sujet Ulm-Cachan de 2002. Le lecteur y trouvera de nombreux résultats supplémentaires. Signalons que l'algèbre des fonctions presque périodiques est strictement incluse dans l'algèbre des fonctions continues bornées (une fonction presque périodique est uniformément continue sur $\mathbb{R}$).

L'énoncé suivant généralise et éclaire la preuve du théorème de Weierstrass par les polynômes de Bernstein donnée dans l'exercice 2.18.

2.23. Théorème de Korovkin (1953)

Soit $\mathrm{E} = \mathcal{C}([0,1].\mathbb{R})$, muni de la norme $\|.\|_\infty$. Pour $u \in \mathrm{L}(\mathrm{E})$, on dit que u est un opérateur positif (ce qu'on note $u \geqslant 0$) si, pour tout $f \in \mathrm{E}$,

$$f \geqslant 0 \Longrightarrow u(f) \geqslant 0.$$

1. Soit $u \in \mathrm{L}(\mathrm{E})$. On suppose $u \geqslant 0$. Montrer que u est continu.

2. Soit $f \in \mathrm{E}$. On fixe $\varepsilon > 0$. Établir l'existence de $c > 0$ tel que, pour tout $(x, y) \in [0,1]^2$,

$$|f(x) - f(y)| \leqslant \varepsilon + c(y-x)^2.$$

3. *Théorème de Korovkin.* Pour $k \in \mathbb{N}$, on note e_k l'élément de E défini par $e_k(x) = x^k$. Soit $(u_n)_{n\in\mathbb{N}} \in \mathrm{L}(\mathrm{E})^{\mathbb{N}}$ une suite d'opérateurs positifs. On suppose que pour $k = 0, 1, 2$ la suite $(u_n(e_k))_{n\in\mathbb{N}}$ converge vers e_k dans $(\mathrm{E}, \|.\|_\infty)$. Montrer que, pour tout $f \in \mathrm{E}$, la suite $(u_n(f))_{n\in\mathbb{N}}$ converge uniformément vers f.

(École polytechnique)

▷ **Solution.**

Notons que si $u \in \mathcal{L}(\mathrm{E})$ est positif, on a, pour $(f, g) \in \mathrm{E}^2$,

$$f \leqslant g \Longrightarrow u(f) \leqslant u(g).$$

et en particulier, pour tout $f \in \mathrm{E}$, $|u(f)| \leqslant u(|f|)$.

1. Pour tout $f \in \mathrm{E}$, des inégalités $-\|f\|_\infty e_0 \leqslant f \leqslant \|f\|_\infty e_0$, avec $e_0 : x \longmapsto 1$, on tire $-\|f\|_\infty u(e_0) \leqslant u(f) \leqslant \|f\|_\infty u(e_0)$, et donc

$$\|u(f)\|_\infty \leqslant \|f\|_\infty \|u(e_0)\|_\infty,$$

ce qui montre la continuité de u (l'application linéaire u est lipschitzienne de rapport $\|u(e_0)\|_\infty$).

2. Puisque f est continue sur le compact $[0,1]$, elle y est uniformément continue. Soit η un module d'uniforme continuité de f pour ε.

- Si $|y-x| \leqslant \eta$, alors $|f(y) - f(x)| \leqslant \varepsilon$.
- Si $|y-x| > \eta$, alors $\dfrac{(y-x)^2}{\eta^2} \geqslant 1$ et on a

$$|f(y) - f(x)| \leqslant 2\|f\|_\infty \leqslant \frac{2\|f\|_\infty}{\eta^2}(y-x)^2.$$

On a donc, pour tout $(x, y) \in [0,1]^2$,

$$|f(y) - f(x)| \leqslant \varepsilon + \frac{2\|f\|_\infty}{\eta^2}(y-x)^2,$$

ce qui est l'inégalité voulue avec $c = \dfrac{2\|f\|_\infty}{\eta^2}$.

3. Soit $f \in \mathrm{E}$ et $\varepsilon > 0$. On choisit c comme dans la question précédente. Fixons x dans $[0,1]$. On applique l'opérateur u_n sur l'inégalité

$$|f(y) - f(x)| \leqslant \varepsilon + c(y^2 - 2xy + x^2)$$

où les fonctions dépendent de la variable y, autrement dit sur l'inégalité entre fonctions : $|f - f(x)e_0| \leqslant \varepsilon e_0 + c(e_2 - 2xe_1 + x^2 e_0)$. On en déduit que la quantité $|u_n(f)(y) - f(x)u_n(e_0)(y)|$ est majorée par

$$\varepsilon u_n(e_0)(y) + c[u_n(e_2)(y) - 2xu_n(e_1)(y) + x^2 u_n(e_0)(y)]$$

pour tout $y \in [0,1]$. Par inégalité triangulaire on a

$$|u_n(f)(x) - f(x)| \leqslant |u_n(f)(x) - f(x)u_n(e_0)(x)| + |f(x)||u_n(e_0)(x) - 1|.$$

On applique l'inégalité précédente avec $y = x$ pour majorer le terme $|u_n(f)(x) - f(x)u_n(e_0)(x)|$ par

$$\varepsilon u_n(e_0)(x) + c[u_n(e_2)(x) - 2xu_n(e_1)(x) + x^2 u_n(e_0)(x)].$$

Par hypothèse, cette quantité converge uniformément vers la fonction constante égale à ε lorsque n tend vers l'infini (le terme entre crochet tend uniformément vers 0). De plus, comme f est bornée, le terme $|f(x)||u_n(e_0)(x) - 1|$ converge uniformément vers 0. Il en résulte qu'il existe un rang N tel que pour tout $n \geqslant \mathrm{N}$ et tout $x \in [0,1]$,

$$|u_n(f)(x) - f(x)| \leqslant 2\varepsilon.$$

Conclusion. La suite $(u_n(f))_{n\in\mathbb{N}}$ converge uniformément vers f. Le résultat de cette question constitue le théorème de Korovkin. ◁

On retrouve en particulier la preuve de Bernstein (voir exercice 2.18) du théorème de Weierstrass en prenant pour B_n *les opérateurs définis par* $\mathrm{B}_n(f)(x) = \displaystyle\sum_{k=0}^{n} \mathrm{C}_n^k f\left(\frac{k}{n}\right) x^k(1-x)^{n-k}$. *Ils sont clairement positifs et il est facile de vérifier que* $\mathrm{B}_n(e_k)$ *converge uniformément vers* e_k *pour* $k = 0, 1, 2$.

L'énoncé suivant concerne encore des questions d'approximation uniforme mais cette fois sur l'intervalle $\mathbb{R}_+$ *qui n'est pas compact. Le fait qu'un polynôme non constant n'est plus borné sur* $\mathbb{R}_+$ *complique grandement les choses.*

2.24. Approximation de Laguerre

Soit $\lambda \geqslant 1$ et $\varepsilon > 0$. Montrer qu'il existe $P \in \mathbb{R}[X]$ tel que $|e^{-\lambda x} - e^{-x}P(x)| \leqslant \varepsilon$ pour tout $x \in \mathbb{R}_+$. On pourra commencer par le cas $1 \leqslant \lambda \leqslant 2$.

(École normale supérieure)

▷ **Solution.**

• Comme le préconise l'énoncé, supposons d'abord $1 \leqslant \lambda \leqslant 2$ et même $\lambda > 1$ puisque le problème est trivial pour $\lambda = 1$. Si on met e^{-x} en facteur, on est conduit à chercher un polynôme P qui approche bien la fonction $e^{-(\lambda-1)x}$. On va prendre son polynôme de Taylor en 0. L'inégalité de Taylor-Lagrange en 0 à l'ordre n, nous permet d'affirmer que pour tout $x \geqslant 0$,

$$\left| e^{-(\lambda-1)x} - \sum_{k=0}^{n} \frac{(-1)^k(\lambda-1)^k}{k!}x^k \right| \leqslant \frac{(\lambda-1)^{n+1}x^{n+1}}{(n+1)!} \leqslant \frac{x^{n+1}}{(n+1)!}$$

car $0 \leqslant \lambda - 1 \leqslant 1$. Si on note $Q_n(x)$ le polynôme de Taylor d'ordre n ci-dessus, on a donc pour tout $x \geqslant 0$,

$$|e^{-\lambda x} - e^{-x}Q_n(x)| \leqslant \frac{e^{-x}x^{n+1}}{(n+1)!}.$$

Une étude de fonction très simple, montre que la fonction majorante atteint son maximum en $n+1$. On a donc

$$\|e^{-\lambda x} - e^{-x}Q_n(x)\|_\infty \leqslant \frac{e^{-n-1}(n+1)^{n+1}}{(n+1)!} \underset{n\to\infty}{\sim} \frac{1}{\sqrt{2\pi n}}$$

d'après l'équivalent de Stirling, ce qui tend vers 0 et donne le résultat demandé. On voit que cette solution ne convient pas pour tout λ car lorsque $\lambda > 2$, on a le terme $(\lambda-1)^{n+1}$ en plus et il rend le majorant divergent.

• On va alors montrer par récurrence sur l'entier $p \geqslant 1$, que le résultat reste vrai pour $\lambda \in]p, p+1]$. Le point précédent concerne le cas $p = 1$. Supposons le résultat prouvé pour $\lambda \in]p-1, p]$ et soit $\lambda \in]p, p+1]$. On dispose donc, par hypothèse de récurrence, d'une suite de polynômes (R_n) telle que $R_n(x)e^{-x}$ converge uniformément sur $\mathbb{R}^+$ vers $e^{-(\lambda-1)x}$. On écrit plutôt $e^{-(\lambda-1)x} = R_n(x)e^{-x} + \varepsilon_n(x)$ où (ε_n) est une suite de fonctions qui converge uniformément vers 0 sur $\mathbb{R}^+$. On multiplie cela par la fonction e^{-x} qui est bornée sur $\mathbb{R}^+$: on a

$$e^{-\lambda x} = R_n(x)e^{-2x} + e^{-x}\varepsilon_n(x) \tag{1}$$

et le reste tend uniformément vers 0. Par le point précédent (appliqué avec $\lambda = 2$), il existe une suite de polynômes (Q_n) telle que $e^{-x}Q_n(x)$ converge uniformément vers e^{-2x}. Ici aussi, on écrit

$$e^{-2x} = e^{-x}Q_n(x) + \tilde{\varepsilon}_n(x) \tag{2}$$

avec $(\tilde{\varepsilon}_n)$ qui tend uniformément vers 0.

Une première idée est de remplacer e^{-2x} donné par (2) dans l'égalité (1). Il vient

$$e^{-\lambda x} = e^{-x}Q_n(x)R_n(x) + R_n(x)\tilde{\varepsilon}_n(x) + e^{-x}\varepsilon_n(x).$$

Mais on ne peut pas conclure car, R_n n'étant pas borné sur $\mathbb{R}^+$, rien ne permet de dire que le terme $R_n\tilde{\varepsilon}_n$ converge uniformément vers 0.

On va procéder autrement pour fabriquer $e^{-\lambda x}$. L'idée est d'écrire (2) en $\dfrac{\lambda x}{2}$ et (1) en $\dfrac{x}{2}$ pour remplacer le terme $e^{-\lambda x/2}$. On a donc

$$e^{-\lambda x} = e^{-\lambda x/2}Q_n(\lambda x/2) + \tilde{\varepsilon}_n(\lambda x/2),$$

et le reste tend uniformément vers 0. Maintenant on remplace le terme $e^{-\lambda x/2}$ par $R_n(x/2)e^{-x} + e^{-x/2}\varepsilon_n(x/2)$. Il vient

$$e^{-\lambda x} = R_n(x/2)Q_n(\lambda x/2)e^{-x} + Q_n(\lambda x/2)e^{-x/2}\varepsilon_n(x/2) + \tilde{\varepsilon}_n(\lambda x/2).$$

Le terme $R_n(x/2)Q_n(\lambda x/2)$ est un polynôme. Le terme $\tilde{\varepsilon}_n(\lambda x/2)$ converge uniformément vers 0 sur $\mathbb{R}^+$. Il reste à regarder le terme $Q_n(\lambda x/2)e^{-x/2}\varepsilon_n(x/2)$. On voit la différence par rapport à la première tentative : il reste ici une exponentielle et la fonction $Q_n(\lambda x/2)e^{-x/2}$ est bornée. Si on note M_n sa borne supérieure, il suffit de faire en sorte que la suite $M_n\|\varepsilon_n\|_\infty$ tende vers 0. Mais, (Q_n) étant donnée, rien n'interdit de choisir la suite (R_n) pour que cela soit vérifié ! ◁

On peut en déduire, à l'aide du théorème de Weierstrass et d'un changement de variable, que le sous-espace des fonctions $x \longmapsto P(x)e^{-x}$ est dense pour la norme uniforme dans l'espace des fonctions continues sur $\mathbb{R}^+$ tendant vers 0 en $+\infty$.

Dans la série d'exercices qui vient, nous proposons des approximations explicites, de la fonction exponentielle pour le premier, de la fonction cotangente pour le second.

2.25. Convergence d'une suite de polynômes

1. Soit $n \in \mathbb{N}^*$ et $k \in [\![1, n]\!]$. Montrer que

$$1 - \frac{n(n-1)\dots(n-k+1)}{n^k} \leqslant \sum_{j=0}^{k-1} \frac{j}{n}.$$

2. En déduire que la suite de fonctions $(f_n)_{n\in\mathbb{N}^*}$ définie par $f_n(z) = \left(1 + \frac{z}{n}\right)^n$ converge uniformément sur tout compact de $\mathbb{C}$.

(École normale supérieure)

▷ **Solution.**

1. L'inégalité à démontrer peut s'écrire

$$1 - \prod_{j=0}^{k-1} \left(1 - \frac{j}{n}\right) \leqslant \sum_{j=0}^{k-1} \frac{j}{n}.$$

Nous démontrerons plus généralement que, pour tout $p \in \mathbb{N}$ et tout $(x_0, \dots, x_p) \in [0,1]^{p+1}$, on a

$$1 - \prod_{i=0}^{p} (1 - x_i) \leqslant \sum_{i=0}^{p} x_i,$$

en raisonnant par récurrence sur p.

La propriété est évidente pour $p = 0$ et si on la suppose vérifiée au rang $p \in \mathbb{N}$, alors on a, pour $(x_0, \dots, x_p, x_{p+1}) \in [0,1]^{p+1}$

$$\begin{aligned} 1 - \prod_{i=0}^{p+1} (1 - x_i) &= (1 - x_{p+1}) \left(1 - \prod_{i=0}^{p} (1 - x_i)\right) + x_{p+1} \\ &\leqslant (1 - x_{p+1}) \left(\sum_{i=0}^{p} x_i\right) + x_{p+1} \leqslant \sum_{i=0}^{p+1} x_i, \end{aligned}$$

ce qui termine la démonstration par récurrence.

2. Il est bien connu que, pour tout réel x, on a

$$\lim_{n\to+\infty} \left(1 + \frac{x}{n}\right)^n = e^x.$$

Montrons que la suite de fonctions $(f_n)_{n\in\mathbb{N}^*}$ converge vers la fonction exponentielle, uniformément sur tout compact de $\mathbb{C}$. Pour $z \in \mathbb{C}$ et $n \in \mathbb{N}^*$, on a

$$e^z = \sum_{k=0}^{n} \frac{z^k}{k!} + \sum_{k=n+1}^{+\infty} \frac{z^k}{k!} \text{ et } \left(1 + \frac{z}{n}\right)^n = \sum_{k=0}^{n} \mathrm{C}_n^k \left(\frac{z}{n}\right)^k.$$

On en déduit que

$$f_n(z) - e^z = \left(1 + \frac{z}{n}\right)^n - e^z = \sum_{k=0}^{n} \left(\frac{C_n^k}{n^k} - \frac{1}{k!}\right) z^k - \sum_{k=n+1}^{+\infty} \frac{z^k}{k!}$$

puis que

$$|f_n(z) - e^z| \leqslant \sum_{k=1}^{n} \left|\frac{n(n-1)\dots(n-k+1)}{n^k} - 1\right| \frac{|z|^k}{k!} + \sum_{k=n+1}^{+\infty} \frac{|z|^k}{k!}.$$

D'après la question 1, on a, pour $k \in [\![1, n]\!]$,

$$\left|\frac{n(n-1)\dots(n-k+1)}{n^k} - 1\right| = 1 - \frac{n(n-1)\dots(n-k+1)}{n^k} \leqslant \sum_{j=0}^{k-1} \frac{j}{n}$$
$$\leqslant \frac{k(k-1)}{2n}.$$

Soit K un compact de $\mathbb{C}$ et $a \in \mathbb{R}_+^*$ tel que, pour tout $z \in K$, on ait $|z| \leqslant a$. Alors, pour $n \in \mathbb{N}^*$ et $z \in K$,

$$|f_n(z) - e^z| \leqslant \frac{1}{2n} \sum_{k=1}^{n} \frac{k(k-1)}{k!} a^k + \sum_{k=n+1}^{+\infty} \frac{a^k}{k!}$$
$$\leqslant \frac{1}{2n} \sum_{k=2}^{n} \frac{a^k}{(k-2)!} + \sum_{k=n+1}^{+\infty} \frac{a^k}{k!}$$
$$\leqslant \frac{1}{2n} a^2 e^a + \sum_{k=n+1}^{+\infty} \frac{a^k}{k!}.$$

Sachant que $\lim_{n \to +\infty} \frac{1}{2n} a^2 e^a + \sum_{k=n+1}^{+\infty} \frac{a^k}{k!} = 0$, on en déduit que la suite $(f_n)_{n \in \mathbb{N}^*}$ converge uniformément sur K vers la fonction exponentielle. ◁

On pouvait montrer directement la question 2 de la manière suivante : prenons $M \geqslant 0$ *et* $z \in \mathbb{C}$ *avec* $|z| \leqslant M$. *À partir de*

$$|f_n(z) - e^z| \leqslant \sum_{k=1}^{n} \left|\frac{n(n-1)\dots(n-k+1)}{n^k} - 1\right| \frac{|z|^k}{k!} + \sum_{k=n+1}^{+\infty} \frac{|z|^k}{k!},$$

on peut écrire

$$|f_n(z) - e^z| \leqslant \sum_{k=1}^{n} \left(1 - \frac{n(n-1)\dots(n-k+1)}{n^k}\right) \frac{|z|^k}{k!} + \sum_{k=n+1}^{+\infty} \frac{|z|^k}{k!},$$

car $\dfrac{n(n-1))\cdots(n-k+1)}{n^k} \leqslant 1$. *Ainsi. on peut majorer*

$$|f_n(z)-e^z| \leqslant \sum_{k=1}^{n}\left(1-\frac{n(n-1)\ldots(n-k+1)}{n^k}\right)\frac{\mathrm{M}^k}{k!}+\sum_{k=n+1}^{+\infty}\frac{\mathrm{M}^k}{k!}$$

$$et\quad |f_n(z)-e^z| \leqslant e^{\mathrm{M}}-\left(1+\frac{\mathrm{M}}{n}\right)^n \xrightarrow[n\to\infty]{} 0,$$

ce qui prouve la convergence uniforme sur tout compact de $\mathbb{C}$.

Notons qu'il est vain d'espérer une convergence uniforme d'une suite de fonctions polynomiales vers l'exponentielle sur tout $\mathbb{C}$ *(voir la remarque qui suit l'exercice 2.18).*

Dans l'exercice suivant on obtient le développement eulérien de la fonction cotangente par des méthodes très élémentaires.

2.26. Développement eulérien de la cotangente

On pose pour x réel, $f(x)=\lim\limits_{n\to+\infty}\sum\limits_{k=-n}^{n}\dfrac{1}{x+k}$.

1. Montrer que f est définie sur $\mathbb{R}\backslash\mathbb{Z}$.

2. Vérifier les égalités ci-dessous :

a. pour $x\in\mathbb{R}\backslash\mathbb{Z}$, $f(-x)=-f(x)$;

b. pour $x\in\mathbb{R}\backslash\mathbb{Z}$, $f(x+1)=f(x)$;

c. pour $x\in\mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$, $f(2x)=\dfrac{1}{2}\left(f(x)+f\left(x+\dfrac{1}{2}\right)\right)$.

3. Montrer que la fonction $x\longmapsto f(x)-\pi\operatorname{cotan}(\pi x)$ admet un prolongement continu à $\mathbb{R}$ tout entier.

4. Montrer que pour tout $x\in\mathbb{R}\backslash\mathbb{Z}$, $f(x)=\pi\operatorname{cotan}\pi x$.

(École polytechnique)

▷ **Solution.**

1. Posons pour x dans $\mathbb{R}\backslash\mathbb{Z}$ et $n\geqslant 1$, $\mathrm{U}_n(x)=\sum\limits_{k=-n}^{n}\dfrac{1}{x+k}$. On peut écrire en regroupant les termes d'indices k et $-k$,

$$\begin{aligned}\mathrm{U}_n(x)&=\frac{1}{x}+\sum_{k=1}^{n}\frac{1}{x+k}+\sum_{k=1}^{n}\frac{1}{x-k}=\frac{1}{x}+\sum_{k=1}^{n}\left(\frac{1}{x+k}+\frac{1}{x-k}\right)\\&=\frac{1}{x}-2x\sum_{k=1}^{n}\frac{1}{k^2-x^2}=\frac{1}{x}-2x\sum_{k=1}^{n}u_k(x)\end{aligned}$$

avec $u_k(x) = \dfrac{1}{k^2 - x^2}$ pour $k \in \mathbb{N}^*$ et $x \in \mathbb{R}\backslash\mathbb{Z}$. Fixons x dans $\mathbb{R}\backslash\mathbb{Z}$. Au voisinage de $+\infty$, $u_k(x) \underset{k\to\infty}{\sim} \dfrac{1}{k^2}$. Le théorème de comparaison des séries à termes positifs garantit alors la convergence de la série $\sum u_k(x)$ et par conséquent l'existence de la limite de $\mathrm{U}_n(x)$. On a donc

$$f(x) = \frac{1}{x} - 2x \sum_{k=1}^{+\infty} \frac{1}{k^2 - x^2}.$$

2. a. L'expression que l'on vient d'établir prouve que f est impaire.

b. Soit $n \geqslant 1$ et $x \in \mathbb{R}\backslash\mathbb{Z}$. On a

$$\begin{aligned}\mathrm{U}_n(x+1) &= \sum_{k=-n}^{n} \frac{1}{x+1+k} \underset{\boxed{j=k+1}}{=} \sum_{j=-n+1}^{n+1} \frac{1}{x+j} \\ &= \mathrm{U}_{n-1}(x) + \frac{1}{x+n} + \frac{1}{x+n+1} \xrightarrow[n\to+\infty]{} f(x).\end{aligned}$$

Donc $f(x+1) = f(x)$ et la fonction f est 1-périodique.

c. Pour $n \geqslant 1$ et $x \in \mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$, on a en séparant les indice pairs et impairs,

$$\begin{aligned}\mathrm{U}_{2n}(2x) &= \sum_{k=-2n}^{2n} \frac{1}{2x+k} = \sum_{j=-n}^{n} \frac{1}{2x+2j} + \sum_{j=-n}^{n-1} \frac{1}{2x+2j+1} \\ &= \frac{1}{2} \sum_{j=-n}^{n} \frac{1}{x+j} + \frac{1}{2} \sum_{j=-n}^{n-1} \frac{1}{x+1/2+j} \\ &= \frac{1}{2}\mathrm{U}_n(x) + \frac{1}{2}\mathrm{U}_n(x+1/2) - \frac{1}{2x+1+2n},\end{aligned}$$

ce qui donne, en faisant tendre n vers $+\infty$,

$$f(2x) = \frac{1}{2}\left(f(x) + f\left(x + \frac{1}{2}\right)\right).$$

3. Notons g la fonction définie sur $\mathbb{R}\backslash\mathbb{Z}$ par $x \longmapsto f(x) - \pi \operatorname{cotan}(\pi x)$.

• Montrons que g est continue sur $\mathbb{R}\backslash\mathbb{Z}$. En tant que différence de deux fonctions 1-périodiques, elle est 1-périodique. Il suffit de prouver que g est continue sur $]0,1[$. Comme cotan est continue sur $]0,\pi[$, cela revient à montrer la continuité de f sur $]0,1[$. Vue l'expression de f trouvée à la question 1, nous avons à prouver la continuité sur $]0,1[$ de la fonction $x \longmapsto \sum_{k=1}^{+\infty} \dfrac{1}{k^2 - x^2}$. Nous allons faire un tout petit peu mieux en prouvant cette continuité sur $]-1,1[$, ce qui nous sera utile pour la suite.

• Considérons la fonction $h : x \in]-1,1[\longmapsto 2x \sum_{k=1}^{+\infty} \frac{1}{k^2 - x^2}$. Cette série de fonctions continues converge normalement sur $[-a,a]$, $a \in]0,1[$, car pour tout $x \in [-a,a]$,

$$0 \leqslant \frac{|2x|}{k^2 - x^2} \leqslant \frac{2}{k^2 - a^2}$$

et $\sum \frac{2}{k^2 - a^2}$ est une série convergente. Il s'ensuit que h est continue sur $[-a,a]$ pour tout $a < 1$, et donc sur $]-1,1[$. Ainsi f, et donc g, est continue sur $]0,1[$, et par 1-périodicité, sur $\mathbb{R}\backslash\mathbb{Z}$.

• Par périodicité de g il suffit de la prolonger par continuité en 0 pour prouver qu'elle se prolonge en une fonction continue sur $\mathbb{R}$. Pour cela, il s'agit donc de prouver que g admet une limite finie en 0. Pour x voisin de 0, on a

$$\pi \operatorname{cotan} \pi x = \pi \frac{\cos \pi x}{\sin \pi x} = \pi \frac{1 + o(x)}{\pi x + o(x^2)} = \frac{1}{x}(1 + o(x)) = \frac{1}{x} + o(1).$$

Pour x voisin de 0, on a donc

$$\begin{aligned} f(x) - \pi \operatorname{cotan} \pi x &= \frac{1}{x} - h(x) - \frac{1}{x} + o(1) \\ &= -h(x) + o(1) \xrightarrow[x \to 0]{} -h(0) + 0 = 0 \end{aligned}$$

puisque h est continue. Donc g se prolonge par continuité en 0, et par 1-périodicité en tout point de $\mathbb{Z}$.

4. • Posons $\tilde{f} : x \in \mathbb{R}\backslash\mathbb{Z} \longmapsto \pi \operatorname{cotan} \pi x$. Montrons que, comme f, pour tout $x \in \mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$, $\tilde{f}(2x) = \frac{1}{2}\left(\tilde{f}(x) + \tilde{f}\left(x + \frac{1}{2}\right)\right)$. Soit $x \in \mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$. On a

$$\begin{aligned} \tilde{f}(2x) &= \pi \frac{\cos 2\pi x}{\sin 2\pi x} = \pi \frac{\cos^2 \pi x - \sin^2 \pi x}{2 \sin \pi x \cos \pi x} = \frac{\pi}{2}\left(\frac{\cos \pi x}{\sin \pi x} - \frac{\sin \pi x}{\cos \pi x}\right) \\ &= \frac{\pi}{2}(\operatorname{cotan} \pi x - \tan \pi x) = \frac{\pi}{2}(\operatorname{cotan} \pi x + \operatorname{cotan}(\pi x + \pi/2)) \\ &= \frac{1}{2}(\tilde{f}(x) + \tilde{f}(x + 1/2)). \end{aligned}$$

• La fonction g qui vaut $g(x) = f(x) - \tilde{f}(x)$ pour $x \in \mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$ et prolongée à $\mathbb{R}$ par continuité, vérifie donc trivialement la même relation pour $x \in \mathbb{R}\backslash\frac{1}{2}\mathbb{Z}$:

$$g(2x) = \frac{1}{2}(g(x) + g(x + 1/2)). \qquad (*)$$

Comme g est continue, en faisant tendre x vers $x_0 \in \frac{1}{2}\mathbb{Z}$, $x \neq x_0$, on trouve que $(*)$ est aussi vraie pour un élément de $\frac{1}{2}\mathbb{Z}$, et donc pour tout $x \in \mathbb{R}$.

• La fonction $|g|$ est continue sur le compact $[0,1]$, donc elle est bornée et atteint ses bornes. En particulier, il existe $c \in [0,1]$ tel que $|g(c)| = \max_{x\in[0,1]} |g(x)|$. Mais alors d'après $(*)$, on a l'inégalité

$$|g(c)| = \left|\frac{g\left(\frac{c}{2}\right) + g\left(\frac{c+1}{2}\right)}{2}\right| \leqslant \frac{\left|g\left(\frac{c}{2}\right)\right| + \left|g\left(\frac{c+1}{2}\right)\right|}{2} \leqslant \frac{\left|g\left(\frac{c}{2}\right)\right|}{2} + \frac{|g(c)|}{2}$$

et donc, $\frac{|g(c)|}{2} \leqslant \frac{|g(c/2)|}{2}$ ce qui entraîne $|g(c/2)| \geqslant \max_{x\in[0,1]} |g(x)|$. On en déduit donc $|g(c/2)| = \max_{x\in[0,1]} |g(x)|$ et par une récurrence immédiate, pour tout $n \in \mathbb{N}$:

$$\max_{x\in[0,1]} |g(x)| = |g(c/2^n)| \xrightarrow[n\to+\infty]{} g(0) = 0$$

puisque g est continue en 0. Par conséquent, $\max_{x\in[0,1]} |g(x)| = 0$ et $g = 0$. On déduit donc que pour $x \in \mathbb{R}\backslash\mathbb{Z}$:

$$\boxed{f(x) = \lim_{n\to+\infty} \sum_{k=-n}^{n} \frac{1}{x+k} = \pi \operatorname{cotan} \pi x}. \quad \lhd$$

Le lecteur trouvera une autre manière d'obtenir ce développement à l'aide des séries de Fourier dans l'exercice 4.9.

Les trois exercices qui suivent concernent les produits infinis. Étant donnée une suite $(a_n) \in \mathbb{C}^{\mathbb{N}}$, on dira que le produit infini $\prod a_n$ converge lorsque la suite des produits partiels $\mathrm{P}_n = \prod_{k=0}^{n} a_k$ converge. On note alors $\prod_{k=0}^{+\infty} a_k = \lim_{n\to\infty} \mathrm{P}_n$. Notons que lorsqu'on développe la théorie des produits infinis, on impose en plus à cette limite d'être non nulle. Avec cette condition supplémentaire on voit facilement que pour qu'un produit infini converge il est nécessaire que le terme général a_n tende vers 1. Nous n'aurons toutefois pas besoin de cette convention dans les exercices qui suivent.

2.27. Produits infinis (1)

On pose, pour $z \in \mathbb{C}$, tel que $|z| < 1$,

$$f(z) = \prod_{k=1}^{+\infty}(1+z^k)\prod_{k=1}^{+\infty}(1-z^{2k-1}).$$

Justifier que f est définie, puis la calculer.

(École normale supérieure)

▷ **Solution.**

Posons, pour tout $n \in \mathbb{N}^*$,

$$a_n(z) = \prod_{k=1}^{n}(1+z^k) \text{ et } b_n(z) = \prod_{k=1}^{n}(1-z^{2k-1}).$$

Pour que z appartienne à l'ensemble de définition de f, il faut et suffit que les deux produits infinis convergent. Il est intéressant d'avoir un critère simple pour justifier qu'un produit infini de la forme $\prod(1+u_n)$ converge. Pour avoir des idées, imaginons que (u_n) soit une suite réelle positive qui converge vers 0. On a $\ln\prod_{k=1}^{n}(1+u_k) = \sum_{k=1}^{n}\ln(1+u_k)$ donc si la série $\sum\ln(1+u_k)$ converge, le produit infini converge. La réciproque est vraie si le produit infini converge vers une valeur non nulle (c'est pour ce genre de chose qu'il est commode de dire qu'un produit infini de limite nulle diverge). De plus comme $\ln(1+u_n) \sim u_n$, la convergence de la série $\sum\ln(1+u_n)$ équivaut à celle de la série $\sum u_n$ (on a pris u_n positif). Dans le cas complexe, c'est un peu plus difficile car on ne peut pas prendre le logarithme. On a tout de même le lemme suivant, fort utile en pratique, et qui va suffire pour l'exercice :

Lemme. *Soit $(u_n) \in \mathbb{C}^{\mathbb{N}}$. Si la série $\sum u_n$ converge absolument, alors le produit $\prod(1+u_n)$ converge. Si, de plus, on a $u_n \neq -1$ pour tout $n \in \mathbb{N}$, alors $\prod_{k=0}^{+\infty}(1+u_k)$ n'est pas nul.*

Démonstration.

• Posons, pour tout $n \in \mathbb{N}$, $P_n = \prod_{k=0}^{n}(1+u_k)$. Notons pour commencer que pour tout n,

$$|P_n| \leqslant \prod_{k=0}^{n}(1+|u_k|) \leqslant \prod_{k=0}^{n} e^{|u_k|} \leqslant e^{\sum_{k=0}^{n}|u_k|} \leqslant e^S,$$

où S est la somme de la série $\sum |u_n|$ (cela résulte de l'inégalité $1+x \leqslant e^x$, valable pour tout réel x). Ainsi la suite (P_n) est bornée. Nous savons que la suite (P_n) converge si et seulement si la série de terme général $\sum (P_n - P_{n-1})$ converge. Or celle-ci est absolument convergente car pour tout $n \geqslant 1$, on a

$$|P_n - P_{n-1}| = |P_{n-1}||u_n| \leqslant e^S |u_n|$$

et la série $\sum |u_n|$ converge par hypothèse. On peut donc conclure que la suite (P_n) converge.

• Supposons maintenant $1+u_n$ non nul pour tout n et montrons que la limite de (P_n) n'est pas nulle. Posons $Q_n = \prod_{k=0}^{n} \frac{1}{1+u_k}$, ce qui est possible puisque les u_k sont différents de -1. On peut alors écrire

$$Q_n = \prod_{k=0}^{n} \left(1 - \frac{u_k}{1+u_k}\right) = \prod_{k=0}^{n} (1+v_k)$$

avec $v_k = -\dfrac{u_k}{1+u_k}$ pour tout k. Comme la suite $(u_n)_{n\in\mathbb{N}}$ converge vers 0, $|v_n| \sim |u_n|$. Le théorème de comparaison des séries à termes positifs nous assure de la convergence de $\sum |v_n|$. D'après ce qui précède il s'ensuit que Q_n converge dans $\mathbb{C}$. Par définition on a $P_n Q_n = 1$ pour tout n. En passant à la limite on a $\lim_{n\to+\infty} P_n \lim_{n\to+\infty} Q_n = 1$ et la limite de la suite $(P_n)_{n\geqslant 0}$ ne peut être nulle. ◇

Il résulte de ce lemme que $f(z)$ est bien défini si $|z| < 1$: en effet, les deux séries $\sum z^k$ et $\sum -z^{2k-1}$ convergent absolument. De plus on a, pour $|z| < 1$,

$$a_n(z) = \prod_{k=1}^{n} (1+z^k) = \frac{\prod_{k=1}^{n} (1-z^{2k})}{\prod_{k=1}^{n} (1-z^k)}.$$

On en déduit que

$$a_n(z) b_n(z) = \frac{\prod_{k=1}^{2n} (1-z^k)}{\prod_{k=1}^{n} (1-z^k)}.$$

Quand n tend vers $+\infty$ le numérateur et le dénominateur de cette expression tendent tous les deux vers $\prod_{k=1}^{+\infty} (1-z^k)$. Le produit $a_n(z)b_n(z)$ tend donc vers 1.

Conclusion. Pour $|z| < 1$, $\boxed{f(z) = 1}$. ◁

L'énoncé suivant est un bon moyen de s'exercer au calcul algébrique !

2.28. Produits infinis (2)

Soit $n \in \mathbb{N}^*$, q, z deux complexes tels que $0 < |q| < 1$ et $z \neq 0$. On pose

$$\mathrm{P}(z) = \prod_{k=1}^{n} (1 + q^{2k-1}z)(1 + q^{2k-1}z^{-1})$$

1. Montrer que $\mathrm{P}(z)$ se met sous la forme

$$\mathrm{P}(z) = c_0 + c_1(z + z^{-1}) + \cdots + c_n(z^n + z^{-n}).$$

2. Calculer les c_i en fonction de n, i et q.

3. Étudier les c_i et P lorsque n tend vers l'infini.

(École normale supérieure)

▷ **Solution.**

1. Il est clair que P est une fraction rationnelle dont le seul pôle est 0, pôle d'ordre n. Il existe donc des nombres complexes $c_{-n}, c_{-(n-1)}, \ldots, c_{n-1}, c_n$ tels que $\mathrm{P}(z) = \sum_{i=-n}^{n} c_i z^i$. C'est tout simplement la décomposition en éléments simples de P. On remarque alors que $\mathrm{P}(z) = \mathrm{P}(z^{-1})$, d'où par unicité de la décomposition en éléments simples les relations $c_{-i} = c_i$ pour tout $i \in [\![1, n]\!]$. Cela donne le résultat voulu.

2. Pour calculer les coefficients c_i, on remarque qu'il y a une relation simple entre $\mathrm{P}(z)$ et $\mathrm{P}(q^2 z)$. En effet, on obtient

$$\begin{aligned}\mathrm{P}(q^2z) &= \prod_{k=1}^{n} (1 + q^{2k+1}z)(1 + q^{2k-3}z^{-1}) \\ &= \mathrm{P}(z)\frac{1 + q^{2n+1}z}{1 + qz}\,\frac{1 + q^{-1}z^{-1}}{1 + q^{2n-1}z^{-1}} = \mathrm{P}(z)\frac{1 + q^{2n+1}z}{qz + q^{2n}}.\end{aligned}$$

De cette égalité qui s'écrit

$$(qz+q^{2n})\Big(c_0+\sum_{i=1}^{n} c_i(q^{2i}z^i+q^{-2i}z^{-i})\Big) = (1+q^{2n+1}z)\Big(c_0+\sum_{i=1}^{n} c_i(z^i+z^{-i})\Big),$$

on déduit, en comparant les coefficients de z^{i+1} dans les deux termes que, pour $0 \leqslant i \leqslant n-1$,

$$c_{i+1}q^{2(i+1)}q^{2n} + c_i q^{2i} q = c_{i+1} + q^{2n+1}c_i,$$

et donc

$$c_i = c_{i+1}\frac{1 - q^{2(n+i+1)}}{q^{2i+1} - q^{2n+1}} = \frac{c_{i+1}(1 - q^{2(n+i+1)})}{q^{2i+1}(1 - q^{2(n-i)})}.$$

On calcule directement

$$c_n = qq^3 \dots q^{2n-1} = q^{1+3+\cdots+(2n-1)} = q^{n^2}.$$

On en déduit que, pour $0 \leqslant i \leqslant n$, on a

$$c_i = \frac{c_n \prod_{k=n-i+1}^{2n} (1-q^{2k})}{q^{(2i+1)+(2i+3)+\cdots+(2n-1)} \prod_{k=1}^{n-i} (1-q^{2k})}.$$

La somme $(2i+1)+(2i+3)+\cdots+(2n-1)$ étant égale à $(n-i)(n+i)$ (c'est la somme de $n-i$ termes consécutifs d'une suite arithmétique), on obtient finalement, en remplaçant c_n par sa valeur,

$$\boxed{c_i = \frac{q^{i^2} \prod_{k=n+i+1}^{2n} (1-q^{2k})}{\prod_{k=1}^{n-i} (1-q^{2k})}}.$$

3. On fait maintenant varier n. Modifions légèrement les notations, en posant, pour tout $n \in \mathbb{N}$,

$$\begin{cases} \mathrm{P}_n = \prod_{k=1}^{n} (1+q^{2k-1}z)(1+q^{2k-1}z^{-1}) \\ c_i(n) = \dfrac{q^{i^2} \prod_{k=n+i+1}^{2n} (1-q^{2k})}{\prod_{k=1}^{n-i} (1-q^{2k})} \quad \text{pour} \quad 0 \leqslant i \leqslant n. \end{cases}$$

Les différentes suites qui apparaissent ici font intervenir des produits. Du lemme prouvé dans l'exercice précédent, on déduit immédiatement que la suite (P_n) converge. En effet, on peut écrire, pour tout $n \in \mathbb{N}$,

$$\mathrm{P}_n = \prod_{k=1}^{n} (1+q^{2k-1}z) \prod_{k=1}^{n} (1+q^{2k-1}z^{-1}).$$

Les séries $\sum |q^{2k-1}z|$ et $\sum |q^{2k-1}z^{-1}|$ convergent puisque $|q| < 1$, donc les produits sont convergents. Ceci montre que (P_n) converge vers

$$\mathrm{P} = \prod_{k=1}^{+\infty} (1+q^{2k-1}z) \prod_{k=1}^{+\infty} (1+q^{2k-1}z^{-1}).$$

Pour $i \in \mathbb{N}$ et $n \geqslant i$, on peut écrire

$$c_i(n) = \frac{q^{i^2} \prod_{k=n+i+1}^{2n} (1-q^{2k})}{\prod_{k=1}^{n-i} (1-q^{2k})} = \frac{q^{i^2} \prod_{k=1}^{2n} (1-q^{2k})}{\prod_{k=1}^{n+i} (1-q^{2k}) \prod_{k=1}^{n-i} (1-q^{2k})}.$$

Le produit $\prod(1-q^{2k})$ converge, puisque $\sum |q^{2k}|$ converge (car $|q|<1$). De plus, sa limite n'est pas nulle car, pour tout $k \in \mathbb{N}$, on a $q^{2k} \neq 1$. On en déduit que la suite $(c_i(n))_{n \geqslant i}$ converge et que sa limite est égale à

$$l_i = \frac{q^{i^2} \prod_{k=1}^{+\infty} (1-q^{2k})}{\left(\prod_{k=1}^{+\infty} (1-q^{2k})\right)^2} = \frac{q^{i^2}}{\prod_{k=1}^{+\infty} (1-q^{2k})}.$$

La suite (P_n) vérifie, pour tout $n \in \mathbb{N}$,

$$\mathrm{P}_n = c_0(n) + \sum_{i=1}^{n} c_i(n)(z^i + z^{-i}).$$

Il est tentant de penser que

$$\mathrm{P} = l_0 + \sum_{i=1}^{+\infty} l_i(z^i + z^{-i}).$$

On a, pour tout $i \in \mathbb{N}$, $\dfrac{l_{i+1}}{l_i} = \dfrac{q^{(i+1)^2}}{q^{i^2}} = q^{2i+1}$ et donc $\lim\limits_{i \to +\infty} \dfrac{l_{i+1}}{l_i} = 0$. Le rayon de convergence de la série entière $\sum l_i z^i$ est donc infini, ce qui montre l'existence de $\mathrm{Q} = l_0 + \sum\limits_{i=1}^{+\infty} l_i(z^i + z^{-i})$. On a alors, pour tout $n \in \mathbb{N}^*$,

$$\mathrm{P}_n - \mathrm{Q} = c_0(n) - l_0 + \sum_{i=1}^{n} (l_i - c_i(n))(z^i + z^{-i}) + \sum_{i=n+1}^{+\infty} l_i(z^i + z^{-i}).$$

Pour démontrer que $\mathrm{P}_n - \mathrm{Q}$ converge vers 0, il faut intervertir sommation et passage à la limite. Pour cela, il suffit que la série définissant P_n converge uniformément par rapport à n. Montrons donc qu'on peut majorer $|c_i(n)|$ pour $0 \leqslant i \leqslant n$, indépendamment de n, par le terme d'une série convergente. On écrit

$$c_i(n) = \frac{q^{i^2} \prod_{k=1}^{2n} (1-q^{2k})}{\prod_{k=1}^{n+i} (1-q^{2k}) \prod_{k=1}^{n-i} (1-q^{2k})}.$$

La suite de terme général $v_n = \prod_{k=0}^{n} (1 - q^{2k})$ convergeant vers une limite non nulle, il existe des réels a et b, strictement positifs, tels que, pour tout $n \in \mathbb{N}$, $a \leqslant |v_n| \leqslant b$. On en déduit que, pour $0 \leqslant i \leqslant n$, on a $|c_i(n)| \leqslant |q|^{i^2} ba^{-2} = \mathrm{C}|q|^{i^2}$. Par passage à la limite, on obtient la même majoration $|l_i| \leqslant \mathrm{C}|q|^{i^2}$ pour l_i. On peut alors écrire, pour $0 \leqslant m \leqslant n$,

$$\begin{aligned} |\mathrm{P}_n - \mathrm{Q}| &\leqslant |c_0(n) - l_0| + \sum_{i=1}^{m} |c_i(n) - l_i| \left(|z|^i + |z|^{-i}\right) + \\ &\quad \sum_{i=m+1}^{n} 2\mathrm{C}|q|^{i^2} \left(|z|^i + |z|^{-i}\right) + \sum_{i=n+1}^{+\infty} \mathrm{C}|q|^{i^2} \left(|z|^i + |z|^{-i}\right) \\ &\leqslant |c_0(n) - l_0| + \sum_{i=1}^{m} |c_i(n) - l_i| \left(|z|^i + |z|^{-i}\right) \\ &\quad + 2 \sum_{i=m+1}^{+\infty} \mathrm{C}|q|^{i^2} \left(|z|^i + |z|^{-i}\right). \end{aligned}$$

Soit $\varepsilon > 0$. Choisissons m tel que $2 \sum_{i=m+1}^{+\infty} \mathrm{C}|q|^{i^2} \left(|z|^i + |z|^{-i}\right) \leqslant \varepsilon$. Alors, pour tout $n \geqslant m$, on a

$$|\mathrm{P}_n - \mathrm{Q}| \leqslant |c_0(n) - l_0| + \sum_{i=1}^{m} |c_i(n) - l_i| \left(|z|^i + |z|^{-i}\right) + \varepsilon.$$

Le terme de droite tend vers ε quand n tend vers $+\infty$. Pour n assez grand, on aura $|\mathrm{P}_n - \mathrm{Q}| \leqslant 2\varepsilon$, ce qui montre que $\lim_{n \to \infty} \mathrm{P}_n = \mathrm{Q}$, c'est-à-dire que $\mathrm{Q} = \mathrm{P}$. On a donc démontré que

$$\prod_{k=1}^{+\infty} (1 + q^{2k-1} z) \prod_{k=1}^{+\infty} (1 + q^{2k-1} z^{-1}) = l_0 + \sum_{i=1}^{+\infty} l_i \left(z^i + z^{-i}\right),$$

c'est-à-dire

$$\prod_{k=1}^{+\infty} (1 + q^{2k-1} z) \prod_{k=1}^{+\infty} (1 + q^{2k-1} z^{-1}) \prod_{k=1}^{+\infty} (1 - q^{2k}) = 1 + \sum_{i=1}^{+\infty} q^{i^2} (z^i + z^{-i}),$$

ce qui peut encore s'écrire

$$\boxed{\prod_{k=1}^{+\infty} (1 + q^{2k-1} z)(1 + q^{2k-1} z^{-1})(1 - q^{2k}) = \sum_{i \in \mathbb{Z}} q^{i^2} z^i} \quad \lhd$$

2.29. Développement eulérien du sinus sur $\mathbb{C}$

1. Soit $z \in \mathbb{C}$. Montrer que

$$\left(1+\frac{iz}{2p}\right)^{2p} - \left(1-\frac{iz}{2p}\right)^{2p} = 2iz\prod_{k=1}^{p-1}\left(1-\frac{z^2}{4p^2\tan^2\dfrac{k\pi}{2p}}\right).$$

2. Montrer que pour tout $z \in \mathbb{C}$, $\sin z = z\prod_{p=1}^{+\infty}\left(1-\dfrac{z^2}{p^2\pi^2}\right)$.

(**École polytechnique**)

▷ **Solution.**

1. Considérons le polynôme $P = (1+iX)^{2p} - (1-iX)^{2p}$. Le coefficient de X^{2p} est nul et celui de X^{2p-1} égal à $4p(-1)^{p-1}i$. Donc P est de degré $2p-1$. Cherchons ses racines. Pour z complexe distinct de $-i$ (qui n'est pas racine de P), on a

$$\begin{aligned}
P(z) = 0 &\iff (1+iz)^{2p} - (1-iz)^{2p} = 0 \iff \left(\frac{1+iz}{1-iz}\right)^{2p} = 0\\
&\iff \exists k \in [\![0, 2p-1]\!],\ \frac{1+iz}{1-iz} = \omega_k,\ \text{où } \omega_k = e^{\frac{2ik\pi}{2p}}\\
&\iff \exists k \in [\![0, 2p-1]\!],\ (1+iz) = \omega_k(1-iz)\\
&\iff \exists k \in [\![0, 2p-1]\!],\ 1-\omega_k = -i(1+\omega_k)z\\
&\iff \exists k \in [\![0, 2p-1]\!],\ \omega_k \neq -1 \text{ et } z = i\frac{1-\omega_k}{1+\omega_k}\\
&\iff \exists k \in [\![0, 2p-1]\!]\backslash\{p\},\ z = \tan\frac{k\pi}{2p}
\end{aligned}$$

après simplification de $i\dfrac{1-e^{i\theta}}{1+e^{i\theta}}$ en $\tan\dfrac{\theta}{2}$. On obtient ainsi $2p-1$ racines deux à deux distinctes. À un scalaire multiplicatif près, P est donc égal à $X\prod_{\substack{1\leqslant k\leqslant 2p-1\\ k\neq p}}\left(X-\tan\dfrac{k\pi}{2p}\right)$. Pour k décrivant $[\![1, p-1]\!]$,

$$\tan\left(\frac{(2p-k)\pi}{2p}\right) = \tan\left(\pi - \frac{k\pi}{2p}\right) = -\tan\left(\frac{k\pi}{2p}\right)$$

et $2p-k$ décrit $[\![p+1, 2p-1]\!]$. En associant les facteurs d'indices k et $2p-k$, on en déduit qu'à un scalaire près, P est égal à $X\prod_{k=1}^{p-1}\left(X^2-\tan^2\dfrac{k\pi}{2p}\right)$.

En divisant chaque facteur par $-\tan^2 \dfrac{k\pi}{2p}$ qui est bien distinct de 0, on en déduit l'existence d'un scalaire $\lambda \in \mathbb{C}^*$ tel que

$$\mathrm{P} = \lambda \mathrm{X} \prod_{k=1}^{p-1} \left(1 - \frac{\mathrm{X}^2}{\tan^2 \dfrac{k\pi}{2p}} \right).$$

La valeur de λ est obtenue en regardant le coefficient de X dans P. On trouve $\lambda = 4ip$. Ainsi,

$$\mathrm{P} = 4ip\mathrm{X} \prod_{k=1}^{p-1} \left(1 - \frac{\mathrm{X}^2}{\tan^2 \dfrac{k\pi}{2p}} \right)$$

et en évaluant P en $\dfrac{z}{2p}$, on arrive à la relation voulue

$$\boxed{\left(1 + \frac{iz}{2p}\right)^{2p} - \left(1 - \frac{iz}{2p}\right)^{2p} = 2iz \prod_{k=1}^{p-1} \left(1 - \frac{z^2}{4p^2 \tan^2 \dfrac{k\pi}{2p}} \right)}.$$

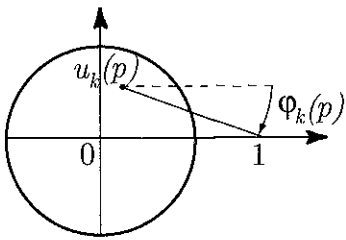
2. Soit z un nombre complexe. L'idée est bien entendu de faire tendre p vers l'infini dans l'égalité ci-dessus. Comme $\lim\limits_{p \to +\infty} \left(1 + \dfrac{iz}{2p}\right)^{2p} = e^{iz}$ et $\lim\limits_{p \to +\infty} \left(1 - \dfrac{iz}{2p}\right)^{2p} = e^{-iz}$ (voir l'exercice 2.25), le terme de gauche de cette égalité tend vers $2i \sin z$. Il reste à regarder le comportement du produit.

Posons, pour $1 \leqslant k < p$, $u_k(p) = \dfrac{z^2}{4p^2 \tan^2 \dfrac{k\pi}{p}} \cdot \dfrac{k\pi}{p}$ et $u_k(p) = 0$ pour $k \geqslant p$. Notons alors

$$\mathrm{F}(p) = \prod_{k=1}^{+\infty} (1 - u_k(p)) = \prod_{k=1}^{p-1} \left(1 - \frac{z^2}{4p^2 \tan^2 \dfrac{k\pi}{2p}} \right)$$

le produit qui nous intéresse. Si k est fixé, le terme $u_k(p)$ tend vers $u_k = \dfrac{z^2}{4k^2\pi^2}$ lorsque p tend vers l'infini. Il est assez naturel d'imaginer que $\lim\limits_{p \to +\infty} \mathrm{F}(p) = \prod\limits_{k=1}^{+\infty} (1 - u_k)$ et c'est ce que nous allons prouver. Il s'agit d'un problème d'interversion de limites.

On va utiliser la forme trigonométrique de $1-u_k(p)$ pour se ramener aux théorèmes sur les séries de fonctions que nous connaissons. Comme pour $x \in \left]0, \frac{\pi}{2}\right[$, $\tan x \geqslant x$, on a $|u_k(p)| \leqslant \frac{|z|^2}{k^2\pi^2}$ pour tout k et p dans $\mathbb{N}^*$. Il existe donc un rang $N \geqslant 1$ tel que pour tout $k \geqslant N$ et tout p, $|u_k(p)| \leqslant \frac{1}{2}$. À partir de ce rang, on peut choisir un argument $\varphi_k(p)$ de $1-u_k(p)$ compris entre $-\frac{\pi}{2}$ et $\frac{\pi}{2}$:



On peut alors écrire pour $k \geqslant N$,

$$1-u_k(p) = \exp\big(\ln|1-u_k(p)| + i\varphi_k(p)\big) \quad \text{avec} \quad -\frac{\pi}{2} \leqslant \varphi_k(p) \leqslant \frac{\pi}{2}.$$

On va maintenant montrer que la série $\sum_{k=N}^{+\infty} (\ln|1-u_k(p)| + i\varphi_k(p))$ converge normalement en p ce qui permettra de passer à la limite.

• On va d'abord majorer le terme $|\ln|1-u_k(p)||$ indépendamment de p. Pour $|u|<1$ on a par croissance du logarithme

$$\ln(1-|u|) \leqslant \ln|1-u| \leqslant \ln(1+|u|).$$

Comme $\left|\frac{\ln(1-|u|)}{|u|}\right|$ et $\left|\frac{\ln(1+|u|)}{|u|}\right|$ tendent vers 1 lorsque $|u|$ tend vers 0, on en déduit l'existence d'un réel $\alpha > 0$ tel que $|\ln|1-u|| \leqslant 2|u|$ pour $|u| \leqslant \alpha$. Quitte à changer N, on peut supposer que $\frac{|z|^2}{k^2\pi^2} \leqslant \alpha$ pour tout $k \geqslant N$. On a alors pour tout $p \geqslant 1$,

$$|\ln|1-u_k(p)|| \leqslant 2|u_k(p)|.$$

• Occupons nous maintenant du terme $\varphi_k(p)$. On constate que $\sin|\varphi_k(p)| \leqslant |u_k(p)|$ (c'est clair sur la figure ci-dessus). Comme $\sin x \geqslant \frac{2}{\pi}x$ pour $0 \leqslant x \leqslant \frac{\pi}{2}$, on a

$$|\varphi_k(p)| \leqslant \frac{\pi}{2}|u_k(p)|.$$

Ainsi, en posant $C = 2 + \dfrac{\pi}{2}$, on a pour tout p,

$$|\ln|1-u_k(p)| + i\varphi_k(p)| \leqslant C|u_k(p)| \leqslant \frac{C|z|^2}{k^2\pi^2},$$

ce qui garantit la convergence normale de $\displaystyle\sum_{k=N}^{+\infty} (\ln|1-u_k(p)| + i\varphi_k(p))$.

Comme la fonction argument $z \in \mathbb{C} \setminus \mathbb{R}_- \longmapsto \arg z \in]-\pi,\pi[$ est continue, la limite de $\ln|1-u_k(p)| + i\varphi_k(p)$ lorsque p tend vers l'infini n'est autre que $\ln|1-u_k| + i\arg(1-u_k)$. Comme on peut passer à la limite dans la série, on a

$$\sum_{k=N}^{+\infty} (\ln|1-u_k(p)| + i\varphi_k(p)) \xrightarrow[p\to+\infty]{} \sum_{k=N}^{+\infty} \ln|1-u_k| + i\arg(1-u_k).$$

La fonction exponentielle étant continue, $\exp(\ln|1-u_k(p)| + i\varphi_k(p))$ tend vers $\exp\left(\displaystyle\sum_{k=N}^{+\infty} \ln|1-u_k| + i\arg(1-u_k)\right)$ et, toujours par continuité de l'exponentielle,

$$\exp\left(\sum_{k=N}^{+\infty} \ln|1-u_k| + i\arg(1-u_k)\right) = \prod_{k=N}^{+\infty}(1-u_k).$$

On en déduit donc que la limite de $F(p)$ lorsque p tend vers l'infini n'est autre $\displaystyle\prod_{k=1}^{+\infty}(1-u_k)$. Nous venons de prouver que

$$\lim_{p\to+\infty} \prod_{k=1}^{p-1} \left(1 - \frac{z^2}{4p^2\tan^2\dfrac{k\pi}{2p}}\right) = \prod_{k=1}^{+\infty}\left(1-\frac{z^2}{k^2\pi^2}\right).$$

En passant à la limite dans la relation établie à la question précédente, on obtient la décomposition en facteur de $\sin z$:

$$\boxed{\sin z = z\prod_{k=1}^{+\infty}\left(1-\frac{z^2}{k^2\pi^2}\right)}. \quad \triangleleft$$

Dans cette dernière partie, on s'intéresse aux conditions suffisantes qui permettent d'assurer la convergence uniforme d'une suite de fonctions. Beaucoup de résultats sont le fruit de recherches des écoles allemande et italienne à la fin du XIX^e siècle. Dini est à l'origine des résultats suivants qui sont très souvent posés aux oraux des concours.

2.30. Théorème de Dini

Soit $a < b$ dans $\mathbb{R}$, $f_n : [a,b] \longrightarrow \mathbb{R}$ une suite de fonctions continues qui converge simplement vers une fonction continue f.

1. *Théorème de Dini :* on suppose que la suite $(f_n)_{n\in\mathbb{N}}$ est croissante, *i.e.* $f_n \leqslant f_{n+1}$. Démontrer que la convergence est uniforme.

2. On suppose que chaque fonction f_n est croissante. Montrer que la convergence est uniforme.

3. On considère la suite de fonctions $(p_n)_{n\geqslant 0}$ de $[0,1]$ dans $\mathbb{R}$ définie par $p_0 = 0$ et $p_{n+1}(x) = p_n(x) + \dfrac{1}{2}(x - p_n^2(x))$ pour tout n et tout x. Montrer que la suite $(p_n)_{n\geqslant 0}$ converge uniformément sur $[0,1]$.

(École polytechnique, école normale supérieure)

▷ **Solution.**

1. Nous proposons deux rédactions pour cette question.

• La première repose sur la propriété selon laquelle l'intersection d'une suite décroissante de compacts non vides n'est pas vide : posons pour $n \geqslant 0$, $g_n = f - f_n \geqslant 0$. Alors (g_n) est une suite décroissante de fonctions continues qui converge simplement vers la fonction nulle. Notons $\alpha_n = \|g_n\|_\infty$. Il s'agit de prouver que cette suite converge vers 0. Comme $\alpha_n = \sup\limits_{x\in[a,b]} g_n(x)$, la suite (α_n) est décroissante et positive. Elle converge donc vers un réel $\alpha \geqslant 0$. Raisonnons par l'absurde et supposons $\alpha > 0$. Considérons pour $n \geqslant 0$ les parties

$$\mathrm{K}_n = \left\{x \in [a,b], g_n(x) \geqslant \frac{\alpha}{2}\right\}.$$

Chaque partie K_n est non vide, bornée car contenue dans $[a,b]$, et en tant qu'image réciproque du fermé $[\alpha/2, +\infty[$ par la fonction continue g_n, K_n est fermée dans $[a,b]$ donc dans $\mathbb{R}$. Comme $g_{n+1} \leqslant g_n$, on a $\mathrm{K}_{n+1} \subset \mathrm{K}_n$. Il s'ensuit, par le théorème des compacts emboîtés, que $\bigcap\limits_{n\geqslant 0} \mathrm{K}_n$ est non vide. Il existe donc c appartenant à tout K_n et par conséquent tel que $g_n(c) \geqslant \dfrac{\alpha}{2}$ pour tout n, ce qui contredit la convergence simple vers 0.

• La seconde rédaction fait appel à la propriété de Borel-Lebesgue. Soit $\varepsilon > 0$. Les parties

$$\Omega_n = \{x \in [a,b], f(x) - \varepsilon < f_n(x)\}$$

constituent une suite croissante d'ouverts de $[a,b]$: en effet, Ω_n est l'image réciproque de l'ouvert $]-\infty, \varepsilon[$ par la fonction continue $f - f_n$, c'est donc un ouvert. Comme $f_n \leqslant f_{n+1}$, Ω_n est inclus dans Ω_{n+1}.

Comme la suite (f_n) converge simplement vers f, la réunion des Ω_n fait $[a,b]$ tout entier. Comme $[a,b]$ est compact, en vertu de la propriété de Borel-Lebesgue, il existe $n_0 \in \mathbb{N}$ tel que $\Omega_{n_0} = [a,b]$. Dans ces conditions, si $n \geqslant n_0$ et $x \in [a,b]$, on a $f(x) - f_{n_0}(x) \leqslant \varepsilon$ et

$$f(x) - \varepsilon < f_{n_0}(x) \leqslant f_n(x) \leqslant f(x).$$

Ainsi $\|f - f_n\| \leqslant \varepsilon$ et la convergence est donc uniforme sur $[a,b]$.

Ce résultat fondamental constitue le théorème de Dini. Le résultat de la question suivante est souvent qualifié de second théorème de Dini, voire même de « faux Dini ».

2. Sur un ensemble fini, la convergence simple équivaut à la convergence uniforme. On va donc se ramener à une subdivision (finie) de $[a,b]$ et contrôler l'écart entre f et f_n sur le reste du segment grâce à l'hypothèse de monotonie.

Soit $\varepsilon > 0$. La fonction f est continue sur le compact $[a,b]$. D'après le théorème de Heine, elle est uniformément continue. Considérons $\eta > 0$ un module d'uniforme continuité de f pour ε et une subdivision $S = (a = a_0 < a_1 < \cdots < a_{n-1} < a_n = b)$ de $[a,b]$ de pas inférieur à η. Il existe $n_0 \in \mathbb{N}$ tel que pour tout entier $n \geqslant n_0$ et tout $i \in [\![0,n]\!]$, $|f(a_i) - f_n(a_i)| \leqslant \varepsilon$. Soit $n \geqslant n_0$ et $x \in [a,b]$. Supposons $a_i \leqslant x \leqslant a_{i+1}$ avec $0 \leqslant i \leqslant n-1$. On a $|f(x) - f(a_i)| \leqslant \varepsilon$ et on peut alors écrire

$$\begin{aligned}
|f(x) - f_n(x)| &\leqslant |f(x) - f(a_i)| + |f(a_i) - f_n(a_i)| + |f_n(a_i) - f_n(x)| \\
&\leqslant \varepsilon + \varepsilon + (f_n(x) - f_n(a_i)) \\
&\leqslant 2\varepsilon + f_n(a_{i+1}) - f_n(a_i) \text{ (car } f_n \text{ est croissante)} \\
&\leqslant 2\varepsilon + |f_n(a_{i+1}) - f(a_{i+1})| + |f(a_{i+1}) - f(a_i)| \\
&\qquad + |f(a_i) - f_n(a_i)| \\
&\leqslant 5\varepsilon.
\end{aligned}$$

La convergence est donc uniforme.

3. Cette question offre une application du (premier) théorème de Dini. Soit $x \in [0,1]$ fixé. La suite $p_n(x)$ est la suite récurrente associée à la fonction $g_x : t \mapsto t + \dfrac{1}{2}(x - t^2)$ avec la condition initiale $p_0(x) = 0$. Les points fixes de g_x sont $-\sqrt{x}$ et $\sqrt{x}$. Il est facile de voir que l'intervalle $[0, \sqrt{x}]$ est stable par g_x et que celle-ci y est croissante avec $g_x(t) \geqslant t$ pour tout $t \in [0, \sqrt{x}]$. On en déduit que la suite $(p_n(x))_{n \geqslant 0}$ est croissante et converge vers $\sqrt{x}$. La fonction racine carrée étant continue sur $[0,1]$ on peut applique le résultat de la question 1 et dire que la convergence de la suite (p_n) est uniforme. ◁

On notera que dans les deux résultats de Dini l'hypothèse de continuité de f est fondamentale. Le lecteur trouvera sans mal des exemples où la convergence n'est pas uniforme si on ôte cette hypothèse.

Dans l'exercice suivant, on utilise le théorème de Dini pour s'assurer de la continuité d'une limite simple.

2.31. Un théorème de point fixe

Soit $H : [0,1]^2 \longmapsto \mathbb{R}_+$ continue telle que, pour tout $x \in [0,1]$, $H(x,x) = 0$. Pour toute application $\varphi : [0,1] \longrightarrow \mathbb{R}$ bornée, on pose

$$\forall t \in [0,1],\ (T\varphi)(t) = \inf_{s\in[0,1]} (\varphi(s) + H(t,s)).$$

1. Montrer que $T\varphi$ est continue. On définit donc ainsi une application T de $\mathcal{C}([0,1],\mathbb{R})$ dans lui-même.

2. Caractériser les points fixes de T (vue comme application de $\mathcal{C}([0,1],\mathbb{R})$ dans lui-même).

3. Montrer que T est continue de $\mathcal{C}([0,1],\mathbb{R})$ dans lui-même pour la norme infinie.

4. Soit $\varphi \in \mathcal{C}([0,1],\mathbb{R})$. Montrer que la suite $(T^n\varphi)_{n\geqslant 0}$ converge uniformément vers un point fixe de T.

(École polytechnique)

▷ **Solution.**

1. D'après le théorème de Heine, la fonction H est uniformément continue sur le compact $[0,1]^2$. Munissons $\mathbb{R}^2$ de la norme définie par $\|(x,y)\| = \max(|x|,|y|)$. Pour tout $\varepsilon > 0$, il existe donc $\eta > 0$ tel que $|H(t,s) - H(t',s')| \leqslant \varepsilon$ si $\|(s,t) - (s',t')\| \leqslant \eta$. Si t et t' sont deux éléments de $[0,1]$ tels que $|t - t'| \leqslant \eta$, on a pour tout $s \in [0,1]$, $\|(s,t) - (s,t')\| \leqslant \eta$ et donc $|H(t,s) - H(t',s)| \leqslant \varepsilon$. On en déduit que, pour tout $s \in [0,1]$, on a

$$T_\varphi(t) \leqslant \varphi(s) + H(t,s) \leqslant \varphi(s) + H(t',s) + \varepsilon.$$

Cela étant vrai pour tout $s \in [0,1]$, on a

$$T_\varphi(t) \leqslant T_\varphi(t') + \varepsilon.$$

En échangeant les rôles de t et t', on obtient de même $T_\varphi(t') \leqslant T_\varphi(t) + \varepsilon$ et finalement

$$|T_\varphi(t) - T_\varphi(t')| \leqslant \varepsilon.$$

Ceci démontre la continuité de T_φ.

2. On cherche les fonctions φ de $\mathcal{C}([0,1],\mathbb{R})$ telles que $T_\varphi(t) = \varphi(t)$ pour tout $t \in [0,1]$. Notons déjà que

$$T_\varphi(t) \leqslant \varphi(t) + H(t,t) = \varphi(t).$$

Pour avoir l'inégalité inverse, il faut $\varphi(t) \leqslant \varphi(s) + \mathrm{H}(s,t)$ pour tout $s \in [0,1]$. Finalement, φ est un point fixe de T si, pour tout $(s,t) \in [0,1]^2$, on a

$$\varphi(t) \leqslant \varphi(s) + \mathrm{H}(t,s).$$

3. Soit φ et ψ dans $\mathcal{C}([0,1],\mathbb{R})$. On a, pour tout $(s,t) \in [0,1]^2$,

$$\begin{aligned}\mathrm{T}_\varphi(t) &\leqslant \varphi(s) + \mathrm{H}(t,s)\\ &\leqslant \varphi(s) - \psi(s) + \psi(s) + \mathrm{H}(t,s)\\ &\leqslant \|\varphi - \psi\|_\infty + \psi(s) + \mathrm{H}(t,s)\end{aligned}$$

On en déduit en prenant la borne inférieure du second membre, que $\mathrm{T}_\varphi(t) \leqslant \mathrm{T}_\psi(t) + \|\varphi - \psi\|_\infty$. Les fonctions φ et ψ jouant le même rôle, on obtient pareillement $\mathrm{T}_\psi(t) \leqslant \mathrm{T}_\varphi(t) + \|\varphi - \psi\|_\infty$ et donc

$$|\mathrm{T}_\varphi(t) - \mathrm{T}_\psi(t)| \leqslant \|\varphi - \psi\|_\infty.$$

Cela étant vrai pour tout $t \in [0,1]$, on a

$$\|\mathrm{T}_\varphi - \mathrm{T}_\psi\|_\infty \leqslant \|\varphi - \psi\|_\infty.$$

L'application T est continue pour la norme infinie, puisque lipschitzienne de rapport 1.

4. Nous avons démontré dans la question 2 que, pour $\varphi \in \mathcal{C}([0,1],\mathbb{R})$ et $t \in [0,1]$, on a $\mathrm{T}_\varphi(t) \leqslant \varphi(t)$. On en déduit que, pour tout $n \in \mathbb{N}$, on a $\mathrm{T}_\varphi^{n+1}(t) \leqslant \mathrm{T}_\varphi^n(t)$. La suite $(\mathrm{T}_\varphi^n(t))_{n\in\mathbb{N}}$ est donc décroissante, pour tout $t \in [0,1]$. Montrons qu'elle est bornée.

L'image de la fonction nulle par T étant la fonction nulle, il résulte de la question 3 que l'on a, pour $\varphi \in \mathcal{C}([0,1],\mathbb{R})$, $\|\mathrm{T}_\varphi\|_\infty \leqslant \|\varphi\|_\infty$. On en déduit que la suite $\left(\left\|\mathrm{T}_\varphi^n\right\|_\infty\right)$ décroît. En particulier, $\left\|\mathrm{T}_\varphi^n\right\|_\infty \leqslant \|\varphi\|_\infty$ pour tout $n \in \mathbb{N}$ et donc pour tout $t \in [0,1]$, $|\mathrm{T}_\varphi^n(t)| \leqslant \|\varphi\|_\infty$. Pour tout $t \in [0,1]$, la suite $(\mathrm{T}_\varphi^n(t))_{n\in\mathbb{N}}$ est bornée ; elle est donc convergente puisqu'elle est décroissante. La suite (T_φ^n) converge donc simplement sur l'intervalle $[0,1]$. Notons f sa limite. Cette fonction est bien continue sur $[0,1]$: en effet, prenons $\varepsilon > 0$. Dans la première question, on a prouvé l'existence de $\eta > 0$ tel que pour tout $t, t' \in [0,1]$ avec $|t - t'| \leqslant \eta$ et toute application ψ bornée sur $[0,1]$, $|\mathrm{T}_\psi(t) - \mathrm{T}_\psi(t')| \leqslant \varepsilon$. En particulier, si $|t - t'| \leqslant \eta$, pour tout $n \geqslant 0$, $|\mathrm{T}_\varphi^n(t) - \mathrm{T}_\varphi^n(t')| \leqslant \varepsilon$. En faisant tendre n vers l'infini, il vient $|f(t) - f(t')| \leqslant \varepsilon$. La fonction f est bien dans $\mathcal{C}([0,1],\mathbb{R})$.

Montrons que f est un point fixe de T. On a, par définition de T, pour $(s,t) \in [0,1]^2$,

$$\mathrm{T}^n(\varphi)(t) = \mathrm{T}(\mathrm{T}^{n-1}(\varphi))(t) \leqslant \mathrm{T}^{n-1}(\varphi)(s) + \mathrm{H}(t,s).$$

En faisant tendre n vers $+\infty$, on obtient, pour tout $(s,t) \in [0,1]^2$,

$$f(t) \leqslant f(s) + \mathrm{H}(t,s).$$

Or nous avons vu dans la question 2 que c'est la condition pour que f soit un point fixe de T. L'application f est donc un point fixe de T. En particulier, il résulte de la première question que $f = \mathrm{T}(f)$ est continue sur $[0,1]$. Il reste à démontrer que la convergence est uniforme. Cela résulte du théorème de Dini (voir l'exercice précédent) appliquée à la suite $(-\mathrm{T}^n(\varphi))$ sur le segment $[0,1]$. ◁

2.32. Suite de fonctions lipschitziennes sur un compact

1. Soit $[a,b]$ un segment de $\mathbb{R}$ et $(f_n)_{n\geqslant 0}$ une suite de fonctions c-lipschitziennes de $[a,b]$ dans $\mathbb{R}$ qui converge simplement vers une fonction f. Montrer que la convergence est uniforme sur $[a,b]$.

2. Plus généralement soit K un compact d'un espace vectoriel normé et $(f_n)_{n\geqslant 0}$ une suite de fonctions c-lipschitziennes de K dans $\mathbb{R}$ qui converge simplement vers $f : \mathrm{K} \to \mathbb{R}$. Montrer que la convergence est uniforme sur K.

(École normale supérieure)

▷ **Solution.**

1. L'idée est de se ramener à un ensemble fini de points, le caractère c-lipschitzien de toutes les fonctions f_n permettant de contrôler ce qui se passe ailleurs. Plus précisément, soit $\varepsilon > 0$ fixé et une subdivision $\mathrm{S} = (x_0 = a < x_1 < \cdots < x_p = b)$ du segment $[a,b]$ de pas inférieur à $\dfrac{\varepsilon}{3c}$. Puisque la suite $(f_n)_{n\geqslant 0}$ converge simplement vers f, on peut trouver un rang N tel que pour tout $n \geqslant \mathrm{N}$ et tout $k \in [\![0,p]\!]$, $|f_n(x_k) - f(x_k)| \leqslant \dfrac{\varepsilon}{3}$. Soit $x \in [a,b]$ un point quelconque et i tel que $x \in [x_i, x_{i+1}]$. Pour tout $n \geqslant \mathrm{N}$ on a

$$\begin{aligned} |f_n(x) - f(x)| &\leqslant |f_n(x) - f_n(x_i)| + |f_n(x_i) - f(x_i)| + |f(x_i) - f(x)| \\ &\leqslant c|x - x_i| + \frac{\varepsilon}{3} + c|x_i - x| \\ &\leqslant \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon \end{aligned}$$

car la fonction f est clairement aussi c-lipschitzienne. On a donc la majoration $\|f_n - f\|_\infty \leqslant \varepsilon$ pour tout $n \geqslant \mathrm{N}$ ce qui prouve le résultat.

2. L'idée est exactement la même : la précompacité de K permet de trouver un nombre fini $x_0, \ldots, x_p$ de points de K tels que les boules ouvertes $\mathrm{B}\left(x_i, \dfrac{\varepsilon}{3c}\right)$ recouvrent K. Cela se montre facilement par l'absurde

(voir le tome analyse 3 pour de nombreux compléments sur la notion de précompacité). Soit $\varepsilon > 0$. On choisit N tel que pour tout $n \geqslant N$ et tout $k \in [\![0, p]\!]$, $|f_n(x_k) - f(x_k)| \leqslant \dfrac{\varepsilon}{3}$ et la fin de la preuve est absolument identique à celle de la question 1. ◁

Ce résultat est utilisé dans la solution de l'exercice suivant.

2.33. Convergence uniforme de suites de fonctions convexes

Soit $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions convexes sur le segment $[a, b]$ qui converge simplement vers f. Montrer que la convergence est uniforme sur tout compact de $]a, b[$.

(École polytechnique)

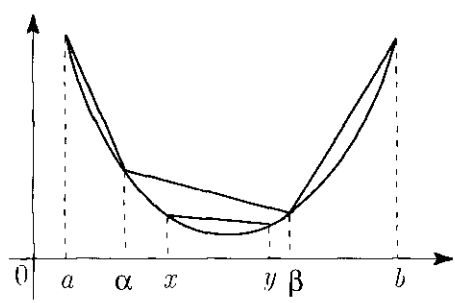
▷ **Solution.**

L'idée est d'utiliser la convexité pour contrôler les variations de f entre deux points, afin de passer de la convergence en un nombre fini de points à une convergence uniforme sur tout un segment $[\alpha, \beta] \subset]a, b[$. Plus précisément, nous allons écrire l'inégalité des pentes croissantes : pour une fonction convexe g sur $[a, b]$, et tout $x < y < z$ dans $[a, b]$, on a donc

$$\frac{g(x) - g(y)}{x - y} \leqslant \frac{g(x) - g(z)}{x - z} \leqslant \frac{g(y) - g(z)}{y - z}.$$

Fixons $\alpha < \beta$ dans $]a, b[$. Pour tout $n \in \mathbb{N}$, et tout $x < y$ dans $[\alpha, \beta]$, on a alors

$$\frac{f_n(a) - f_n(\alpha)}{a - \alpha} \leqslant \frac{f_n(x) - f_n(y)}{x - y} \leqslant \frac{f_n(b) - f_n(\beta)}{b - \beta}.$$



Les quantités $\dfrac{f_n(a) - f_n(\alpha)}{a - \alpha}$ et $\dfrac{f_n(b) - f_n(\beta)}{b - \beta}$ convergent respectivement vers $\dfrac{f(a) - f(\alpha)}{a - \alpha}$ et $\dfrac{f(b) - f(\beta)}{b - \beta}$. Elles sont en particulier bornées. Il

existe donc M réel strictement positif tel que

$$-\mathrm{M} \leqslant \frac{f_n(a) - f_n(\alpha)}{a - \alpha} \leqslant \frac{f_n(x) - f_n(y)}{x - y} \leqslant \frac{f_n(b) - f_n(\beta)}{b - \beta} \leqslant \mathrm{M}.$$

Par conséquent, toutes les fonctions f_n sont M-lipschitziennes sur $[\alpha, \beta]$. On se reportera alors à l'exercice précédent pour en déduire la convergence uniforme des f_n vers f sur $[\alpha, \beta]$. ◁

Les trois exercices suivants sont difficiles et nécessitent la mise en œuvre du procédé d'extraction diagonale dû à Cantor.

2.34. Théorème d'Ascoli

Soit (f_n) une suite d'applications continues de $\mathbb{R}$ dans $\mathbb{R}$ et A une partie de $\mathbb{R}$.

1. On suppose que,

$$\forall x \in \mathrm{A}, \quad \exists \mathrm{M}_x \in \mathbb{R}, \quad \forall n \in \mathbb{N}, \quad |f_n(x)| \leqslant \mathrm{M}_x.$$

Montrer qu'il existe une suite extraite de (f_n) qui converge simplement sur A lorsque A est fini ou dénombrable.

2. On ajoute l'hypothèse suivante : pour tout $\varepsilon > 0$, et tout $x \in \mathbb{R}$, il existe $\eta > 0$ tel que

$$\forall n \in \mathbb{N}, \quad \forall y \in \mathbb{R}, \quad |x - y| \leqslant \eta \longrightarrow |f_n(x) - f_n(y)| \leqslant \varepsilon$$

(on parle *d'équicontinuité* de la suite (f_n)).

Montrer qu'il existe une suite extraite de (f_n) qui converge simplement sur $\mathbb{R}$ et que la convergence est uniforme sur tout compact.

3. Réciproquement, si la suite (f_n) converge uniformément sur tout compact, montrer qu'elle est uniformément bornée sur tout compact et équicontinue.

(École normale supérieure)

▷ **Solution.**

1. Supposons A fini, $\mathrm{A} = \{a_1, a_2, \ldots, a_p\}$. La suite $(f_n(a_1))$ est bornée par hypothèse. D'après le théorème de Bolzano-Weierstrass, on peut en extraire une sous-suite convergente. Soit donc φ_1 une extraction telle que la sous-suite $(f_{\varphi_1(n)}(a_1))$ soit convergente, vers un réel ℓ_1. La suite $(f_{\varphi_1(n)}(a_2))$ est bornée et on peut choisir une extraction φ_2 de telle manière que la suite $(f_{\varphi_1 \circ \varphi_2(n)}(a_2))$ converge, vers un réel ℓ_2. On réitère ce procédé pour construire par récurrence des extractions φ_r telles que

$$\lim_{n\to+\infty} f_{\varphi_1\circ\cdots\circ\varphi_p(n)}(a_i) - \ell_i \in \mathbb{R}.$$

On pose $\psi(n) = \varphi_1 \circ \cdots \circ \varphi_n(n)$. Pour $1 \leqslant i \leqslant p$, la suite $(f_{\psi(n)}(a_i))$ est une suite extraite de $(f_{\varphi_1\circ\ldots\circ\varphi_n(n)}(a_i))$ qui converge donc vers ℓ_i. La suite $(f_{\psi(n)})$ converge simplement sur A.

• Lorsque $A = \{u_p\}_{p\geqslant 1}$ est infini dénombrable, on définit comme précédemment les extractions φ_i et on pose $\psi(n) = \varphi_1 \circ \cdots \circ \varphi_n(n)$. Alors, pour tout $p \geqslant 1$, la suite $(f_{\psi(n)}(a_p))_{n\geqslant p}$ est une suite extraite de $(f_{\varphi_1\circ\cdots\circ\varphi_p(n)}(a_p))_{n\geqslant p}$, qui converge vers ℓ_p. En effet, $n-1 < n \leqslant \varphi(n)$, donc

$$\varphi_{p+1} \circ \cdots \circ \varphi_{n-1}(n-1) < \varphi_{p+1} \circ \cdots \circ \varphi_{n-1} \circ \varphi_n(n)$$

et $f_{\psi(n)}(a_p) = f_{\varphi_1\circ\cdots\circ\varphi_p(\varphi_{p+1}\circ\cdots\circ\varphi_n(n))}(a_p)$ converge vers ℓ_p. La suite $(f_{\psi(n)})$ converge simplement sur A.

Cette construction d'une sous-suite convergente, extraite d'une famille dénombrable de suites, porte le nom de procédé diagonal de Cantor. On en trouvera une application dans l'exercice suivant.

2. Prenons $A = \mathbb{Q}$. D'après la question précédente, il existe une extraction ψ telle que la suite $(f_{\psi(n)})$ converge simplement. Notons pour $n \geqslant 0$, $g_n = f_{\psi(n)}$. La suite (g_n) est équicontinue et converge simplement sur $\mathbb{Q}$.

• Soit $\varepsilon > 0$ et $x \in \mathbb{R}$. Il existe $\eta > 0$ tel que si $|y - x| \leqslant \eta$, alors $|g_n(x) - g_n(y)| \leqslant \varepsilon$ pour tout $n \in \mathbb{N}$. Montrons que la suite $(g_n(x))$ converge à l'aide du critère de Cauchy. Soit q un rationnel avec $|x-q| \leqslant \eta$. Alors pour n et n' entiers naturels,

$$\begin{aligned}|g_n(x) - g_{n'}(x)| &\leqslant |g_n(x) - g_n(q)| + |g_n(q) - g_{n'}(q)| + |g_{n'}(q) - g_{n'}(x)|\\ &\leqslant 2\varepsilon + |g_n(q) - g_{n'}(q)|.\end{aligned}$$

Or, la suite $(g_n(q))$ étant convergente, elle est de Cauchy : il existe n_0 entier naturel tel que si $n \geqslant n_0$ et $n' \geqslant n_0$, $|g_n(q) - g_{n'}(q)| \leqslant 3\varepsilon$. Dans ces conditions, la suite $((g_n(x))$ est de Cauchy, donc elle converge puisque $\mathbb{R}$ est complet.

Ainsi, la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement sur $\mathbb{R}$.

• Montrons que la convergence est uniforme sur tout compact de $\mathbb{R}$. Soit K un compact de $\mathbb{R}$. Prenons $\varepsilon > 0$. Il existe alors pour tout $x \in \mathbb{R}$, un réel strictement positif η_x tel que

$$\forall n \in \mathbb{N}, \forall y \in \mathbb{R},\ |x - y| \leqslant \eta_x \Longrightarrow |g_n(x) - g_n(y)| \leqslant \varepsilon.$$

En passant à la limite dans la dernière inégalité, on a pour $|x - y| \leqslant \eta_x$, $|g(x) - g(y)| \leqslant \varepsilon$. Les ouverts $]x-\eta_x, x+\eta_x[$ recouvrent le compact K. Par le théorème de Borel-Lebesgue, on peut en extraire un sous-recouvrement fini : il existe $x_1, \ldots, x_p$ tel que

$$\mathrm{K} \subset \bigcup_{i=1}^{p}]x_i - \eta_{x_i}, x_i + \eta_{x_i}[.$$

Il existe un rang n_0 tel si $n \geqslant n_0$, on a pour chaque i,

$$|g_n(x_i) - g(x_i)| \leqslant \varepsilon.$$

Si $x \in \mathrm{K}$, on peut trouver i tel que $x \in]x_i - \eta_{x_i}, x_i + \eta_{x_i}[$. Si $n \geqslant n_0$, on a

$$\begin{aligned} |g_n(x) - g(x)| &\leqslant |g_n(x) - g_n(x_i)| + |g_n(x_i) - g(x_i)| + |g(x_i) - g(x)| \\ &\leqslant 3\varepsilon. \end{aligned}$$

Il y a donc convergence uniforme de la suite $(g_n)_{n\in\mathbb{N}}$ sur K.

3. Supposons que $(f_n)_{n\in\mathbb{N}}$ converge uniformément sur tout compact vers f. La fonction f est alors continue. Soit K un compact.

• Pour $g : \mathbb{R} \longrightarrow \mathbb{R}$ continue, posons $\|g\|_\infty = \sup_{x\in\mathrm{K}} |g(x)|$. Par hypothèse, la suite $(\|f_n - f\|_\infty)$ converge vers 0, elle est donc bornée par M. Si $x \in \mathrm{K}$ et $n \in \mathbb{N}$, on a

$$|f_n(x)| \leqslant |f(x)| + |f_n(x) - f(x)| \leqslant \|f\|_\infty + \mathrm{M}.$$

Donc la suite $(f_n)_{n\in\mathbb{N}}$ est uniformément bornée sur K.

• Soit $x \in \mathbb{R}$ et $\varepsilon > 0$. Il existe $\eta > 0$ tel que pour $|x - y| \leqslant \eta$, $|f(x) - f(y)| \leqslant \varepsilon$. Pour tout $n \in \mathbb{N}$, il existe $\eta_n > 0$ que l'on peut prendre inférieur ou égal à η tel que

$$\forall y \in \mathbb{R},\ |x - y| \leqslant \eta_n \Longrightarrow |f_n(x) - f_n(y)| \leqslant \varepsilon.$$

Il y a convergence uniforme de la suite $(f_n)_{n\in\mathbb{N}}$ sur $[x-\eta, x+\eta]$. Prenons $n_0 \in \mathbb{N}$ tel que pour tout $n \geqslant n_0$,

$$\forall y \in [x - \eta, x + \eta],\ |f_n(y) - f(y)| \leqslant \varepsilon.$$

Si on pose $\eta' = \min(\eta, \eta_0, \eta_1, \ldots, \eta_{n_0})$, on a pour $n \geqslant n_0$, et tout y réel tel que $|x - y| \leqslant \eta'$,

$$\begin{aligned} |f_n(x) - f_n(y)| &\leqslant |f_n(x) - f(x)| + |f(x) - f(y)| + |f_n(y) - f(y)| \\ &\leqslant 3\varepsilon \end{aligned}$$

Si $n \leqslant n_0$, on a pour y réel tel que $|x - y| \leqslant \eta'$, $|f_n(x) - f_n(y)| \leqslant \varepsilon \leqslant 3\varepsilon$. La suite $(f_n)_{n\in\mathbb{N}}$ est donc équicontinue. ◁

Ce théorème permet de caractériser les parties compactes dans l'espace des fonctions continues muni de la norme de la convergence uniforme[1] *: pour qu'une partie $\mathcal{H}$ de $\mathcal{C}([a,b],\mathbb{R})$ muni de la norme $\|\ \|_\infty$ soit compacte, il faut et suffit qu'elle soit fermée bornée et équicontinue.*

1. Voir par exemple SCHWARTZ (L.), *Analyse hilbertienne*, Hermann, 1979, p. 129.

L'exercice suivant traite d'un cas particulier du théorème d'Ascoli. En effet, une suite de fonctions M*-lipschitziennes, constitue clairement une famille équicontinue.*

2.35. Un cas particulier du théorème d'Ascoli

Soit $M \geqslant 0$ et $(f_n)_{n\geqslant 0}$ une suite de fonctions de classe $\mathcal{C}^1$ sur $[0,1]$ vérifiant $|f_n(x)| + |f_n'(x)| \leqslant M$ pour tout $x \in [0,1]$. Montrer qu'il existe une suite extraite de $(f_n)_{n\geqslant 0}$ qui converge uniformément.

(**École normale supérieure**)

Solution.

On va dans un premier temps construire une sous-suite de $(f_n)_{n\in\mathbb{N}}$ qui converge simplement sur les rationnels de $[0,1]$. Pour cela, on applique le procédé diagonal de Cantor : notons $(r_p)_{p\in\mathbb{N}}$ les éléments de $\mathbb{Q}\cap[0,1]$. Par hypothèse la suite $(f_n(r_0))_{n\geqslant 0}$ est bornée. On peut donc en extraire une sous-suite convergente $(f_{\varphi_0(n)}(r_0))_{n\geqslant 0}$. De la suite bornée $(f_{\varphi_0(n)}(r_1))_{n\geqslant 0}$ on peut également extraire une sous-suite convergente $(f_{\varphi_0\circ\varphi_1(n)}(r_1))_{n\geqslant 0}$. Bien entendu la suite $(f_{\varphi_0\circ\varphi_1(n)}(r_0))_{n\geqslant 0}$ converge toujours. On itère ce procédé. On construit donc une suite $(\varphi_p)_{p\geqslant 0}$ de fonctions strictement croissantes de $\mathbb{N}$ dans $\mathbb{N}$ telle que pour tout entier p, et tout $k \in [\![0,p]\!]$, la suite

$$f_{\varphi_0\circ\varphi_1\circ\cdots\circ\varphi_n(n)}(r_k)$$

converge. Si on pose $g_n = f_{\varphi_0\circ\varphi_1\circ\cdots\circ\varphi_n(n)}$, on a une suite extraite de $(f_n)_{n\in\mathbb{N}}$ telle que pour tout k, la suite $(g_n(r_k))$ converge (on pourra se reporter à la première question de l'exercice précédent pour les détails).

Nous allons prouver que la suite $(g_n)_{n\in\mathbb{N}}$ répond au problème. Tout d'abord, en montrant qu'elle converge simplement sur $[0,1]$, puis de manière uniforme. Les fonctions g_n sont de classe $\mathcal{C}^1$ et en particulier, elles sont M-lipschitziennes. Prenons $x \in [0,1]$ et vérifions que $(g_n(x))_{n\in\mathbb{N}}$ est une suite réelle de Cauchy. Soit $\varepsilon > 0$. On choisit $q \in [0,1]$ rationnel tel que $|x-q| \leqslant \varepsilon$. Pour tout $n, m \in \mathbb{N}$, on a

$$\begin{aligned}|g_n(x) - g_m(x)| &\leqslant |g_n(x) - g_n(q)| + |g_n(q) - g_m(q)| + |g_m(q) - g_m(x)| \\ &\leqslant 2M\varepsilon + |g_n(q) - g_m(q)|.\end{aligned}$$

Comme $(g_n(q))_{n\in\mathbb{N}}$ converge, il existe $n_0 \in \mathbb{N}$ tel que si $n, m \geqslant n_0$, $|g_n(q) - g_m(q)| \leqslant \varepsilon$. Dans ces conditions, si $n, m \geqslant n_0$, on a la majoration $|g_n(x) - g_m(x)| \leqslant (2M+1)\varepsilon$: la suite $(g_n(x))_{n\in\mathbb{N}}$ est bien de Cauchy, et $\mathbb{R}$ étant complet, la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement sur $[0,1]$ vers une fonction que nous noterons g.

Le fait que la convergence est uniforme est classique et découle de ce que les fonctions g_n sont toutes M-lipschitziennes. Le lecteur trouvera cela dans la question 1 de l'exercice 2.32 qui avait été posée comme première question lors de l'oral. ◁

Voici un autre situation où apparaît le procédé diagonal de Cantor.

2.36. Théorème de sélection de Helly

1. Soit $E \subset \mathbb{R}$ dénombrable et (f_n) une suite de fonctions de E dans $\mathbb{R}$ telle que pour tout n et tout x de E, $|f_n(x)| \leqslant 1$. Montrer qu'il existe une sous-suite de (f_n) qui converge simplement vers une fonction $f : E \to \mathbb{R}$.

2. Soit (f_n) une suite de fonctions croissantes de $\mathbb{R}$ dans $[-1, 1]$. Montrer qu'il existe une sous-suite de (f_n) qui converge simplement vers une fonction $f : \mathbb{R} \to [-1, 1]$.

(École normale supérieure)

▷ **Solution.**

1. On applique le procédé diagonal de Cantor comme dans l'exercice précédent avec $(r_p)_{p\in\mathbb{N}}$ la suite des éléments de E pour construire une sous-suite de (f_n) simplement convergente.

2. Utilisons le résultat de la question précédente en prenant pour E l'ensemble des nombres rationnels. On peut supposer que notre suite (f_n) converge simplement sur $\mathbb{Q}$ vers une fonction $f : \mathbb{Q} \to \mathbb{R}$. Cette fonction est croissante sur $\mathbb{Q}$. Soit x un nombre réel. Supposons que f admette la même limite ℓ à gauche et à droite en x. Soit $\varepsilon > 0$. Il existe donc $\eta > 0$ tel que pour $t \in [x - \eta, x + \eta]$ (t rationnel), $|f(t) - \ell| \leqslant \varepsilon$. On choisit deux rationnels α et β tels que $\alpha \in [x - \eta, x]$ et $\beta \in [x, x + \eta]$. On a pour tout entier n, $f_n(\alpha) \leqslant f_n(x) \leqslant f_n(\beta)$. Pour n assez grand, $f_n(\beta) \leqslant f(\beta) + \varepsilon$ et $f_n(\alpha) \geqslant f(\alpha) - \varepsilon$. Il en résulte que pour n assez grand,

$$\ell - 2\varepsilon \leqslant f(\alpha) - \varepsilon \leqslant f_n(x) \leqslant f(\beta) + \varepsilon \leqslant \ell + 2\varepsilon.$$

Ainsi, la suite $(f_n(x))_{n\geqslant 0}$ converge vers ℓ.

L'ensemble D des points x tels que $\lim\limits_{x^-} f < \lim\limits_{x^+} f$ est au plus dénombrable : en effet, pour chaque $x \in D$, on peut choisir un rationnel q_x compris entre $\lim\limits_{x^-} f$ et $\lim\limits_{x^+} f$. L'application $x \in D \longmapsto q_x$ est clairement injective puisque f est croissante. Il s'ensuit donc que D est dénombrable. On peut donc, par une nouvelle application de la question 1 extraire de la suite (f_n) une sous-suite qui converge simplement sur D. Bien entendu cette sous-suite converge encore sur $\mathbb{R} \setminus D$. ◁

Dans l'exercice suivant on montre que pour une suite de fonctions prises dans un sous-espace de dimension finie de $\mathcal{C}^0([0,1],\mathbb{R})$ la convergence simple implique automatiquement la convergence uniforme.

2.37. Suite de fonctions d'un sous-espace de dimension finie de $\mathcal{C}^0([0,1])$

1. Soit $d \in \mathbb{N}^*$ et $(P_n)_{n\geqslant 0}$ une suite de $\mathbb{R}_d[X]$. Montrer que $(P_n)_{n\geqslant 0}$ converge uniformément sur $[0,1]$ si et seulement si $(P_n)_{n\geqslant 0}$ converge simplement sur $[0,1]$.

2. Soit V un sous-espace de dimension finie de $\mathcal{C}^0([0,1],\mathbb{R})$ et $(f_n)_{n\geqslant 0}$ une suite de V qui converge simplement sur $[0,1]$. Montrer que la convergence est uniforme sur $[0,1]$.

(École normale supérieure)

▷ **Solution.**

1. On sait qu'un polynôme de $\mathbb{R}_d[X]$ est déterminé de manière unique par ses valeurs en $d+1$ points. Soit $x_0, x_1, \ldots, x_d$ des éléments distincts de $[0,1]$. En notant, pour $0 \leqslant i \leqslant d$, le i-ième polynôme interpolateur en ces points, défini par $L_i = \dfrac{\prod\limits_{k\neq i}(X - x_k)}{\prod\limits_{k\neq i}(x_i - x_k)}$, on obtient $P_n = \sum\limits_{i=0}^{d} P_n(x_i)L_i$.

Supposons que $(P_n)_{n\geqslant 0}$ converge simplement vers une fonction f. Pour tout $x \in [0,1]$, on a

$$f(x) = \lim_{n\to+\infty} P_n(x) = \sum_{i=0}^{d} \lim_{n\to+\infty} P_n(x_i)L_i(x) = \sum_{i=0}^{d} f(x_i)L_i(x).$$

On en déduit que, pour tout $x \in [0,1]$,

$$|P_n(x) - f(x)| \leqslant \sum_{i=0}^{d} |P_n(x_i) - f(x_i)|\,|L_i(x)| \leqslant \sum_{i=0}^{d} |P_n(x_i) - f(x_i)|\,\|L_i\|_\infty$$

et donc

$$\|P_n - f\| \leqslant \sum_{i=0}^{d} |P_n(x_i) - f(x_i)|\,\|L_i\|_\infty.$$

Ainsi $(\|P_n - f\|)$ converge vers 0 et la suite $(P_n)_{n\geqslant 0}$ converge uniformément vers f. La réciproque est évidemment triviale.

2. On s'inspire de la question précédente et on va montrer qu'un élément de V est déterminé par ses valeurs en un nombre fini de points. Notons p la dimension de V et soit $(g_1, g_2, \ldots, g_p)$ une base de V. C'est

un résultat classique qu'il existe une suite $(x_1, x_2, \dots, x_p)$ de points de $[0,1]$ telle que $\det(g_i(x_j))_{1\leqslant i,j\leqslant d} \neq 0$ (le lecteur en trouvera une preuve dans l'exercice 1.16 du tome algèbre 2 et une autre dans l'exercice 6.28 du tome algèbre 1).

Exprimons chaque fonction f_n dans la base de V. Pour tout $n \in \mathbb{N}$, il existe $(a_{n,1}, a_{n,2}, \dots, a_{n,d})$ tel que $f_n = \sum_{i=1}^{p} a_{n,i} g_i$. En écrivant cette égalité aux points x_j, on obtient

$$\forall i \in [\![1, p]\!], \quad f_n(x_j) = \sum_{i=1}^{p} a_{n,i} g_i(x_j),$$

que l'on considère comme un système linéaire de p équations dont les inconnues sont $a_{n,1}, a_{n,2}, \dots, a_{n,d}$. Le déterminant $\det(g_i(x_j))_{1\leqslant i,j\leqslant n}$ du système n'est pas nul, donc il possède une solution unique, donnée par les formules de Cramer. En notant $C_1, C_2, \dots, C_p$ les colonnes de la matrice $(g_i(x_j))_{1\leqslant i,j\leqslant n}$ et X_n le vecteur-colonne $\begin{pmatrix} f_n(x_1) \\ \vdots \\ f_n(x_p) \end{pmatrix}$, on obtient, pour tout $1 \leqslant i \leqslant p$,

$$a_{n,i} = \frac{\det(C_1, \dots, C_{i-1}, X_n, C_{i+1}, \dots, C_p)}{\det(g_i(x_j))_{1\leqslant i,j\leqslant n}}.$$

À i fixé, la suite $(a_{n,i})$ apparaît comme une combinaison linéaire des suites $(f_n(x_j))$ $(1 \leqslant j \leqslant p)$. Chacune de ces suites est convergente, car (f_n) converge simplement, donc la suite $(a_{n,i})$ converge. On note a_i sa limite. De l'égalité

$$\forall x \in [0,1], \quad f_n(x) = \sum_{i=1}^{d} a_{n,i} g_i(x),$$

on tire par passage à la limite,

$$\forall x \in [0,1], \quad f(x) = \sum_{i=1}^{d} a_i g_i(x).$$

On en déduit, pour tout $x \in [0,1]$,

$$|f_n(x) - f(x)| \leqslant \sum_{i=1}^{d} |a_{n,i} - a_i| |g_i(x)| \leqslant \sum_{i=1}^{d} |a_{n,i} - a_i| \, \|g_i\|_\infty$$

et donc

$$\|f_n(x) - f(x)\|_\infty \leqslant \sum_{i=1}^{d} |a_{n,i} - a_i| \, \|g_i\|_\infty.$$

Cela montre que $(\|f_n(x) - f(x)\|_\infty)$ tend vers 0 et donc que la suite (f_n) converge uniformément vers f sur $[0,1]$. ◁

On a montré de plus que la limite f appartient à $\mathbb{R}_d[X]$ dans le premier cas et à V dans le second. Ce n'est pas étonnant puisqu'un sous-espace de dimension finie d'un espace vectoriel normé, ici $(\mathcal{C}^0([0,1],\mathbb{R}), \|\ \|_\infty)$, est fermé.

Le dernier exercice du chapitre est plus anecdotique que les précédents mais pas facile pour autant.

2.38. Critère de convergence uniforme

Soit $f_n : [0,1] \longrightarrow \mathbb{R}$ une suite de fonctions telle que pour toute suite convergente $(x_n)_{n\in\mathbb{N}}$ de $[0,1]$ la suite $(f_n(x_n))_{n\in\mathbb{N}}$ converge.

1. Montrer que la suite (f_n) converge simplement sur $[0,1]$ vers une fonction f.

2. Prouver la continuité de f.

3. En déduire que la convergence de (f_n) vers f est uniforme.

(École polytechnique)

Solution.

1. Soit $x \in [0,1]$. En appliquant l'hypothèse avec la suite constante égale à x on peut affirmer que la suite $f_n(x)$ converge. On note $f(x)$ sa limite. Notons alors que pour toute suite (x_n) qui converge vers x la suite $f_n(x_n)$ converge vers $f(x)$. En effet, pour le montrer il suffit d'appliquer l'hypothèse avec la suite (y_n) définie par $y_{2p} = x_{2p}$ et $y_{2p+1} = x$:

$$\lim_{n\to+\infty} f_n(y_n) = \lim_{p\to+\infty} f_{2p}(y_{2p}) = \lim_{p\to+\infty} f_{2p}(x_{2p}) = \lim_{n\to+\infty} f_n(x_n)$$

et

$$\lim_{n\to+\infty} f_n(y_n) = \lim_{p\to+\infty} f_{2p+1}(y_{2p+1}) = \lim_{p\to+\infty} f_{2p+1}(x) = f(x).$$

L'unicité de la limite permet de conclure.

2. Raisonnons par l'absurde et supposons qu'il existe $a \in [0,1]$ tel que f ne soit pas continue en a. Il existe alors $\varepsilon > 0$ et une suite (x_n) qui converge vers a telle que $|f(x_n) - f(a)| \geqslant \varepsilon$ pour tout n. Par hypothèse, pour tout n, on a

$$\lim_{k\to+\infty} f_k(x_n) = f(x_n).$$

On peut donc choisir $\varphi(0) \in \mathbb{N}$ tel que $|f_{\varphi(0)}(x_0) - f(x_0)| \leqslant \dfrac{\varepsilon}{2}$. On peut choisir ensuite $\varphi(1) > \varphi(0)$ tel que $|f_{\varphi(1)}(x_1) - f(x_1)| \leqslant \dfrac{\varepsilon}{2}$. On construit

ainsi par récurrence une fonction $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ strictement croissante en choisissant $\varphi(p) > \varphi(p-1)$ tel que

$$|f_{\varphi(p)}(x_p) - f(x_p)| \leqslant \frac{\varepsilon}{2}.$$

Il en résulte pour tout $p \in \mathbb{N}$ que $|f_{\varphi(p)}(x_p) - f(a)| \geqslant \dfrac{\varepsilon}{2}$. Posons pour $n \in \mathbb{N}$, $\mathrm{X}_n = x_p$ si n s'écrit $n = \varphi(p)$ avec $p \in \mathbb{N}$ et $\mathrm{X}_n = a$ sinon. On vérifie aisément que (X_n) converge vers a. D'autre part, on a pour tout $p \in \mathbb{N}$,

$$|f_{\varphi(p)}(\mathrm{X}_{\varphi(p)}) - f(a)| = |f_{\varphi(p)}(x_p) - f(a)| \geqslant \frac{\varepsilon}{2},$$

si bien que la suite $(f_{\varphi(p)}(\mathrm{X}_{\varphi(p)}))_{p\in\mathbb{N}}$ ne converge pas vers $f(a)$. Comme c'est une suite extraite de la suite $(f_p(\mathrm{X}_p))_{p\in\mathbb{N}}$, cette dernière ne peut converger vers $f(a)$, ce qui est contraire à l'hypothèse.

On conclut donc que la fonction f est continue.

3. Raisonnons par l'absurde et supposons que la convergence n'est pas uniforme. Il existe alors une sous-suite de $(\|f - f_n\|_\infty)_{n\in\mathbb{N}}$ ne convergeant pas vers 0 : autrement dit, il existe $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$ strictement croissante avec pour tout $n \in \mathbb{N}$

$$\sup_{x\in[0,1]} |f(x) - f_{\varphi(n)}(x)| = \|f - f_{\varphi(n)}\|_\infty \geqslant \varepsilon.$$

Par définition de la borne supérieure, il existe $x_n \in [0,1]$ tel que

$$|f(x_n) - g_n(x_n)| \geqslant \frac{\varepsilon}{2} \text{ où } g_n = f_{\varphi(n)}.$$

Comme $[0,1]$ est compact, on peut extraire de $(x_n)_{n\in\mathbb{N}}$ une sous-suite $(x_{\psi(n)})_{n\geqslant 0}$ (avec $\psi : \mathbb{N} \longrightarrow \mathbb{N}$ strictement croissante) convergente. Notons x la limite. On a

$$|f(x_{\psi(n)}) - g_{\psi(n)}(x_{\psi(n)})| \geqslant \frac{\varepsilon}{2}.$$

La suite $(g_{\psi(n)})$ vérifiant les mêmes conditions que $(f_n)_{n\in\mathbb{N}}$ et f étant continue, en faisant tendre n vers l'infini, on a $|f(x) - f(x)| \geqslant \varepsilon$ ce qui est absurde. On conclut donc que la convergence est uniforme. ◁

Chapitre 3

Séries entières

La série géométrique est connue en substance depuis l'Antiquité. Au XVII^e^ *siècle, on doit à Gregory le développement de* $\tan x$ *et* $\arctan x$ *et à Mercator celui de* $\ln(1+x)$*. Vers 1665, Newton en même temps qu'il découvre le calcul infinitésimal, détermine de nombreux développements en séries entières, entre autres celui de* $(1-x^2)^{\frac{1}{2}}$*, par des méthodes indirectes, partant de problèmes d'intégration. À la suite de Newton, les mathématiciens ne cherchèrent plus à éviter les sommes infinies, comme avaient pu le faire les Grecs. Au* XVIII^e^ *siècle, après qu'a été énoncée la formule de Taylor, la plupart des mathématiciens en viennent à ne plus considérer que des fonctions analytiques, c'est-à-dire sommes au voisinage de chaque point d'une série entière. Mais cette approche est essentiellement formelle et ignore très souvent les problèmes de convergence. Dans* Introductio in analysin infinitorum *(1748), Euler développe cette « analyse algébrique » qui étend aux domaines des séries et des produits infinis les règles usuelles de manipulations algébriques. Les séries entières apparaissent alors comme des polynômes à une infinité de termes.*

Les fondements de la théorie des fonctions entières sont posés par Cauchy dans son Cours d'Analyse *(1821). Il fait une étude très précise de la convergence d'une série entière, met en évidence le rayon de convergence. En 1831, il établit la formule connue sous le nom de formule intégrale de Cauchy, donne une expression intégrale des coefficients du développement d'où il déduit les inégalités qui portent son nom. Riemann et Weierstrass développeront cette théorie en s'appuyant sur la vision géométrique, prenant comme champ d'étude les fonctions de la variable complexe dérivables (contrairement aux fonctions de la variable réelle, les fonctions* $\mathbb{C}$*-dérivables sont automatiquement analytiques, autrement dit développables en série de Taylor au voisinage de tout point).*

Les premiers exercices concernent le calcul du rayon de convergence de séries entières et l'étude de la convergence sur le disque de convergence.

3.1. Rayon de convergence (1)

Soit (a_n) une suite de nombres complexes et R le rayon de convergence de la série entière $\sum a_n z^n$. Déterminer le rayon de convergence de la série entière $\sum a_n^2 z^n$.

(École polytechnique)

▷ **Solution.**

Si la limite du rapport $\left|\dfrac{a_n}{a_{n+1}}\right|$ existe, elle est égale à R. Dans ces conditions, la limite de $\left|\dfrac{a_n^2}{a_{n+1}^2}\right| = \left|\dfrac{a_n}{a_{n+1}}\right|^2$ existe aussi et vaut R^2 qui est alors le rayon de convergence de $\sum a_n^2 z^n$. Il est raisonnable de penser que dans le cas général, le rayon de convergence est toujours R^2.

Prenons $r < \mathrm{R}^2$ un réel positif. Alors $\sqrt{r} < \mathrm{R}$ et ainsi la suite de terme général

$$|a_n^2 r^n| = \left[|a_n|\left(\sqrt{r}\right)^n\right]^2$$

est bornée. Par conséquent, le rayon de convergence R' de $\sum a_n^2 z^n$ est supérieur ou égal à r. Comme cela vaut pour tout $r < \mathrm{R}^2$ on en déduit que $\mathrm{R}' \geqslant \mathrm{R}^2$ (et $\mathrm{R}' = +\infty$ si $\mathrm{R} = +\infty$). Prenons maintenant un réel $r > \mathrm{R}^2$. La suite de terme général $a_n\left(\sqrt{r}\right)^n$ ne tend pas vers 0 et il en va donc de même de la suite de terme général $\left[a_n\left(\sqrt{r}\right)^n\right]^2 = a_n^2 r^n$. Ainsi on a $r \geqslant \mathrm{R}'$. Comme cela vaut pour tout $r > \mathrm{R}^2$, on a $\mathrm{R}' \leqslant \mathrm{R}^2$ et on conclut que $\boxed{\mathrm{R}' = \mathrm{R}^2}$. ◁

On montre de même que, pour tout entier $p \geqslant 2$, la série $\sum a_n^p z^n$ a pour rayon de convergence R^p.

3.2. Rayon de convergence (2)

Déterminer le rayon de convergence de $\sum e^{n \sin n} z^n$.

(École polytechnique)

▷ **Solution.**

Désignons par R le rayon de convergence de cette série. Pour tout réel $r > 0$, on a la majoration

$$0 \leqslant e^{n \sin n} r^n \leqslant e^n r^n = (er)^n.$$

Donc si $r < \dfrac{1}{e}$, la série de terme général $e^{n \sin n} r^n$ converge, ce qui montre que $\mathrm{R} \geqslant \dfrac{1}{e}$.

Prenons maintenant un réel $r > \dfrac{1}{e}$. Puisque 1 est valeur d'adhérence de la suite $(\sin n)_{n \geqslant 0}$ (on pourra se reporter à l'exercice 1.13 du premier tome d'analyse), il existe une suite $(k_n)_{n \in \mathbb{N}}$ d'entiers divergente vers l'infini telle que la limite de $\sin k_n$ soit égale à 1. Il s'ensuit que

$$\ln(e^{k_n \sin k_n} r^{k_n}) = k_n(\sin k_n + \ln r) \underset{n \to \infty}{\sim} k_n(1 + \ln r) \xrightarrow[n \to +\infty]{} +\infty.$$

La limite de $e^{n \sin n} r^n$ ne peut donc être nulle et $\mathrm{R} \leqslant \dfrac{1}{e}$.

Conclusion. On a $\boxed{\mathrm{R} = \dfrac{1}{e}}$. ◁

Le rayon de convergence R *de* $\sum a_n z^n$ *est donné dans tous les cas par la formule d'Hadamard* $\dfrac{1}{\mathrm{R}} = \limsup\limits_{n \to +\infty} \sqrt[n]{a_n} \in \overline{\mathbb{R}}$ *(le lecteur consultera le premier tome d'analyse page 78 pour la définition des limites supérieures et inférieures). Ici, on obtient immédiatement*

$$\frac{1}{\mathrm{R}} = \limsup_{n \to +\infty} \sqrt[n]{e^{n \sin n}} = \limsup_{n \to +\infty} e^{\sin n} = e \quad \textit{et} \quad \mathrm{R} = \frac{1}{e},$$

puisque 1 *est la plus grande valeur d'adhérence de la suite* $(\sin n)_{n \geqslant 0}$.

3.3. Rayon de convergence (3)

Déterminer le rayon de convergence de $\sum \dfrac{z^n}{\sin(n\pi\sqrt{3})}$.

(École normale supérieure)

▷ **Solution.**

On observe pour commencer que $\sin(n\pi\sqrt{3}) \neq 0$ pour tout entier n car $\sqrt{3}$ est irrationnel.

Pour $z = 1$ la série diverge car $|\sin(n\pi\sqrt{3})| \leqslant 1$ pour tout n. Donc le rayon de convergence R vérifie $\mathrm{R} \leqslant 1$.

Pour obtenir une minoration de R on va majorer $|\sin(n\pi\sqrt{3})|^{-1}$ c'est-à-dire minorer $|\sin(n\pi\sqrt{3})|$. On introduit l'entier le plus proche de $n\sqrt{3}$ pour se ramener au sinus d'un réel de $\left[-\dfrac{\pi}{2}, \dfrac{\pi}{2}\right]$. Soit $p_n \in \mathbb{N}$ l'unique entier tel que $p_n - \dfrac{1}{2} \leqslant n\sqrt{3} < p_n + \dfrac{1}{2}$ et $\varepsilon_n = n\sqrt{3} - p_n \in [-1/2, 1/2[$. On a

$$|\sin(n\pi\sqrt{3})| = |\sin(\pi\varepsilon_n)| \geqslant \frac{2}{\pi}|\pi\varepsilon_n| = 2|\varepsilon_n|,$$

car $|\sin x| \geqslant \frac{2}{\pi}|x|$ si $x \in [-\pi/2, \pi/2]$. En passant à l'inverse cela donne,

$$\frac{1}{\left|\sin(n\pi\sqrt{3})\right|} \leqslant \frac{1}{2|\varepsilon_n|} = \frac{1}{2|n\sqrt{3} - p_n|} = \frac{n\sqrt{3} + p_n}{2|3n^2 - p_n^2|} \leqslant \frac{n\sqrt{3} + p_n}{2}.$$

Ce dernier terme est un $O(n)$. Il en résulte que, pour tout nombre complexe z vérifiant $|z| < 1$, la série $\sum \frac{z^n}{\sin(n\pi\sqrt{3})}$ converge.

Conclusion. On a $\boxed{\mathrm{R} = 1}$. ◁

3.4. Rayon de convergence (4)

Déterminer le rayon de convergence R de la série entière $\sum \left(1 + \frac{(-1)^n}{n}\right)^{n^2} z^n$. Étudier la convergence lorsque $|z| = \mathrm{R}$.

(École polytechnique)

▷ **Solution.**

Pour tout $r > 0$, on a la majoration

$$0 \leqslant \sqrt[n]{\left(1 + \frac{(-1)^n}{n}\right)^{n^2} r^n} \leqslant \left(1 + \frac{1}{n}\right)^n r \xrightarrow[n \to +\infty]{} re.$$

La règle de Cauchy assure que la série entière converge pour $|z| = r < \frac{1}{e}$, ce qui montre que $\mathrm{R} \geqslant \frac{1}{e}$. En revanche, si $r > \frac{1}{e}$, la suite correspondant aux indices pairs vérifie

$$\sqrt[2n]{\left(1 + \frac{1}{2n}\right)^{4n^2} r^{2n}} = \left(1 + \frac{1}{2n}\right)^{2n} r \xrightarrow[n \to +\infty]{} re > 1,$$

si bien que la suite $\left(1 + \frac{(-1)^n}{n}\right)^{n^2} r^n$ n'est pas bornée. On en conclut que $\mathrm{R} = \frac{1}{e}$. Enfin, si $|z| = \mathrm{R} = \frac{1}{e}$, le terme général de la série entière ne tend pas vers 0. En effet, le module du terme général correspondant à un indice n pair vaut

$$\left(1 + \frac{1}{n}\right)^{n^2} \frac{1}{e^n} = \exp\left(n^2 \ln\left(1 + \frac{1}{n}\right) - n\right) = \exp\left(-\frac{1}{2} + o(1)\right)$$
$$\xrightarrow[n \to +\infty]{} \frac{1}{\sqrt{e}}.$$

Il y a donc divergence de la série entière en tout point du cercle de convergence. ◁

Si une série entière a pour rayon de convergence R, *l'ensemble des points du cercle de convergence* $\{z, |z| = \mathrm{R}\}$ *où la série converge peut aller de l'ensemble vide au cercle entier en passant par des cas nettement plus complexes. Considérons par exemple la série* $\sum n^\alpha z^n$, *dont le rayon de convergence est* 1 *pour tout* α. *Si* $\alpha < -1$, *la série converge normalement sur* $\overline{\mathrm{D}} = \{z \in \mathbb{C}, |z| \leqslant 1\}$. *Si* $\alpha \geqslant 0$, *la série ne converge en aucun point du cercle de convergence, puisque son terme général ne tend pas vers* 0. *Enfin, si* $\alpha \in [-1, 0[$, *la série diverge pour* $z = 1$, *mais on peut montrer, grâce à une transformation d'Abel, qu'elle converge pour* $z = e^{i\theta}$ *si* $\theta \in]0, 2\pi[$.

L'exercice suivant, qui étudie la convergence d'une série entière en certains points du cercle de convergence, est nettement plus difficile que les énoncés précédents.

3.5. Étude d'une série entière sur le cercle de convergence

On considère la série entière $\sum c_n x^n$ où, pour $n \geqslant 1$, $c_n = \dfrac{p_n}{n}$, p_n étant le nombre d'entiers $k \geqslant 1$ tels que $k!$ divise n.

1. Donner le rayon de convergence de cette série.

2. Étudier la convergence de la série $\sum c_n x^n$ lorsque $x = e^{2i\pi r}$, avec r rationnel, puis lorsque $x = e^{2i\pi\theta}$. On pourra montrer que

$$\sum_{i=1}^{n} c_i x^i = \sum_{\substack{k \geqslant 1 \\ k! \leqslant n}} \frac{1}{k!} \left(\sum_{\ell=1}^{\frac{n}{k!}} \frac{x^{k!\,\ell}}{\ell} \right).$$

(École polytechnique)

Solution.

1. On a, pour $n \geqslant 1$, $1 \leqslant p_n \leqslant n$, donc $\dfrac{1}{n} \leqslant c_n \leqslant 1$. Les séries $\sum \dfrac{x^n}{n}$ et $\sum x^n$ ayant pour rayon de convergence 1, on en déduit que le rayon de convergence de $\sum c_n x^n$ est 1.

Dans la question suivante, on étudie la convergence de la série en certains points du cercle de convergence.

2. Pour $n \in \mathbb{N}^*$, on pose $S_n = \sum_{i=1}^{n} c_i x^i$. On peut écrire $p_i = \sum_{k! \mid i} 1$, pour tout $i \in \mathbb{N}^*$. On en déduit que

$$S_n = \sum_{i=1}^{n} \sum_{k!|i} \frac{x^i}{i} = \sum_{\substack{(i,k,\ell) \\ 1 \leqslant i \leqslant n \\ i = k!\ell}} \frac{x^i}{i} = \sum_{\substack{(k,\ell) \\ 1 \leqslant k! \leqslant n}} \frac{x^{k!\ell}}{k!\ell}$$

$$= \sum_{1 \leqslant k! \leqslant n} \frac{1}{k!} \left(\sum_{\ell=1}^{\frac{n}{k!}} \frac{x^{k!\ell}}{\ell} \right) = \sum_{1 \leqslant k! \leqslant n} \frac{1}{k!} \sigma_{k,n},$$

où $\sigma_{k,n} = \sum_{\ell=1}^{\frac{n}{k!}} \frac{x^{k!\ell}}{\ell}$.

a. Supposons que r est rationnel : $r = \frac{p}{q}$, avec $q \in \mathbb{N}^*$ et que $x = e^{2i\pi r}$. On obtient $x^{k!} = e^{2i\pi \frac{pk!}{q}}$. Dès que $k \geqslant q$, on a $x^{k!} = 1$. L'ensemble des k tels que $x^{k!} \neq 1$ est donc un ensemble fini I. Pour tous les k tels que $x^{k!} = 1$, la série $\sum \frac{x^{k!\ell}}{\ell}$ diverge. Cela fait pressentir que $\sum c_n x^n$ diverge. Nous le justifierons en montrant que $|S_n|$ tend vers $+\infty$ quand n tend vers $+\infty$.

Si $k \notin I$, alors $\sigma_{k,n} = \sum_{\ell=1}^{\frac{n}{k!}} \frac{1}{\ell} > 0$. On en déduit que, pour $n \geqslant q!$, on a

$$\sum_{\substack{1 \leqslant k! \leqslant n \\ k! \notin I}} \frac{1}{k!} \sigma_{k,n} \geqslant \frac{1}{q!} \sigma_{q,n} \geqslant \frac{1}{q!} \sum_{\ell=1}^{\frac{n}{q!}} \frac{1}{\ell}.$$

Cette quantité tend vers $+\infty$ avec n, puisque q est fixé et que la série harmonique $\sum \frac{1}{\ell}$ diverge. Pour montrer que $|S_n|$ tend vers $+\infty$, il suffit de montrer que $\sum_{\substack{1 \leqslant k! \leqslant n \\ k! \in I}} \frac{1}{k!} \sigma_{k,n}$ est borné (indépendamment de n).

Pour $k \in I$, posons $u = x^{k!}$. On a donc $u \neq 1$. Pour tout $p \in \mathbb{N}^*$, il s'agit de majorer $A_p = \sum_{\ell=1}^{p} \frac{u^\ell}{\ell}$. On utilise une transformation d'Abel. On pose, pour $p \in \mathbb{N}$, $B_p = \sum_{\ell=0}^{p} u^\ell = \frac{1 - u^{p+1}}{1 - u}$. On a $|B_p| \leqslant \frac{2}{|u-1|}$. On écrit alors

$$A_p = \sum_{\ell=1}^{p} \frac{B_\ell - B_{\ell-1}}{\ell} = \sum_{\ell=1}^{p} \frac{B_\ell}{\ell} - \sum_{\ell=0}^{p-1} \frac{B_\ell}{\ell+1} = \sum_{\ell=1}^{p-1} B_\ell \left(\frac{1}{\ell} - \frac{1}{\ell+1} \right) + \frac{B_p}{p} - B_0.$$

De la majoration de B_p, on déduit que

$$|A_p| \leqslant \frac{2}{|u-1|} \left(\sum_{\ell=1}^{p-1} \left(\frac{1}{\ell} - \frac{1}{\ell+1} \right) + \frac{1}{p} + 1 \right) \leqslant \frac{4}{|u-1|}$$

et en particulier $|\sigma_{k,n}| \leqslant \frac{4}{|u-1|} \leqslant \frac{4}{|x^{k!} - 1|}$. Chaque terme $\frac{1}{k!} \sigma_{k,n}$, pour $k \in I$, est majoré indépendamment de n. L'ensemble I étant fini,

on en déduit que $\sum_{\substack{1\leqslant k!\leqslant n\\ k\in\mathbb{N}}} \frac{1}{k!}\sigma_{k,n}$ est borné, quand n varie. Ceci permet de conclure que $\lim_{n\to+\infty} |S_n| = +\infty$.

b. Nous allons montrer que pour $x = e^{2i\pi e}$ la série converge. On reprend les notations utilisées dans le premier cas. On a, pour tout $k \in \mathbb{N}^*$, $x^{k!} \neq 1$, car $e \notin \mathbb{Q}$. On en déduit, toujours en posant $u = x^{k!}$, que, pour tout $n \geqslant 1$, $|\sigma_{k,n}| \leqslant \frac{1}{|u-1|}$. Il s'agit de trouver un majorant de $\frac{1}{|u-1|}$. La formule de Taylor-Lagrange à l'ordre k appliquée à la fonction exponentielle entre 0 et 1 montre qu'il existe $c \in]0,1[$ tel que

$$e = \sum_{j=0}^{k} \frac{1}{j!} + \frac{e^c}{(k+1)!}.$$

On note n_k l'entier $k! \sum_{j=0}^{k} \frac{1}{j!}$ et on pose $r_k = k!e - n_k = \frac{e^c}{k+1}$. On a donc $\frac{1}{k+1} \leqslant r_k \leqslant \frac{e}{k+1}$, ce qui montre en particulier que $\lim_{k\to+\infty} r_k = 0$. On en déduit que $x^{k!} = e^{2i\pi e k!} = e^{2i\pi(n_k+r_k)} = e^{2i\pi r_k}$. Il en résulte que $|\sigma_{k,n}| \leqslant \frac{4}{|e^{2i\pi r_k}-1|} = \frac{2}{|\sin \pi r_k|}$. De la minoration $\sin x \geqslant \frac{2}{\pi}x$ si $x \in \left[0, \frac{\pi}{2}\right[$, on déduit que, pour k assez grand ($k \geqslant k_0$), on a, pour tout $n \in \mathbb{N}^*$, $|\sigma_{k,n}| \leqslant \frac{1}{r_k} \leqslant k+1$ et donc $\left|\frac{\sigma_{k,n}}{k!}\right| \leqslant \frac{k+1}{k!}$.

On note, d'autre part, que pour tout $k \geqslant 1$, $\sigma_{k,n}$ a une limite quand n tend vers $+\infty$. En effet, en reprenant la méthode employée précédemment, on obtient, pour tout $k \geqslant 1$ et pour $1 \leqslant n \leqslant m$,

$$\begin{aligned}\sigma_{k,m} - \sigma_{k,n} &= \sum_{\ell=n_0+1}^{m_0} \frac{u^\ell}{\ell} = \sum_{\ell=n_0+1}^{m_0} \frac{B_\ell - B_{\ell-1}}{\ell} \\ &= \frac{B_{m_0}}{m_0} + \sum_{n_0+1}^{m_0-1} B_\ell \left(\frac{1}{\ell} - \frac{1}{\ell+1}\right) - \frac{B_{n_0}}{n_0+1},\end{aligned}$$

où on a posé $m_0 = E\left(\frac{m}{k!}\right)$ et $n_0 = E\left(\frac{n}{k!}\right)$. On en déduit que

$$|\sigma_{k,m} - \sigma_{k,n}| \leqslant \frac{2}{|u-1|} \cdot \frac{2}{n_0+1} \leqslant \frac{4}{|u-1|} \cdot \frac{k!}{n}.$$

On en déduit que, pour tout $k \geqslant 1$, la suite $(\sigma_{k,n})_{n\in\mathbb{N}^*}$ est de Cauchy et donc convergente. On note σ_k sa limite. En fait, on a $\sigma_k = \sum_{\ell=1}^{+\infty} \frac{x^{k!\ell}}{\ell}$. Il ne s'agit plus que d'intervertir une limite et une sommation. Il résulte

de ce qui précède, en faisant tendre n vers $+\infty$, que, pour $k \geqslant k_0$, on a $\left|\frac{\sigma_k}{k}\right| \leqslant \frac{k+1}{k!}$. On en déduit que $\sum \frac{\sigma_k}{k!}$ converge. Soit $\varepsilon > 0$ et k_1 tel que $\sum_{k=k_1+1}^{+\infty} \frac{k+1}{k!} \leqslant \varepsilon$. On a, pour $n \geqslant k_1$,

$$\left|S_n - \sum_{k=1}^{+\infty} \frac{1}{k!}\sigma_k\right| \leqslant \sum_{k=1}^{k_1} \frac{1}{k!}|\sigma_k - \sigma_{k,n}| + \sum_{k=k_1+1}^{n} \frac{1}{k!}|\sigma_{k,n}| + \sum_{k=k_1+1}^{+\infty} \frac{1}{k!}|\sigma_k|$$
$$\leqslant \sum_{k=1}^{k_1} \frac{1}{k!}|\sigma_k - \sigma_{k,n}| + 2\varepsilon.$$

Pour n assez grand, cela sera inférieur à 3ε. On en déduit que la suite (S_n), c'est-à-dire que la série $\sum c_n x^n$, converge. On a, de plus,

$$\sum_{n=1}^{+\infty} c_n x^n = \sum_{k=1}^{+\infty} \left(\frac{1}{k!} \sum_{\ell=1}^{+\infty} \frac{x^{k!\,\ell}}{\ell}\right). \quad \lhd$$

Le lecteur sait que la convergence d'une série entière est uniforme sur tout compact inclus dans le disque de convergence. La seconde question de l'exercice qui suit donne un théorème d'Abel qui caractérise les cas où la série converge uniformément sur tout le disque de convergence. Dans la première question on démontre le classique théorème d'Abel-Dirichlet. La résolution de ce type de questions repose sur des transformations d'Abel.

3.6. Le théorème d'Abel-Dirichlet

1. Soit $\sum a_n x^n$ une série entière de rayon de convergence égal à $R \in]0, +\infty[$. On suppose que la série $\sum a_n R^n$ converge. Montrer que la convergence de la série $\sum a_n x^n$ est uniforme sur le segment $[0, R]$.

En déduire que

$$\sum_{n=0}^{+\infty} a_n R^n = \lim_{x \to R} \sum_{n=0}^{+\infty} a_n x^n.$$

2. Soit $\sum a_n z^n$ une série entière de rayon de convergence égal à 1. On pose $\mathrm{D} = \{z \in \mathbb{C},\ |z| < 1\}$ et $\mathrm{U} = \{z \in \mathbb{C},\ |z| = 1\}$. Montrer que les assertions suivantes sont équivalentes :

(*i*) la série converge uniformément sur $\overline{\mathrm{D}}$;

(*ii*) la série converge uniformément sur D ;

(*iii*) la série converge uniformément sur U.

(**École normale supérieure**)

Solution.

1. Soit $\varepsilon > 0$. On note $\rho_{\mathrm{N}} = \sum_{n=\mathrm{N}}^{+\infty} a_n \mathrm{R}^n$ pour tout $\mathrm{N} \geqslant 0$. Nous allons majorer le reste de la série entière à l'aide d'une transformation d'Abel : pour $x \in [0, \mathrm{R}[$, on a

$$\begin{aligned}
\sum_{n=\mathrm{N}}^{+\infty} a_n x^n &= \sum_{n=\mathrm{N}}^{+\infty} a_n \mathrm{R}^n \left(\frac{x}{\mathrm{R}}\right)^n = \sum_{n=\mathrm{N}}^{+\infty} (\rho_n - \rho_{n+1}) \left(\frac{x}{\mathrm{R}}\right)^n \\
&= \sum_{n=\mathrm{N}}^{+\infty} \rho_n \left(\frac{x}{\mathrm{R}}\right)^n - \sum_{n=\mathrm{N}}^{+\infty} \rho_{n+1} \left(\frac{x}{\mathrm{R}}\right)^n \\
&\quad \text{(ces séries convergent car } (\rho_n) \text{ est bornée)} \\
&= \sum_{n=\mathrm{N}}^{+\infty} \rho_n \left(\frac{x}{\mathrm{R}}\right)^n - \sum_{n=\mathrm{N}+1}^{+\infty} \rho_n \left(\frac{x}{\mathrm{R}}\right)^{n-1} \\
&= \rho_{\mathrm{N}} \left(\frac{x}{\mathrm{R}}\right)^{\mathrm{N}} + \sum_{n=\mathrm{N}+1}^{+\infty} \rho_n \left(\left(\frac{x}{\mathrm{R}}\right)^n - \left(\frac{x}{\mathrm{R}}\right)^{n-1} \right).
\end{aligned}$$

Il existe un rang n_0 tel que si $n \geqslant n_0$, $|\rho_n| \leqslant \varepsilon$. On obtient alors pour $\mathrm{N} \geqslant n_0$ la majoration

$$\left| \sum_{n=\mathrm{N}}^{+\infty} a_n x^n \right| \leqslant \varepsilon + \sum_{n=\mathrm{N}+1}^{+\infty} \varepsilon \underbrace{\left(\left(\frac{x}{\mathrm{R}}\right)^{n-1} - \left(\frac{x}{\mathrm{R}}\right)^n \right)}_{\geqslant 0} \leqslant \varepsilon + \varepsilon \left(\frac{x}{\mathrm{R}}\right)^{\mathrm{N}} \leqslant 2\varepsilon.$$

Cette majoration qui est encore vraie pour $x = \mathrm{R}$ traduit la convergence uniforme de la série entière sur $[0, \mathrm{R}]$.

Il en résulte que la fonction $x \longmapsto \sum_{n=0}^{+\infty} a_n x^n$ est continue sur $[0, \mathrm{R}]$ en tant que limite uniforme de fonctions continues. On obtient alors la limite demandée.

Ce théorème peut s'avérer commode pour calculer les sommes de séries entières au point x égal au rayon de convergence. Par exemple, les séries entières $\sum_{n=1}^{+\infty} \frac{(-1)^n}{n} x^n$ et $\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1} x^{2n}$ convergent en $x = 1$

par simple application du théorème spécial de convergence des séries alternées. Il en résulte que $\displaystyle\sum_{n=1}^{+\infty}\frac{(-1)^n}{n} = -\lim_{x\to 1^-}\ln(1+x) = -\ln 2$ *et de même que* $\displaystyle\sum_{n=0}^{+\infty}\frac{(-1)^n}{2n+1} = \lim_{x\to 1^-}\arctan x = \frac{\pi}{4}$.

2. • L'implication $(i)\Longrightarrow(ii)$ est triviale car D est contenu dans $\overline{\mathrm{D}}$.

• Montrons que $(ii)\Longrightarrow(iii)$. On va utiliser le critère de Cauchy. Soit $\varepsilon>0$. Par hypothèse, il existe un entier N tel que pour tout $n\geqslant \mathrm{N}$, tout $p\geqslant 0$ et tout $z\in\mathrm{D}$,

$$\left|\sum_{k=n}^{n+p} a_k z^k\right| \leqslant \varepsilon.$$

Par continuité du polynôme et passage des inégalités larges à la limite, cette majoration reste vraie pour tout point z de U. Le critère de Cauchy uniforme assure alors que la série converge uniformément sur U.

• Montrons que $(iii)\Longrightarrow(i)$. Introduisons naturellement pour $\mathrm{N}\geqslant 0$, $\theta\in\mathbb{R}$, $\rho_{\mathrm{N}}(\theta)=\displaystyle\sum_{n=\mathrm{N}}^{+\infty} a_n e^{in\theta}$, le reste de la série sur U, dont on sait qu'il converge uniformément vers 0 sur U.

Soit $\varepsilon>0$. Pour $z\in\mathrm{D}$, écrivons $z=re^{i\theta}$ avec $0\leqslant r<1$ et $\theta\in\mathbb{R}$. Majorons le reste à l'aide d'une transformation d'Abel. Pour $\mathrm{N}\geqslant 0$, on obtient,

$$\begin{aligned}
\sum_{n=\mathrm{N}}^{+\infty} a_n z^n &= \sum_{n=\mathrm{N}}^{+\infty} a_n e^{in\theta} r^n = \sum_{n=\mathrm{N}}^{+\infty}\Big(\rho_n(\theta)-\rho_{n+1}(\theta)\Big)r^n\\
&= \sum_{n=\mathrm{N}}^{+\infty}\rho_n(\theta)r^n - \sum_{n=\mathrm{N}}^{+\infty}\rho_{n+1}(\theta)r^n\\
&= \sum_{n=\mathrm{N}}^{+\infty}\rho_n(\theta)r^n - \sum_{n=\mathrm{N}+1}^{+\infty}\rho_n(\theta)r^{n-1}\\
&= \sum_{n=\mathrm{N}+1}^{+\infty}\rho_n(\theta)(r^n-r^{n-1}) + \rho_{\mathrm{N}}(\theta)r^{\mathrm{N}}.
\end{aligned}$$

On obtient, grâce à l'inégalité triangulaire,

$$\begin{aligned}
\left|\sum_{n=\mathrm{N}}^{+\infty} a_n z^n\right| &\leqslant \sum_{n=\mathrm{N}+1}^{+\infty}|\rho_n(\theta)|(r^{n-1}-r^n) + |\rho_{\mathrm{N}}(\theta)|\\
&\leqslant \sup_{n\geqslant \mathrm{N}}|\rho_n(\theta)|\underbrace{\sum_{n=\mathrm{N}+1}^{+\infty}(r^{n-1}-r^n)}_{=r^{\mathrm{N}}\leqslant 1} + |\rho_{\mathrm{N}}(\theta)| \leqslant 2\sup_{n\geqslant \mathrm{N}}|\rho_n(\theta)|.
\end{aligned}$$

Cette inégalité reste vraie si $z\in\mathrm{U}$.

Par hypothèse, il existe un rang n_0 tel que si $n \geqslant n_0$, pour tout $\theta \in \mathbb{R}$, $|\rho_n(\theta)| \leqslant \varepsilon$. Donc, pour $N \geqslant n_0$ et pour tout $z \in \overline{D}$, $\left|\sum_{n=N}^{+\infty} a_n z^n\right| \leqslant 2\varepsilon$.

La convergence est donc uniforme sur $\overline{D}$. ◁

La dernière implication peut se traiter beaucoup plus rapidement à l'aide du principe du maximum (voir exercice 3.40) en utilisant le critère de Cauchy uniforme : en effet, si la tranche de Cauchy $\left|\sum_{k=n}^{n+p} a_k z^k\right|$ *est majorée par* $\varepsilon > 0$ *pour tout* z *de* U, *le principe du maximum permet de dire que cette majoration est réalisée sur tout le disque* $\overline{D}$.

Les prochains exercices concernent le calcul de la somme de quelques séries entières.

3.7. Calcul de la somme d'une série entière

Calculer $\sum_{n=1}^{+\infty} \frac{\sin(na)x^n}{n}$, où a et x sont des nombres réels.

(École polytechnique)

Solution.

Les coefficients de la série proposée sont des fonctions impaires et 2π-périodiques de a. On peut donc supposer que $a \in [0, \pi]$. Lorsque a vaut 0 ou π, la série est clairement nulle. Dans la suite de la solution, on supposera donc que $a \in]0, \pi[$.

La série $\sum_{n=1}^{+\infty} \frac{\sin(na)x^n}{n}$ a clairement un rayon de convergence supérieur ou égal à 1. Pour x réel tel que $|x| < 1$, posons

$$G_a(x) = \sum_{n=1}^{+\infty} \frac{\sin(na)x^n}{n}.$$

Le n du dénominateur est gênant. Pour s'en débarrasser, on dérive G_a. On obtient

$$G_a'(x) = \sum_{n=1}^{+\infty} \sin(na)x^{n-1} = \operatorname{Im}\left(\sum_{n=1}^{+\infty} e^{ina}x^{n-1}\right)$$
$$= \operatorname{Im}\left(e^{ia}\sum_{n=0}^{+\infty}(e^{ia}x)^n\right) = \operatorname{Im}\left(\frac{e^{ia}}{1-e^{ia}x}\right) = \operatorname{Im}\left(\frac{1}{e^{-ia}-x}\right).$$

Comme G_a est de classe $\mathcal{C}^1$ et nulle en 0, on a

$$G_a(x) = \int_0^x \operatorname{Im}\left(\frac{1}{e^{-ia}-t}\right) dt.$$

On va simplement calculer cette intégrale. Il vient

$$\begin{aligned} \mathrm{G}_a(x) &= \int_0^x \mathrm{Im}\left(\frac{1}{\cos a - t - i\sin a}\right) \mathrm{d}t = \int_0^x \frac{\sin a}{(\cos a - t)^2 + \sin^2 a}\,\mathrm{d}t \\ &= \left[-\arctan\left(\frac{\cos a - t}{\sin a}\right)\right]_0^x \\ &= \arctan\left(\frac{x - \cos a}{\sin a}\right) + \arctan(\operatorname{cotan} a). \end{aligned}$$

Or, $\operatorname{cotan} a = \tan\left(\frac{\pi}{2} - a\right)$ et comme a est dans l'intervalle $]0, \pi[$, $\frac{\pi}{2} - a$ est dans $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ et $\arctan(\operatorname{cotan} a) = \frac{\pi}{2} - a$. On a donc, pour $a \in]0, \pi[$, et $|x| < 1$,

$$\boxed{\mathrm{G}_a(x) = \arctan\left(\frac{x - \cos a}{\sin a}\right) - a + \frac{\pi}{2}}.$$

Les symétries exposées au début de la solution permettent alors d'écrire $\mathrm{G}_a(x)$ pour tout réel a et tout x de $]-1, 1[$. ◁

Une transformation d'Abel montre que la série $\sum \frac{\sin na}{n}$ *(obtenue pour* $x = 1$*) converge. Le théorème d'Abel-Dirichlet (voir page 178) montre alors que* $\mathrm{G}_a(x)$ *tend vers* $\mathrm{G}_a(1)$ *lorsque* x *tend vers* 1*. On en déduit donc que, pour* $0 < a < \pi$*,*

$$\sum_{n=1}^{+\infty} \frac{\sin na}{n} = \arctan\left(\frac{1 - \cos a}{\sin a}\right) - a + \frac{\pi}{2} = \frac{\pi - a}{2}$$

car $\frac{1 - \cos a}{\sin a} = \tan\frac{a}{2}$*. Ce résultat est classique et se retrouve facilement à l'aide d'un développement en série de Fourier.*

Il n'y a pas de calculs supplémentaires pour trouver $\mathrm{G}_a(-1)$ *: il suffit de noter que* $\mathrm{G}_a(-1) = \mathrm{G}_{a+\pi}(1)$*.*

3.8. Calcul de la somme de séries de fonctions

Déterminer le domaine de définition et calculer

$$f(x) = \sum_{n=1}^{+\infty} \frac{1}{n^2}\left(x^n + \left(\frac{-x}{1-x}\right)^n\right) \text{ et } g(x) = \sum_{n=1}^{+\infty} \frac{1}{n^2}(x^n + (1-x)^n).$$

(École polytechnique)

▷ **Solution.**

Le calcul de $f(x)$ et $g(x)$ se ramène à celui de la somme de la série entière $\sum \frac{x^n}{n^2}$, dont le rayon de convergence est 1. On pose, pour

$x \in [-1,1]$, $\varphi(x) = \sum_{n=1}^{+\infty} \frac{x^n}{n^2}$. Cette fonction est continue sur $[-1,1]$ et dérivable sur $]-1,1[$. On a, pour $x \in]-1,1[$ et $x \neq 0$,

$$\varphi'(x) = \sum_{n=1}^{+\infty} \frac{x^{n-1}}{n} = -\frac{\ln(1-x)}{x}$$

et de plus $\varphi'(0) = 1$.

• Recherchons le domaine de définition de f. Posons, pour $x \neq 1$ et $n \geqslant 1$,

$$u_n(x) = \frac{1}{n^2}\left(x^n + \left(\frac{-x}{1-x}\right)^n\right).$$

Si n est pair, on a

$$|u_n(x)| = \frac{1}{n^2}\left(|x|^n + \left|\frac{x}{1-x}\right|^n\right).$$

Pour qu'il y ait convergence de la série définissant f, il est donc nécessaire que $|x| \leqslant 1$ et que $\left|\frac{x}{1-x}\right| \leqslant 1$. C'est le cas si et seulement si $-1 \leqslant x \leqslant \frac{1}{2}$. Réciproquement, lorsque $-1 \leqslant x \leqslant \frac{1}{2}$, on a $\left|\frac{x}{1-x}\right| \leqslant 1$ et $|u_n(x)|$ est majoré par $\frac{1}{n^2}\left(|x|^n + \left|\frac{x}{1-x}\right|^n\right)$ qui est le terme général d'une série convergente. La fonction f est donc définie sur $\left[-1, \frac{1}{2}\right]$.

Pour $x \in \left[-1, \frac{1}{2}\right]$, on peut écrire $f(x) = \varphi(x) + \varphi\left(\frac{-x}{1-x}\right)$. La fonction f est donc dérivable sur $\left]-1, \frac{1}{2}\right[$ et sa dérivée vaut

$$f'(x) = -\frac{\ln(1-x)}{x} - \frac{-1}{(1-x)^2}\frac{\ln\left(1+\frac{x}{1-x}\right)}{\frac{-x}{1-x}} = \frac{\ln(1-x)}{(1-x)},$$

après simplifications. Comme $f(0) = 0$, il vient

$$\boxed{f(x) = -\frac{1}{2}\ln^2(1-x)}.$$

Cette formule reste valable en -1 et $\frac{1}{2}$ par continuité de φ et donc de f.

• Pour que $g(x)$ soit défini il apparaît nécessaire, en regardant les termes d'indices pair, que $|x| \leqslant 1$ et $|1-x| \leqslant 1$, autrement dit que $x \in [0,1]$. Il est alors clair que le segment $[0,1]$ est l'ensemble de définition de g.

Pour $x \in [0,1]$, on a $g(x) = \varphi(x) + \varphi(1-x)$. La fonction g est donc continue sur $[0,1]$ et dérivable sur $]0,1[$, avec pour $x \in]0,1[$,

$$g'(x) = -\frac{\ln(1-x)}{x} + \frac{\ln(1-(1-x))}{1-x} = -\frac{\ln(1-x)}{x} + \frac{\ln(x)}{1-x}$$

et donc $g(x) = -\ln(x)\ln(1-x) + C$, où C est une constante. Puisque, pour x tendant vers 0, on a

$$\ln(x)\ln(1-x) \sim -x\ln x \xrightarrow[x\to 0]{} 0,$$

on en déduit, par continuité de g, que $C = g(0) = \frac{\pi^2}{6}$. On obtient

$$g(x) = \frac{\pi^2}{6} - \ln(x)\ln(1-x) \text{ pour } x \in]0,1[\text{ et } g(0) = g(1) = \frac{\pi^2}{6}. \quad \triangleleft$$

L'exercice suivant s'intéresse à la fonction obtenue à partir d'une série entière en ne gardant que les termes dont l'indice décrit une suite arithmétique.

3.9. Somme d'une série entière extraite

Soit $\sum a_n z^n$ une série entière de rayon de convergence $R > 0$. On pose $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ pour $|z| < R$. Soit $p \in \mathbb{N}^*$ et $k \in [\![0, p-1]\!]$.

1. Montrer que le rayon de convergence de $\sum a_{np+k} z^{np+k}$ est au moins R.

2. Pour $|z| < R$, exprimer $\sum_{n=0}^{+\infty} a_{np+k} z^{np+k}$ à l'aide de f.

(École polytechnique)

▷ **Solution.**

1. Pour tout $|z| < R$, la suite $(a_n z^n)_{n\geqslant 0}$ est bornée. Soit $K > 0$ tel que $|a_n z^n| \leqslant K$ pour tout $n \in \mathbb{N}$. On a, en particulier, $|a_{np+k} z^{np+k}| \leqslant K$. Donc le rayon de convergence de $\sum a_{np+k} z^{np+k}$ est supérieur ou égal à R.

2. Commençons par regarder le cas $p = 2$. On a clairement

$$f(z) + f(-z) = 2\sum_{n=0}^{+\infty} a_{2n} z^{2n} \text{ et } f(z) - f(-z) = 2\sum_{n=0}^{+\infty} a_{2n+1} z^{2n+1}.$$

Cela donne le résultat pour $p = 2$. Il s'agit de le généraliser.

Puisque -1 et 1 sont les racines carrées de 1, on songe à calculer dans le cas général $f(\omega z)$, où $\omega^p = 1$. Posons $\omega = e^{\frac{2i\pi}{p}}$ et fixons $z \in \mathbb{C}$.

$|z| < 1$. On a

$$f(z) + f(\omega z) + \cdots + f(\omega^{p-1}z) = \sum_{n=0}^{+\infty} a_n(1 + \omega^n + \cdots + (\omega^{p-1})^n)z^n.$$

Or, si $n \in \mathbb{N}$, nous savons que

$$\begin{aligned} 1 + \omega^n + \cdots + (\omega^{p-1})^n &= 1 + \omega^n + (\omega^n)^2 + \cdots + (\omega^n)^{p-1} \\ &= \begin{cases} \dfrac{1-(\omega^n)^p}{1-\omega^n} = \dfrac{1-(\omega^p)^n}{1-\omega^n} = 0 \text{ si } \omega^n \neq 1 \\ p \text{ si } \omega^n = 1 \end{cases} \\ &= \begin{cases} 0 \text{ si } n \notin p\mathbb{N} \\ 1 \text{ si } n \in p\mathbb{N}. \end{cases} \end{aligned}$$

Ainsi, on obtient $f(z) + f(\omega z) + \cdots + f(\omega^{p-1}z) = p\sum_{n=0}^{+\infty} a_{pn}z^{pn}$ et la formule

$$\boxed{\sum_{n=0}^{+\infty} a_{pn}z^{pn} = \frac{f(z) + f(\omega z) + \cdots + f(\omega^{p-1}z)}{p}}.$$

Pour k quelconque dans $[\![0, p-1]\!]$, on calcule

$$\begin{aligned} f(z) + \omega^{-k}f(\omega z) + \cdots + \omega^{-(p-1)k}f(\omega^{p-1}z) &= \sum_{n=0}^{+\infty} a_n\left(\sum_{\ell=0}^{p-1} \omega^{-\ell k + n\ell}\right)z^n \\ &= \sum_{n=0}^{+\infty} a_n\left(\sum_{\ell=0}^{p-1} (\omega^{n-k})^\ell\right)z^n. \end{aligned}$$

Sachant que $\sum_{\ell=0}^{p-1}(\omega^{n-k})^\ell = \begin{cases} p & \text{si } p|(n-k) \text{ i.e. } n \in p\mathbb{N}+k \\ 0 & \text{sinon} \end{cases}$, on obtient

$$\boxed{\sum_{n=0}^{+\infty} a_{pn+k}z^{pn+k} = \frac{f(z) + \omega^{-k}f(\omega z) + \cdots + \omega^{-k(p-1)}f(\omega^{p-1}z)}{p}}. \quad \lhd$$

L'exercice suivant est un cas particulier du précédent, mais nous en donnons une autre solution fondée sur l'utilisation d'une équation différentielle.

3.10. Calcul d'une somme

Calculer $\displaystyle\sum_{n=0}^{+\infty} \frac{1}{(5n)!}$.

(École polytechnique)

▷ **Solution.**

Pour calculer cette somme, introduisons la fonction f définie par $f(x) = \displaystyle\sum_{n=0}^{+\infty} \frac{x^{5n}}{(5n)!}$. Il s'agit de calculer $S = f(1)$. Le rayon de convergence de la série définissant f est infini, donc f est définie et de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$. D'après les propriétés des séries entières, on a, pour tout $x \in \mathbb{R}$,

$$\begin{aligned} f^{(5)}(x) &= \sum_{n=1}^{+\infty} 5n(5n-1)(5n-2)(5n-3)(5n-4)\frac{x^{5n-5}}{(5n)!} \\ &= \sum_{n=1}^{+\infty} \frac{x^{5n-5}}{(5n-5)!} = f(x). \end{aligned}$$

La fonction f est donc solution de l'équation différentielle linéaire à coefficients constants (E) $y^{(5)} = y$. Le polynôme caractéristique associé à cette équation est $X^5 - 1$. Ses racines sont les racines cinquièmes de l'unité que nous noterons λ_1, λ_2, λ_3, λ_4, λ_5. Ces racines étant distinctes, toute solution de (E) s'écrit de manière unique sous la forme $x \longmapsto \displaystyle\sum_{k=1}^{5} a_k e^{\lambda_k x}$, où a_1, a_2, a_3, a_4, a_5 sont des constantes complexes. La fonction f vérifie les conditions $f(0) = 1$ et $f'(0) = f''(0) = f^{(3)}(0) = f^{(4)}(0) = 0$, c'est-à-dire $\displaystyle\sum_{k=1}^{5} a_k\lambda_k = 1$ et $\displaystyle\sum_{k=1}^{5} a_k\lambda_k^i = 0$ pour $1 \leqslant i \leqslant 4$. Sachant que la somme des puissances i-ièmes des racines cinquièmes de l'unité vaut 0 si i n'est pas multiple de 5 et 5 sinon, on en déduit qu'il faut prendre tous les a_k égaux à $\dfrac{1}{5}$ (ce système est de Cramer). On a donc, pour tout $x \in \mathbb{R}$, $f(x) = \dfrac{1}{5}\displaystyle\sum_{k=1}^{5} e^{\lambda_k x}$ et en particulier $S = \dfrac{1}{5}\displaystyle\sum_{k=1}^{5} e^{\lambda_k}$. C'est bien le résultat que donne l'exercice précédent.

On peut préciser ce résultat en déterminant les racines cinquièmes de l'unité. Si λ est une telle racine, différente de 1, elle vérifie l'équation $\lambda^4 + \lambda^3 + \lambda^2 + \lambda + 1 = 0$. C'est une équation réciproque. On divise par λ^2 et on prend comme inconnue auxiliaire $\mu = \lambda + \dfrac{1}{\lambda}$. On obtient $\mu^2 + \mu - 1 = 0$, soit $\mu = \dfrac{-1 + \varepsilon\sqrt{5}}{2}$, où $\varepsilon = \pm 1$. On détermine ensuite les valeurs de λ. On trouve $\lambda = \dfrac{-1 + \varepsilon\sqrt{5}}{4} \pm i\dfrac{\sqrt{10 + 2\varepsilon\sqrt{5}}}{4}$. En regroupant

les racines conjuguées, on trouve finalement

$$\boxed{S = \frac{1}{5}\left(e + 2e^{-\frac{1+\sqrt{5}}{4}} \cos\frac{\sqrt{10-2\sqrt{5}}}{4} + 2e^{\frac{-1+\sqrt{5}}{4}} \cos\frac{\sqrt{10+2\sqrt{5}}}{4}\right)}. \quad \triangleleft$$

On a, plus généralement, pour $p \in \mathbb{N}^*$, $\sum_{n=0}^{+\infty} \frac{1}{(pn)!} = \sum_{k=1}^{p} e^{\lambda_k}$, *où* $\lambda_1, \ldots, \lambda_p$ *sont les racines* p*-ièmes de l'unité (cf. exercice 3.9).*

Dans l'exercice suivant, les séries entières servent d'outil pour le calcul de la somme de certaines séries alternées.

3.11. Calculs de sommes de séries alternées

1. Calculer la somme de la série $\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1}$.

2. Étudier la convergence et calculer la somme de la série de terme général $\frac{(-1)^n}{n+1}\left(\sum_{k=0}^{n} \frac{1}{2k+1}\right)$.

(École polytechnique)

▷ **Solution.**

1. La série $\sum (-1)^n \frac{1}{2n+1}$ converge, car elle vérifie le critère spécial de convergence des séries alternées. Nous allons donner deux solutions pour calculer sa somme.

• Une première idée consiste à introduire la fonction f définie par $f(x) = \sum_{n=0}^{+\infty} (-1)^n \frac{x^{2n+1}}{2n+1}$. Le rayon de convergence de la série entière définissant f est 1. La fonction f est donc définie et dérivable sur $]-1,1[$. On a, pour tout $x \in]-1,1[$,

$$f'(x) = \sum_{n=0}^{+\infty} (-1)^n x^{2n} = \frac{1}{1+x^2}.$$

Puisque de plus, $f(0) = 0$, on en déduit que, pour tout $x \in]-1,1[$, $f(x) = \arctan(x)$. On a vu que f est définie en 1 et nous cherchons justement à calculer $f(1)$. Montrons la continuité de f en 1. Pour tout $x \in [0,1]$, la série définissant $f(x)$ vérifie le critère spécial des séries alternées. On en déduit que le reste $R_n(x) = \sum_{k=n}^{+\infty} (-1)^k \frac{x^{2k+1}}{2k+1}$ vérifie

$$|R_n(x)| \leqslant \frac{x^{2n+1}}{2n+1} \leqslant \frac{1}{2n+1}.$$

La série qui définit f converge donc uniformément sur $[0,1]$ et f qui est limite uniforme de fonctions continues sur $[0,1[$ est donc continue sur $[0,1]$ et en particulier en 1. On conclut donc que

$$\boxed{\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1} = \frac{\pi}{4}}.$$

Notons que l'on pouvait appliquer directement le théorème d'Abel-Dirichlet (voir page 178) pour montrer la continuité de f en 1.

• La seconde solution que nous proposons repose sur le fait que $\dfrac{(-1)^n}{2n+1} = \displaystyle\int_0^1 (-x^2)^n \mathrm{d}x$. On a alors envie d'écrire que

$$\sum_{n=0}^{+\infty} \frac{(-1)^n}{2n+1} = \sum_{n=0}^{+\infty} \int_{[0,1[} (-x^2)^n \mathrm{d}x = \int_{[0,1[} \sum_{n=0}^{+\infty} (-x^2)^n \mathrm{d}x$$
$$= \int_{[0,1[} \frac{\mathrm{d}x}{1+x^2} = [\arctan x]_0^1 = \frac{\pi}{4}.$$

Le seul point de ce calcul qui mérite justification est l'interversion de la sommation et de l'intégration. Elle est légitimée par le théorème de convergence dominée, puisque pour tout N, la fonction $\left|\displaystyle\sum_{n=0}^{N} (-x^2)^n\right|$ est majorée par 1 sur $[0,1[$.

2. Posons, pour tout $n \in \mathbb{N}$, $u_n = \dfrac{1}{n+1} \displaystyle\sum_{k=0}^{n} \frac{1}{2k+1}$. On a

$$u_{n+1} = \frac{1}{n+2} \sum_{k=0}^{n+1} \frac{1}{2k+1} = \frac{1}{n+2} \left((n+1)u_n + \frac{1}{2n+3} \right).$$

On en déduit que $u_{n+1} - u_n = \dfrac{1}{n+2} \left(\dfrac{1}{2n+3} - u_n \right)$. Or u_n est la moyenne de $n+1$ termes supérieurs à $\dfrac{1}{2n+1}$. On a donc $u_n \geqslant \dfrac{1}{2n+1}$ et $u_{n+1} - u_n \leqslant 0$: la suite (u_n) décroît. De plus, on sait que

$$\sum_{k=0}^{n} \frac{1}{2k+1} \leqslant \sum_{k=1}^{2n+1} \frac{1}{k} \underset{n\to\infty}{\sim} \ln(2n+1) \underset{n\to\infty}{\sim} \ln n.$$

On en déduit que $u_n = O\left(\dfrac{\ln n}{n}\right)$ et en particulier $\lim\limits_{n\to+\infty} u_n = 0$. La série $\sum (-1)^n u_n$ converge d'après le critère spécial des séries alternées.

Pour calculer la somme S de cette série, on procède comme dans la première solution de la question 1 et on pose $g(x) = \displaystyle\sum_{n=0}^{+\infty} (-1)^n u_n x^n$. Le

rayon de convergence de la série définissant g est supérieur ou égal à 1, car la suite (u_n) converge vers 0. La série converge en 1 et on vérifie que g est continue sur $[0,1]$; en effet, pour $x \in [0,1]$ la série définissant g vérifie le critère spécial des séries alternées : la suite $(u_n x^n)$ est à termes positifs, décroissante (car produit de suites positives décroissantes) et converge vers 0. Le reste $R'_n(x)$ de la série vérifie, pour tout $x \in [0,1]$, $|R'_n(x)| \leqslant u_n x^n \leqslant u_n$. La série converge donc uniformément sur $[0,1]$ ce qui implique la continuité de la fonction g.

Notons que là encore, le théorème d'Abel-Dirichlet (voir page 178) permet de démontrer directement la continuité de g.

Déterminons maintenant la valeur de $g(x)$. L'expression $(-x)^n \sum_{k=0}^{n} \frac{1}{2k+1}$ fait penser à un produit de Cauchy. Effectivement, on a, pour $x \in]-1,1[$,

$$\left(\sum_{n=0}^{+\infty} \frac{(-x)^n}{2n+1}\right)\left(\sum_{n=0}^{+\infty}(-x)^n\right) = \sum_{n=0}^{+\infty}(n+1)u_n(-x)^n.$$

Compte tenu de la première question, on obtient pour $x \in]-1,1[$,

$$\sum_{n=0}^{+\infty}(n+1)u_n(-x)^n = \frac{\arctan(\sqrt{x})}{\sqrt{x}(1+x)}.$$

En intégrant, il vient, pour tout $x \in]-1,1[$,

$$xg(x) = \sum_{n=0}^{+\infty} u_n(-1)^n x^{n+1} = \int_0^x \frac{\arctan(\sqrt{t})}{\sqrt{t}(1+t)}\,dt.$$

La continuité de g en 1 conduit à

$$S = g(1) = \int_0^1 \frac{\arctan(\sqrt{t})}{\sqrt{t}(1+t)}dt.$$

Le changement de variable $u = \sqrt{t}$ donne

$$S = \int_0^1 \frac{2\arctan u}{1+u^2}\,du = \left[(\arctan u)^2\right]_0^1 = \left(\frac{\pi}{4}\right)^2.$$

On conclut que

$$\boxed{\sum_{n=0}^{+\infty}\left(\frac{(-1)^n}{n+1}\sum_{k=0}^{n}\frac{1}{2k+1}\right) = \frac{\pi^2}{16}}. \quad \triangleleft$$

Le problème inverse du calcul de la somme d'une série est le développement en série entière d'une fonction donnée. Les fonctions

les plus simples qu'on sait être développables en série entière sont les fractions rationnelles. Comme on le redémontre dans l'exercice suivant, toute fraction rationnelle dont 0 n'est pas un pôle est développable en série entière au voisinage de 0, avec un rayon de convergence égal au plus petit module des pôles de f.

3.12. Développement en série entière des fractions rationnelles

1. Soit P et Q deux polynômes à coefficients complexes, avec $Q(0) \neq 0$. Pour $z \in \mathbb{C}$ tel que $Q(z) \neq 0$, on pose $f(z) = \dfrac{P(z)}{Q(z)}$. Montrer que f est développable en série entière au voisinage de 0 et que les coefficients de ce développement vérifient une relation de récurrence linéaire à coefficients constants. Préciser le rayon de convergence.

2. Réciproquement, si la suite de nombres complexes (u_n) vérifie une relation de récurrence linéaire à coefficients constants, montrer que la série entière $\sum u_n z^n$ a un rayon de convergence $R > 0$ et qu'il existe deux polynômes à coefficients complexes P et Q avec $Q(0) \neq 0$ tels que $\sum_{n=0}^{+\infty} u_n z^n = \dfrac{P(z)}{Q(z)}$.

(École polytechnique)

▷ **Solution.**

1. Si Q est constant, f est un polynôme. Elle est donc développable en une série entière avec un rayon infini. Supposons Q non constant. En décomposant f en éléments simples sur $\mathbb{C}$, on montre que f est somme d'un polynôme correspondant à sa partie entière et de fonctions de la forme $z \longmapsto \dfrac{\alpha}{(z - z_0)^p}$, où $\alpha \in \mathbb{C}$, $z_0 \neq 0$ (car 0 n'est pas un pôle de f) et $p \in \mathbb{N}^*$. La fonction $z \longmapsto \dfrac{1}{z - z_0}$ est développable en série entière au voisinage de 0, car pour $|z| < |z_0|$, on a

$$\frac{1}{z - z_0} = -\frac{1}{z_0} \frac{1}{1 - \dfrac{z}{z_0}} = -\frac{1}{z_0} \sum_{k=0}^{+\infty} \left(\frac{z}{z_0}\right)^k.$$

On en déduit que pour tout $p \geqslant 1$ la fonction $z \longmapsto \dfrac{1}{(z - z_0)^p}$ est développable en série entière au voisinage de 0, comme produit de telles fonctions et qu'il en est de même de $z \longmapsto \dfrac{\alpha}{(z - z_0)^p}$. Enfin f est développable en série entière au voisinage de 0, comme somme de telles fonctions. Notons $\sum_{n=0}^{+\infty} u_n z^n$ ce développement en série entière, R son

rayon de convergence et ρ le plus petit module des pôles de f. Ce qui précède montre que $R \geqslant \rho$. En fait on peut noter qu'il y a forcément égalité. En effet, supposons $R > \rho$ et soit z_0 un pôle de f de module ρ. Pour $|z| < \rho$ on peut écrire $f(z) = \sum_{n=0}^{+\infty} u_n z^n$. On prend $z = tz_0$ avec $t \in [0, 1[$ et on fait tendre t vers 1. La quantité $|f(tz_0)|$ tend vers $+\infty$ car z_0 est un pôle, alors que la série admet une limite finie en z_0 puisque, comme $R > \rho$, z_0 est dans son disque ouvert de convergence. L'hypothèse $R > \rho$ est absurde et on a bien $R = \rho$.

Pour $|z| < R$ on peut donc écrire $f(z) = \sum_{n=0}^{+\infty} u_n z^n$. Notons p et q les degrés respectifs de P et Q et posons $P = \sum_{k=0}^{p} a_k X^k$ et $Q = \sum_{k=0}^{q} b_k X^k$. On a, pour $|z| < R$,

$$P(z) = Q(z)f(z) = \sum_{k=0}^{q} b_k z^k \sum_{n=0}^{+\infty} u_n z^n.$$

Le développement en série de Qf est donné par un produit de Cauchy et par unicité du développement en série entière, les termes d'indice $> p$ sont nuls. Si on a $q > p$, les termes d'indice $\geqslant q$ sont nuls. On a alors, pour $n \geqslant q$, $\sum_{k=0}^{q} b_k u_{n-k} = 0$, ce qui peut s'écrire, puisque $b_0 = Q(0) \neq 0$,

$$u_n = -\sum_{k=1}^{q} \frac{b_k}{b_0} u_{n-k}.$$

La suite (u_n) vérifie donc une relation de récurrence linéaire à coefficients constants d'ordre q. Dans le cas où $q \leqslant p$, il suffit d'écrire $Q = \sum_{k=0}^{p+1} b_k X^k$ avec $b_{q+1} = \cdots = b_{p+1} = 0$ pour se ramener au cas précédent.

2. Supposons, réciproquement, que la suite (u_n) vérifie une relation linéaire à coefficients constants. Il existe donc $q \in \mathbb{N}^*$ et $(\alpha_1, \alpha_2, \ldots, \alpha_q) \in \mathbb{C}^q$, tels que, pour $n \geqslant q$, $u_n = \sum_{k=1}^{q} \alpha_k u_{n-k}$.

On va montrer que $\sum u_n z^n$ a un rayon de convergence non nul. Ce résultat est immédiat si on connaît la forme générale d'une suite récurrente linéaire (c'est une combinaison linéaire de suites de la forme $n^k \rho^n$) mais nous allons en donner une preuve n'utilisant pas ce fait. On va majorer $|u_n|$ par une suite géométrique. Posons $K = \max\left(\sum_{k=1}^{n} |\alpha_k|, 1\right)$ et $M = \max_{0 \leqslant k \leqslant q-1} |u_k|$. On va montrer par récurrence sur n que, pour tout $n \in \mathbb{N}$, $|u_n| \leqslant K^n M$. C'est vrai pour $n \leqslant q-1$, car $K^n \geqslant 1$, par définition. Si la propriété est vraie jusqu'au rang $n-1$, avec $n \geqslant q$, alors

$$|u_n| \leqslant \sum_{k=1}^{q} |\alpha_k||u_{n-k}| \leqslant \sum_{k=1}^{q} |\alpha_k| \mathrm{M}\mathrm{K}^{n-k} \leqslant \mathrm{M}\mathrm{K}^{n-1} \sum_{k=1}^{q} |\alpha_k| \leqslant \mathrm{M}\mathrm{K}^n.$$

On en déduit que le rayon de convergence de $\sum u_n z^n$ vérifie $\mathrm{R} \geqslant \dfrac{1}{\mathrm{K}} > 0$. On pose, pour $|z| < \mathrm{R}$, $f(z) = \sum_{n=0}^{+\infty} u_n z^n$. On écrit alors la relation de récurrence sous la forme $-u_n + \sum_{k=1}^{q} \alpha_k u_{n-k} = 0$. Le calcul de la première question nous conduit à poser $\mathrm{Q} = \sum_{k=1}^{q} \alpha_k \mathrm{X}^k - 1$ et à considérer la fonction $\mathrm{Q}f$. Elle est développable en série entière de rayon $\geqslant \mathrm{R}$ et son développement $\sum_{n=0}^{+\infty} a_n z^n$ est donné par un produit de Cauchy. On obtient, pour $n \geqslant q$, $a_n = -u_n + \sum_{k=1}^{q} \alpha_k u_{n-k} = 0$. On en déduit que, pour $|z| < \mathrm{R}$, $\mathrm{Q}(z)f(z) = \sum_{n=0}^{q-1} a_n z^n$ et la fonction $\mathrm{Q}f$ est donc un polynôme P. On a $\mathrm{Q}(0) = -1 \neq 0$, donc dans un voisinage de 0, $\mathrm{Q}(z)$ est non nul et on peut écrire $\sum_{n=0}^{+\infty} u_n z^n = \dfrac{\mathrm{P}(z)}{\mathrm{Q}(z)}$. $\lhd$

L'exercice suivant est l'occasion de s'exercer à écrire le développement en série entière d'une fraction rationnelle, dans différents cas – pôles réels ou non.

3.13. Signe des coefficients d'un développement en série entière

Donner une condition nécessaire et suffisante sur le réel a pour que les coefficients du développement en série entière en 0 de la fraction $\mathrm{F}(x) = \dfrac{1}{1 - ax + ax^2 - x^3}$ soient tous positifs.

(École polytechnique)

▷ **Solution.**

Puisque 0 n'est pas pôle de la fraction rationnelle F, celle-ci est développable en série entière autour de zéro. On a $\mathrm{F}(0) = 1$ et $\mathrm{F}'(0) = a$. Il est nécessaire, pour que les coefficients soient tous positifs, que $a \geqslant 0$, condition qu'on suppose réalisée par la suite. On a

$$\mathrm{F}(x) = \frac{1}{(1-x)(1-(a-1)x+x^2)}.$$

Nous allons décomposer F en éléments simples. Regardons le trinôme

$1-(a-1)x+x^2$. Son discriminant est $\Delta=(1-a)^2-4=a^2-2a-3$ qui se factorise en $\Delta=(a+1)(a-3)$.

• Étudions le cas où $a\in[0,3[$. Le trinôme admet alors deux racines complexes conjuguées dont le produit fait 1 : elles sont de module 1. Posons $\theta=\arccos\dfrac{a-1}{2}\in\left]0,\dfrac{2\pi}{3}\right]$. Ainsi,

$$\begin{aligned}1-(a-1)x+x^2&=1-2\cos\theta x+x^2=(x-e^{i\theta})(x-e^{-i\theta})\\&=(1-xe^{i\theta})(1-xe^{-i\theta}).\end{aligned}$$

Il existe alors $(A,B,C)\in\mathbb{C}^3$ avec A réel et $\overline{B}=C$ tel que

$$F(x)=\frac{A}{1-x}+\frac{B}{1-xe^{i\theta}}+\frac{C}{1-xe^{-i\theta}}.$$

On multiplie par $1-x$ et l'on fait $x=1$ pour obtenir

$$A=\frac{1}{(1-e^{-i\theta})(1-e^{i\theta})}=\frac{1}{4\sin^2\dfrac{\theta}{2}}.$$

On multiplie par $1-xe^{i\theta}$ et l'on évalue en $x=e^{-i\theta}$ pour obtenir

$$B=\frac{1}{(1-e^{-i\theta})(1-e^{-2i\theta})}=\frac{e^{i\frac{3\theta}{2}}}{2i\sin\dfrac{\theta}{2}\times 2i\sin\theta}=-\frac{e^{i\frac{3\theta}{2}}}{4\sin\theta\sin\dfrac{\theta}{2}}=\overline{C}.$$

Pour $|x|<1$, on peut écrire

$$\begin{aligned}F(x)&=A\sum_{n=0}^{+\infty}x^n+B\sum_{n=0}^{+\infty}e^{in\theta}x^n+C\sum_{n=0}^{+\infty}e^{-in\theta}x^n\\&=\sum_{n=0}^{+\infty}\underbrace{\left(A+2\operatorname{Re}(Be^{in\theta})\right)}_{=a_n}x^n.\end{aligned}$$

Pour tout $n\in\mathbb{N}$, on a donc

$$a_n=\frac{1}{4\sin^2\dfrac{\theta}{2}}+\operatorname{Re}\left(-\frac{e^{i\frac{3\theta}{2}}e^{in\theta}}{4\sin\theta\sin\dfrac{\theta}{2}}\right)=-\frac{1}{4\sin^2\dfrac{\theta}{2}}-\frac{2\cos[(n+3/2)\theta]}{4\sin\theta\sin\dfrac{\theta}{2}}$$

$$a_n=\frac{\sin\theta-2\sin\dfrac{\theta}{2}\cos[(n+3/2)\theta]}{4\sin^2\dfrac{\theta}{2}\sin\theta}=\frac{2\sin\dfrac{\theta}{2}\left[\cos\dfrac{\theta}{2}-\cos[(n+3/2)\theta]\right]}{4\sin^2\dfrac{\theta}{2}\sin\theta}$$

$$=\frac{\sin\left[\left(\dfrac{n+1}{2}\right)\theta\right]\sin\left[\left(\dfrac{n+2}{2}\right)\theta\right]}{\sin\dfrac{\theta}{2}\sin\theta}.$$

On passe de $\left(\frac{n+1}{2}\right)\theta$ à $\left(\frac{n+2}{2}\right)\theta$ en faisant un saut de $\frac{\theta}{2} \in \left]0, \frac{\pi}{3}\right]$. On aura $a_n < 0$ dès lors que ces deux réels sont de part et d'autre strictement d'un multiple de π (le sinus changeant de signe en tout point de $\pi\mathbb{Z}$).

Supposons que $\frac{2\pi}{\theta} \notin \mathbb{N}$. Si on pose $N = E(2\pi/\theta) \geqslant 3$, on a

$$N < \frac{2\pi}{\theta} < N+1 \text{ et } 0 < \frac{N\theta}{2} < \pi < \frac{(N+1)\theta}{2} < 2\pi$$

et donc

$$a_{N-1} = \frac{\sin\left[\frac{N}{2}\theta\right]\sin\left[\frac{N+1}{2}\theta\right]}{\sin\frac{\theta}{2}\sin\theta} < 0.$$

Supposons $\frac{2\pi}{\theta} \in \mathbb{N}$ et écrivons $\theta = \frac{2\pi}{N}$ avec $N \in \mathbb{N}^*$. On a alors $N \geqslant 3$. On considère les produits des sinus de deux termes consécutifs de la suite $\left(\frac{k\theta}{2}\right)_{k\in\mathbb{N}^*} = \left(\frac{k\pi}{N}\right)_{k\in\mathbb{N}^*}$. Ces produits sont toujours positifs ou nuls (car les deux termes consécutifs sont dans un intervalle de la forme $[l\pi, (l+1)\pi]$ avec $l \in \mathbb{N}$). Dans ce cas là, les a_n sont tous positifs.

• Étudions le cas $a = 3$. Pour $|x| < 1$, on a

$$F(x) = \frac{1}{(1-x)^3} = \sum_{n=0}^{+\infty} \frac{(n+2)(n+1)}{2} x^n$$

(développement obtenu en dérivant deux fois celui de $x \mapsto \frac{1}{1-x}$). Dans ce cas là, les coefficients sont tous positifs (strictement).

• Traitons enfin le cas $a > 3$. Le trinôme $1-(a-1)x+x^2$ admet alors deux racines réelles distinctes dont le produit fait 1, et la somme $a-1>0$: ces deux racines sont donc de la forme λ et $\frac{1}{\lambda}$ avec $\lambda > 1$. Il existe $(A, B, C) \in \mathbb{R}^3$ tel que

$$F(x) = \frac{1}{(1-x)(1-\lambda x)(1-\frac{x}{\lambda})} = \frac{A}{1-x} + \frac{B}{1-\lambda x} + \frac{C}{1-\frac{x}{\lambda}}.$$

Les réels A, B et C sont tous non nuls et pour $|x| < \frac{1}{\lambda}$,

$$F(x) = \sum_{n=0}^{+\infty} \left(A + B\lambda^n + C\frac{1}{\lambda^n}\right) x^n.$$

On trouve en multipliant successivement par $(1-x)$, $(1-\lambda x)$ et $(1-x/\lambda)$ et en substituant respectivement à x la valeur 1, $1/\lambda$ et λ :

$$A = -\frac{\lambda}{(1-\lambda)^2}, \quad B = \frac{\lambda^3}{(\lambda-1)^2(\lambda+1)} \quad \text{et} \quad C = \frac{1}{(\lambda-1)^2(\lambda+1)}.$$

Ainsi, on obtient

$$\mathrm{F}(x) = \frac{1}{(\lambda-1)^2(\lambda+1)} \sum_{n=0}^{+\infty} \left(-\lambda(\lambda+1) + \lambda^{n+3} + \lambda^{-n}\right) x^n.$$

Étudions la fonction $h : t \geqslant 0 \longmapsto \lambda^{3+t} + \lambda^{-t}$. Elle est de classe $\mathcal{C}^\infty$ et pour $t \geqslant 0$, $h'(t) = \ln\lambda\,(\lambda^{3+t} - \lambda^{-t}) > 0$ car $\lambda > 1$. Donc h est croissante. Ainsi pour tout $n \in \mathbb{N}$,

$$\begin{aligned} -\lambda(\lambda+1) + \lambda^{n+3} + \lambda^{-n} &\geqslant -\lambda(\lambda+1) + h(0) = \lambda^3 - \lambda^2 - \lambda + 1 \\ &= (\lambda-1)(\lambda^2-1) > 0. \end{aligned}$$

Donc tous les coefficients de la série entière sont positifs (strictement).

Conclusion. Pour que les coefficients de la série entière de F soient tous positifs (ou nuls), il faut et il suffit que $a \geqslant 3$ ou que a soit de la forme $1 + 2\cos\left(\dfrac{2\pi}{\mathrm{N}}\right)$ avec $\mathrm{N} \geqslant 3$. ◁

3.14. Séries génératrices de deux suites récurrentes

Soit $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ deux suites de $\mathbb{C}$ vérifiant $u_{n+1} = u_n - v_n$ et $v_{n+1} = u_n - 2v_n$ pour tout $n \in \mathbb{N}$. Donner le rayon de convergence et la somme des séries entières $\sum u_n z^n$ et $\sum v_n z^n$.

(École polytechnique)

▷ **Solution.**

Il est possible d'exprimer u_n et v_n en fonction de n en introduisant $\mathrm{X}_n = \begin{pmatrix} u_n \\ v_n \end{pmatrix}$ et $\mathrm{A} = \begin{pmatrix} 1 & -1 \\ 1 & -2 \end{pmatrix}$. On a la relation $\mathrm{X}_{n+1} = \mathrm{A}\mathrm{X}_n$ et donc $\mathrm{X}_n = \mathrm{A}^n\mathrm{X}_0$. Comme A est diagonalisable, sa diagonalisation donne l'expression de A^n, puis de X_n. Les suites $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ s'exprimant comme combinaison linéaire de deux suites géométriques (A possède deux valeurs propres distinctes), on peut calculer la somme des séries entières correspondantes. Cependant, il s'avère plus rapide d'exprimer le système linéaire dont $\displaystyle\sum_{n=0}^{+\infty} u_n z^n$ et $\displaystyle\sum_{n=0}^{+\infty} v_n z^n$ doivent être solutions.

Posons $f(z) = \displaystyle\sum_{n=0}^{+\infty} u_n z^n$ et $g(z) = \displaystyle\sum_{n=0}^{+\infty} v_n z^n$ et notons R_f et R_g les rayons de convergence respectifs de ces deux séries entières. Supposons que ces rayons sont strictement positifs. On a alors, lorsque $|z| < \min(\mathrm{R}_f, \mathrm{R}_g)$,

$$\begin{aligned} f(z) &= u_0 + \sum_{n=0}^{+\infty} u_{n+1} z^{n+1} = u_0 + \sum_{n=0}^{+\infty} (u_n - v_n) z^{n+1} \\ &= u_0 + z\big(f(z) - g(z)\big), \end{aligned}$$

c'est-à-dire

$$(1-z)f(z)+zg(z)=u_0. \tag{1}$$

On a de même

$$\begin{aligned} g(z) &= v_0+\sum_{n=0}^{+\infty} v_{n+1}z^{n+1} = v_0+\sum_{n=0}^{+\infty}(u_n-2v_n)z^{n+1} \\ &= v_0+z(f(z)-2g(z)) \end{aligned}$$

et donc

$$-zf(z)+(1+2z)g(z)=v_0. \tag{2}$$

Les équations (1) et (2) constituent un système linéaire de déterminant $(1-z)(1+2z)+z^2=1+2z-z-2z^2+z^2=1+z-z^2$. Le polynôme X^2-X-1 a deux racines réelles. $\dfrac{1+\sqrt{5}}{2}$ et $\dfrac{1-\sqrt{5}}{2}$. Par conséquent, si $|z|<\dfrac{\sqrt{5}-1}{2}$, le système est de Cramer et il donne

$$f(z)=\frac{1}{1+z-z^2}\begin{vmatrix} u_0 & z \\ v_0 & 1+2z \end{vmatrix}=\frac{z(2u_0-v_0)+u_0}{1+z-z^2},$$

$$g(z)=\frac{1}{1+z-z^2}\begin{vmatrix} 1-z & u_0 \\ -z & v_0 \end{vmatrix}=\frac{z(u_0-v_0)+v_0}{1+z-z^2}.$$

Considérons maintenant les fonctions $f : z \longmapsto \dfrac{z(2u_0-v_0)+u_0}{1+z-z^2}$ et $g : z \longmapsto \dfrac{z(u_0-v_0)+v_0}{1+z-z^2}$. Ce sont des fractions rationnelles définies pour $z \in \mathbb{C}$, $|z|<\dfrac{\sqrt{5}-1}{2}$ qui vérifient les équations (1) et (2). Puisque 0 n'est un pôle ni de f, ni de g, f et g sont développables en série entière autour de zéro, avec un rayon égal au plus petit module de leurs pôles (au moins $\dfrac{\sqrt{5}-1}{2}$). Si on écrit $f(z)=\sum_{n=0}^{\infty} u_n z^n$ et $g(z)=\sum_{n=0}^{\infty} v_n z^n$, les équations (1) et (2) entraînent que, pour $n \in \mathbb{N}$, $u_{n+1}=u_n-v_n$ et $v_{n+1}=u_n-2v_n$. On retrouve les suites du départ.

Conclusion. Les séries entières $f(z)=\sum_{n=0}^{\infty} u_n z^n$ et $g(z)=\sum_{n=0}^{\infty} v_n z^n$ avec $(u_n)_{n\in\mathbb{N}}$ et $(v_n)_{n\in\mathbb{N}}$ définies par l'énoncé ont un rayon de convergence au moins égal à $\dfrac{\sqrt{5}-1}{2}$ et s'expriment ainsi

$$\boxed{f(z)=\frac{z(2u_0-v_0)+u_0}{1+z-z^2} \quad \text{et} \quad g(z)=\frac{z(u_0-v_0)+v_0}{1+z-z^2}}.$$

En ce qui concerne leur rayon respectif R_f et R_g, on a

- $R_f=R_g=+\infty$ si $u_0=v_0=0$;

• $R_f = \dfrac{\sqrt{5}+1}{2}$ si $2u_0 - v_0 \neq 0$ et $\dfrac{u_0}{2u_0 - v_0} = \dfrac{\sqrt{5}-1}{2}$ (le facteur $z - \dfrac{1-\sqrt{5}}{2}$ se simplifie au numérateur et au dénominateur) ; dans ce cas, le facteur $z - \dfrac{1-\sqrt{5}}{2}$ se simplifie aussi au numérateur et au dénominateur dans l'expression de $g(z)$ et $R_g = R_f$;

• dans les autres cas, $R_f = R_g = \dfrac{\sqrt{5}-1}{2}$. ◁

On peut facilement en déduire les expressions de u_n et v_n en fonction de n, u_0 et v_0. Il suffit pour cela d'écrire le développement en série entière des fractions $\dfrac{1}{1+z-z^2}$ et $\dfrac{z}{1+z-z^2}$.

Les deux exercices suivants sont liés à des problèmes de dénombrement. Dans les deux cas, pour étudier une suite (u_n), on considère sa série génératrice, définie par $f(z) = \sum_{n=0}^{+\infty} u_n z^n$. Pour d'autres exemples d'utilisation de cette notion, le lecteur se reportera à notre tome 1 d'algèbre (exercices 1.3, 1.5 et 5.26). Ici, l'idée fondamentale qui permet de calculer les sommes des séries génératrices est la suivante : si A *et* B *sont deux parties non vides de* $\mathbb{N}$, *on a, si* $|z| < 1$, $\sum_{n\in A} z^n \times \sum_{n\in B} z^n = \sum_{n\in\mathbb{N}} \alpha_n z^n$, *où α_n est le nombre de façons d'écrire n comme somme d'un élément de* A *et d'un élément de* B. *En effet, ces trois séries ont un rayon de convergence au moins égal à* 1 *et l'égalité revient à faire le produit de Cauchy des séries entières.*

3.15. Partitions d'un entier en parts fixées

Soient $a_1, \dots, a_k$ des entiers naturels non nuls premiers entre eux dans leur ensemble. Pour $n \geqslant 1$, on note u_n le nombre de k-uplets $(x_1, \dots, x_k) \in \mathbb{N}^k$ tels que $a_1x_1 + \cdots + a_kx_k = n$. Déterminer un équivalent de u_n en $+\infty$. On pourra considérer $\prod_{i=1}^{k} \dfrac{1}{1-z^{a_i}}$.

(École polytechnique)

▷ **Solution.**

Les séries $\sum_{x_i\in\mathbb{N}} z^{x_i a_i}$ (pour $1 \leqslant i \leqslant k$) ont toutes pour rayon de convergence 1. Considérons le produit de Cauchy de ces k séries entières. Le coefficient de z^n dans ce produit de Cauchy est $\sum_{\substack{(x_1,\dots,x_k)\in\mathbb{N}^k \\ a_1x_1+\cdots+a_kx_k=n}} 1 = u_n$.

On a donc, pour $|z| < 1$,

$$f(z) = \sum_{n=0}^{+\infty} u_n z^n = \prod_{i=1}^{k} \left(\sum_{r_i=0}^{+\infty} z^{r_i a_i} \right) = \prod_{i=1}^{k} \frac{1}{1-z^{a_i}}.$$

La fonction f est la série génératrice de la suite (u_n). C'est une fraction rationnelle dont les pôles sont les racines a_1-ièmes, $\ldots$, a_k-ièmes de l'unité. Le pôle 1 est de multiplicité k et tous les autres sont de multiplicité $< k$. En effet, d'après le théorème de Bezout, il existe des entiers $u_1, \ldots, u_k$ tels que $\sum_{i=1}^{k} a_i u_i = 1$. Si ω est une racine de l'unité telle que $\omega^{a_1} = \cdots = \omega^{a_k} = 1$, alors

$$\omega = \omega^{\sum_{i=1}^{k} a_i u_i} = \prod_{i=1}^{k} (\omega^{a_i})^{u_i} = 1.$$

Si on note $\mathrm{P} = \{\omega_1, \omega_2, \ldots, \omega_p\}$ l'ensemble des pôles de f, avec $\omega_1 = 1$, il existe des constantes complexes c_{ij} $(1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant p-1)$ et α telles que, pour $|z| < 1$,

$$f(z) = \frac{\alpha}{(1-z)^k} + \sum_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant p-1}} \frac{c_{ij}}{(\omega_i - z)^j}.$$

Pour $\omega \in \mathrm{P}$ et $j \in \mathbb{N}$, la fonction $z \longmapsto \dfrac{1}{(\omega - z)^j}$ est développable en série entière de rayon de convergence 1. Ses coefficients s'obtiennent en dérivant $j-1$ fois le développement en série entière de la fonction $z \mapsto \dfrac{1}{\omega - z}$. On a, pour $|z| < 1$, $\dfrac{1}{\omega - z} = \dfrac{1}{\omega} \dfrac{1}{1 - \dfrac{z}{\omega}} = \sum_{n=0}^{\infty} \dfrac{z^n}{\omega^{n+1}}$, puis, pour $j \leqslant k-1$,

$$\frac{(j-1)!}{(\omega - z)^j} = \sum_{n=j-1}^{+\infty} \frac{n!}{(n-j+1)!} \frac{z^{n-j+1}}{\omega^{n+1}}$$

$$\frac{1}{(\omega - z)^j} = \sum_{n=0}^{+\infty} \mathrm{C}_{n+j-1}^{n} \frac{z^n}{\omega^{n+j}}.$$

On déduit de ce qui précède, en prenant les coefficients de z^n dans les différents développements en série entière, que

$$u_n = \alpha \mathrm{C}_{n+k-1}^{n} + \sum_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant p-1}} c_{ij} \mathrm{C}_{n+j-1}^{n} \omega^{-n-j},$$

le premier terme correspondant au développement de $\dfrac{1}{(1-z)^k}$. Celui-ci est équivalent, quand n tend vers $+\infty$, à $\alpha \dfrac{n^{k-1}}{(k-1)!}$. On remarque que tous les autres termes de la somme sont négligeables devant n^{k-1}. On

donc $u_n \underset{n\to\infty}{\sim} \alpha \dfrac{n^{k-1}}{(k-1)!}$. Il nous reste à calculer α. Classiquement, on multiplie $f(z)$ par $(1-z)^k$ et on fait $z = 1$. On obtient

$$(1-z)^k f(z) = \prod_{i=1}^{k} \frac{1-z}{1-z^{a_i}} = \prod_{i=1}^{k} \frac{1}{(1+z+\cdots+z^{a_i-1})},$$

et donc $\alpha = \dfrac{1}{a_1 a_2 \dots a_k}$. On conclut finalement que

$$\boxed{u_n \underset{n\to\infty}{\sim} \frac{1}{a_1 a_2 \dots a_k} \frac{n^{k-1}}{(k-1)!}}. \quad \triangleleft$$

Dans un cas concret, le développement en série entière de la fonction f permet d'obtenir une expression explicite de u_n. Si par exemple on cherche de combien de manières il est possible de faire n euros avec des pièces (ou des billets !) de 1, 2 ou 5 euros, il suffit de déterminer le coefficient u_n de x^n dans le développement en série entière de $\dfrac{1}{(1-x)(1-x^2)(1-x^5)}$. Après quelques calculs que nous invitons le lecteur à effectuer, on obtient pour cet exemple $u_n = \dfrac{13}{40} + \dfrac{n+1}{4} + \dfrac{(n+1)(n+2)}{20} + \dfrac{(-1)^n}{8} + \dfrac{\varepsilon_n}{5}$ avec $\varepsilon_n = 1$ si $n \equiv 0$ ou $n \equiv 2$ modulo 5, $\varepsilon_n = -1$ si $n \equiv 3$ ou $n \equiv 4$ modulo 5 et $\varepsilon_n = 0$ si $n \equiv 1$ modulo 5.

L'exercice suivant s'intéresse à la généralisation naturelle de ce qui précède en étudiant la série génératrice de la suite $p(n)$, où $p(n)$ est le nombre de partitions de n, c'est-à-dire le nombre de manières d'écrire n comme somme d'entiers non nuls (sans tenir compte de l'ordre).

3.16. Série génératrice du nombre de partitions de n

Pour $|t| < 1$, on pose $f(t) = \displaystyle\prod_{k=1}^{+\infty} \frac{1}{1-t^k}$.

1. Montrer que f est bien définie.

2. Démontrer, pour $|t| < 1$, l'identité $f(t) = 1 + \displaystyle\sum_{n=1}^{+\infty} p(n) t^n$, où $p(n)$ est le nombre de suites $y = (y_k)_{k\in\mathbb{N}^*}$ d'entiers naturels telles que $\displaystyle\sum_{k=1}^{+\infty} k y_k = n$.

(École polytechnique)

▷ **Solution.**

1. Posons pour $t \in]-1,1[$ et $n \geqslant 1$, $f_n(t) = \prod_{k=1}^{n} \frac{1}{1-t^k}$. On a

$$f_n(t) = \exp\left(-\sum_{k=1}^{n} \ln(1-t^k)\right).$$

Comme pour k tendant vers l'infini, $\ln(1-t^k) \sim -t^k$, et comme la série géométrique de raison $|t|$ converge, le théorème de comparaison des séries à termes positifs assure la convergence absolue de la série de terme général $\ln(1-t^k)$. Par continuité de l'exponentielle,

$$f_n(t) \xrightarrow[n\to+\infty]{} f(t) = \exp\left(-\sum_{k=1}^{+\infty} \ln(1-t^k)\right).$$

Donc, f est bien définie sur $]-1,1[$.

2. D'après la règle du produit de Cauchy des séries entières, f_n est développable en série entière sur $]-1,1[$. Écrivons pour $t \in]-1,1[$

$$f_n(t) = \sum_{\ell=0}^{+\infty} a_n(\ell) t^\ell.$$

On a alors

$$f_n(t) = \prod_{k=1}^{n} \frac{1}{1-t^k} = \prod_{k=1}^{n} \left(\sum_{\ell=0}^{+\infty} t^{k\ell}\right) = \sum_{\ell=0}^{+\infty} \left(\sum_{y_1+2y_2+\cdots+ny_n=\ell} 1\right) t^\ell.$$

Ainsi, $a_n(\ell)$ est le nombre de n-uplets $(y_1,\dots,y_n)$ d'entiers $\geqslant 0$ vérifiant $y_1+2y_2+\cdots+ny_n = \ell$. On remarque que pour $n \geqslant \ell \geqslant 1$, $a_n(\ell) = p(\ell)$, puisque les termes y_k des suites qui interviennent dans la définition de $p(\ell)$ sont nécessairement nuls pour $k > \ell$. On pose $p(0) = 1$.

On a, pour $t \in [0,1[$,

$$0 \leqslant \sum_{\ell=0}^{n} p(\ell) t^\ell = \sum_{\ell=0}^{n} a_n(\ell) t^\ell \leqslant f_n(t) \leqslant f(t).$$

Donc, la série à termes positifs $\sum p(\ell) t^\ell$ converge. Cette série entière est de rayon de convergence 1.

Pour t dans $]-1,1[$, notons $\delta_n(t) = \left| f_n(t) - \sum_{\ell=0}^{+\infty} p(\ell) t^\ell \right|$. Comme pour tout n et tout ℓ, $0 \leqslant a_n(\ell) \leqslant p(\ell)$, on a

$$\delta_n(t) = \left| \sum_{\ell=n+1}^{+\infty} (a_n(\ell) - p(\ell)) t^\ell \right| \leqslant \sum_{\ell=n+1}^{+\infty} (p(\ell) - a_n(\ell)) |t|^\ell$$

$$\leqslant \sum_{\ell=n+1}^{+\infty} p(\ell) |t|^\ell \xrightarrow[n\to+\infty]{} 0.$$

On conclut donc que pour $t \in]-1,1[$,

$$\boxed{f(t) = \prod_{k=1}^{+\infty} \frac{1}{1-t^k} = \sum_{n=0}^{+\infty} p(n)t^n}. \lhd$$

La formule démontrée dans la question 2 qui décompose f comme un produit infini est due à Euler. Hardy et Ramanujan ont démontré en 1918 que $p(n) \underset{n\to\infty}{\sim} \frac{1}{4n\sqrt{3}} e^{\pi\sqrt{\frac{2n}{3}}}$, introduisant pour cela une méthode appelée aujourd'hui méthode du cercle, qui s'est révélée très puissante en théorie analytique des nombres.

3.17. Signe des dérivées successives d'une fonction

Soit $a > 0$. On pose, pour $x \in]-a,a[$, $f(x) = \ln(a-x)\ln(x+a)$. Déterminer, selon les valeurs de a et pour n assez grand, le signe de $f^{(n)}$. On pourra utiliser le développement de f en série entière.
(École polytechnique)

Solution.

Remarquons que f est paire. Si $x \in]-a,a[$, on obtient

$$\begin{aligned} f(x) &= (\ln a)^2 + \ln a \ln\left(1+\frac{x}{a}\right) + \ln a \ln\left(1-\frac{x}{a}\right) \\ &\qquad + \ln\left(1+\frac{x}{a}\right)\ln\left(1-\frac{x}{a}\right) \\ &= (\ln a)^2 + \ln a \ln\left(1-\frac{x^2}{a^2}\right) + \ln\left(1+\frac{x}{a}\right)\ln\left(1-\frac{x}{a}\right) \\ &= (\ln a)^2 - \ln a \sum_{n=1}^{+\infty} \frac{1}{n}\frac{x^{2n}}{a^{2n}} - \left(\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n}\frac{x^n}{a^n}\right)\left(\sum_{n=1}^{+\infty} \frac{1}{n}\frac{x^n}{a^n}\right), \end{aligned}$$

puisque $\left|\frac{x}{a}\right| < 1$ et que la fonction $y \longmapsto \ln(1+y)$ est développable en 0 avec un rayon égal à 1. Le dernier terme est un produit de séries entières de rayon a. Ce produit est alors développable en série entière (de rayon au moins a) et s'exprime à l'aide du produit de Cauchy des séries :

$$\begin{aligned} \left(\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n}\frac{x^n}{a^n}\right)\left(\sum_{n=1}^{+\infty} \frac{1}{n}\frac{x^n}{a^n}\right) &= -\left(\sum_{n=1}^{+\infty} \frac{(-1)^n}{n}\frac{x^n}{a^n}\right)\left(\sum_{n=1}^{+\infty} \frac{1}{n}\frac{x^n}{a^n}\right) \\ &= -\sum_{n=2}^{+\infty}\left(\sum_{k=1}^{n-1} \frac{(-1)^k}{k(n-k)}\right)\frac{x^n}{a^n}. \end{aligned}$$

La fonction f est donc développable en série entière en 0 avec un rayon

de convergence égal à a et, f étant paire, les coefficients des puissances impaires sont nuls :

$$f(x) = (\ln a)^2 - \ln a \sum_{n=1}^{+\infty} \frac{1}{n} \frac{x^{2n}}{a^{2n}} + \sum_{n=1}^{+\infty} \left(\sum_{k=1}^{2n-1} \frac{(-1)^k}{k(2n-k)} \right) \frac{x^{2n}}{a^{2n}}.$$

Arrangeons $\displaystyle\sum_{k=1}^{2n-1} \frac{(-1)^k}{k(2n-k)}$. On obtient

$$\begin{aligned}
\sum_{k=1}^{2n-1} \frac{(-1)^k}{k(2n-k)} &= \sum_{k=1}^{2n-1} \frac{(-1)^k}{2n} \left(\frac{1}{k} + \frac{1}{2n-k} \right) \\
&= \frac{1}{2n} \left(\sum_{k=1}^{2n-1} \frac{(1)^k}{k} + \frac{(-1)^{2n-k}}{2n-k} \right) \\
&= \frac{2}{2n} \left(\sum_{k=1}^{2n-1} \frac{(-1)^k}{k} \right) = \frac{1}{n} \left(\sum_{k=1}^{2n-1} \frac{(-1)^k}{k} \right).
\end{aligned}$$

C'est un résultat classique que $\displaystyle\sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} = \ln(2)$. Pour le voir, il suffit d'écrire que

$$\sum_{k=1}^{+\infty} \frac{(-1)^{k+1}}{k} = \sum_{k=1}^{+\infty} \int_0^1 (-x)^{k-1} \mathrm{d}x = \int_0^1 \sum_{k=0}^{+\infty} (-x)^k \mathrm{d}x = \int_0^1 \frac{\mathrm{d}x}{1+x} = \ln 2.$$

L'interversion de l'intégration et de la sommation se justifie par le théorème de convergence dominée puisque pour tout $n \in \mathbb{N}$

$$\left| \sum_{k=0}^{n} (-x)^k \right| = \left| \frac{1-(-x)^{n+1}}{1+x} \right| \leqslant \frac{2}{1+x}.$$

Notons qu'il est aussi possible d'utiliser le théorème d'Abel-Dirichlet (voir page 178). On a donc

$$\sum_{k=1}^{2n-1} \frac{(-1)^k}{k} = -\ln 2 - \sum_{k=2n}^{+\infty} \frac{(-1)^k}{k}.$$

Exprimons alors $f(x)$ pour $x \in]-a, a[$:

$$\begin{aligned}
f(x) &= (\ln a)^2 + \sum_{n=1}^{+\infty} \frac{1}{n} \left(-\ln a - \ln 2 - \sum_{k=2n}^{+\infty} \frac{(-1)^k}{k} \right) \frac{x^{2n}}{a^{2n}} \\
&= (\ln a)^2 + \sum_{n=1}^{+\infty} \underbrace{\frac{1}{n} \left(-\ln 2a - \sum_{k=2n}^{+\infty} \frac{(-1)^k}{k} \right)}_{=\mathrm{C}_n} \frac{x^{2n}}{a^{2n}} \\
&= (\ln a)^2 + \sum_{n=1}^{+\infty} \mathrm{C}_n \frac{x^{2n}}{a^{2n}}.
\end{aligned}$$

Si $m \geqslant 1$, la dérivée d'ordre m de f s'écrit

$$f^{(m)}(x) = \sum_{n=\mathrm{E}(m/2)+1}^{+\infty} \mathrm{C}_n 2n(2n-1)\dots(2n-m+1)\frac{x^{2n-m}}{a^{2n}}.$$

Si $a \neq \frac{1}{2}$, $\mathrm{C}_n = \frac{1}{n}(-\ln 2a + o(1)) \sim \frac{-\ln 2a}{n}$ au voisinage de l'infini et il existe donc $m_0 > 0$ tel que pour $n \geqslant m_0$, $\mathrm{C}_n > 0$ (resp. $\mathrm{C}_n < 0$) si $a < \frac{1}{2}$ (resp. $a > \frac{1}{2}$). Si $a = \frac{1}{2}$, $\mathrm{C}_n = -\frac{1}{n}\left(\sum_{k=2n}^{+\infty} \frac{(-1)^k}{k}\right) < 0$. En effet, d'après le théorème spécial des séries alternées, le reste $\sum_{k=2n}^{+\infty} \frac{(-1)^k}{k}$ est du signe du premier terme de la série. Posons dans ce cas $m_0 = 1$.

On en tire le signe de $f^{(2m)}(x)$ et $f^{(2m+1)}(x)$ pour $m \geqslant m_0$:

Cas $a \geqslant \frac{1}{2}$

x	$-a$		0		a
$f^{(2m)}(x)$		$-$		$-$	
$f^{(2m+1)}(x)$		$+$	0	$-$	

Cas $a < \frac{1}{2}$

x	$-a$		0		a
$f^{(2m)}(x)$		$+$		$+$	
$f^{(2m+1)}(x)$		$-$	0	$+$	

Dans les exercices suivants, on détermine des développements en série entière. L'existence de ces développements ne pose pas de problème en général. Mais pour ceux qui résultent d'un produit de Cauchy, les coefficients sont souvent malaisés à obtenir et on préfère montrer que la fonction à développer vérifie une équation différentielle.

3.18. Développement en série entière (1)

Déterminer le développement en série entière en 0 de la fonction $x \longmapsto e^{-x^2/2}\int_0^x e^{t^2/2}\mathrm{d}t$.

(École polytechnique)

Solution.

Notons pour x réel, $\mathrm{F}(x) = e^{-x^2/2}\int_0^x e^{t^2/2}\mathrm{d}t$. Pour tout réel x,

$$e^{-x^2/2} = \sum_{n=0}^{+\infty} \frac{(-1)^n x^{2n}}{2^n n!} \quad \text{et} \quad e^{x^2/2} = \sum_{n=0}^{+\infty} \frac{x^{2n}}{2^n n!},$$

ce qui garantit que les fonctions $x \longmapsto e^{-x^2/2}$ et $x \longmapsto e^{x^2/2}$ sont développables en série entière sur $\mathbb{R}$ tout entier. Il en va de même de leurs primitives et donc comme le produit de deux fonctions développables en

série entière sur $\mathbb{R}$ est encore développable en série entière sur $\mathbb{R}$, F vérifie également cette propriété.

Écrivons pour $x \in \mathbb{R}$, $\mathrm{F}(x) = \sum_{n=0}^{+\infty} a_n x^n$. En dérivant F, on constate que F vérifie l'équation différentielle $y' + xy = 1$, ce qui donne pour tout réel x

$$\sum_{n=0}^{+\infty} (n+1)a_{n+1}x^n + \sum_{n=1}^{+\infty} a_{n-1}x^n = 1$$

et, par unicité du développement en série entière, $a_{n+1} = -\dfrac{a_{n-1}}{n+1}$ et $a_1 = 1$. Comme $a_0 = \mathrm{F}(0) = 0$, tous les termes d'indice pair sont nuls (c'était prévisible puisque F est impaire) et pour les autres,

$$\begin{aligned} a_{2n+1} &= -\frac{a_{2n-1}}{2n+1} = \cdots = \frac{(-1)^n a_1}{(2n+1)(2n-1)\cdots 3} \\ &= \frac{(-1)^n (2n)(2n-2)\ldots 2}{(2n+1)!}, \end{aligned}$$

soit finalement, $a_{2n+1} = \dfrac{(-1)^n 2^n n!}{(2n+1)!}$. On conclut donc que, pour tout $x \in \mathbb{R}$,

$$\boxed{\mathrm{F}(x) = \sum_{n=0}^{+\infty} \frac{(-1)^n 2^n n!}{(2n+1)!} x^{2n+1}}. \quad \lhd$$

3.19. Développement en série entière (2)

Développer en série entière la fonction $x \longmapsto \arcsin x$, puis les fonctions $x \longmapsto \dfrac{\arcsin x}{\sqrt{1-x^2}}$ et $x \longmapsto (\arcsin x)^2$.

(École polytechnique)

▷ **Solution.**

• Soit $f : x \longmapsto \arcsin x$. La fonction f est de classe $\mathcal{C}^\infty$ sur $]-1,1[$ et sa dérivée $x \longmapsto \dfrac{1}{\sqrt{1-x^2}}$ est développable en série entière avec un rayon égal à 1. Il en va donc de même de f. Le cours donne, pour $x \in]-1,1[$,

$$\begin{aligned} \frac{1}{\sqrt{1+x}} &= 1 + \sum_{n=1}^{+\infty} (-1)^n \frac{1\cdot 3\ldots(2n-1)}{2\cdot 4\ldots(2n)} x^n = \sum_{n=0}^{+\infty} (-1)^n \frac{(2n)!}{2^{2n}(n!)^2} x^n \\ &= \sum_{n=0}^{+\infty} (-1)^n \frac{C_{2n}^n}{4^n} x^n \end{aligned}$$

et donc

$$\frac{1}{\sqrt{1-x^2}} = \sum_{n=0}^{+\infty} \frac{C_{2n}^n}{4^n} x^{2n}.$$

On obtient alors le développement de arcsin en intégrant celui de sa dérivée terme à terme et, compte tenu de $\arcsin 0 = 0$,

$$\boxed{\arcsin x = \sum_{n=0}^{+\infty} \frac{C_{2n}^n}{(2n+1)4^n} x^{2n+1}}.$$

• Comme $x \longmapsto \arcsin x$ et $x \longmapsto \dfrac{1}{\sqrt{1-x^2}}$ sont développables en série entière avec un rayon égal à 1, leur produit l'est aussi avec un rayon supérieur ou égal à 1. En fait, ce rayon est exactement 1 puisque $\lim\limits_{x \to 1^-} \dfrac{\arcsin x}{\sqrt{1-x^2}} = +\infty$.

On constate vite que l'expression du coefficient de x^{2n+1} obtenue par la formule du produit de Cauchy ne se simplifie pas facilement. On va plutôt tenter de trouver une équation différentielle vérifiée par la fonction $g : x \in]-1,1[\longmapsto \dfrac{\arcsin x}{\sqrt{1-x^2}}$. Si $x \in]-1,1[$,

$$g'(x) = \frac{1}{1-x^2} + \frac{x \arcsin x}{(\sqrt{1-x^2})^3}.$$

Donc g vérifie l'équation différentielle linéaire $y'(1-x^2) - xy = 1$. Notons $\sum\limits_{n=0}^{+\infty} a_n x^n$ le développement en série entière de g. La relation vérifiée par g s'écrit

$$(1-x^2) \sum_{n=1}^{+\infty} n a_n x^{n-1} - \sum_{n=0}^{+\infty} a_n x^{n+1} = 1,$$

$$\sum_{n=1}^{+\infty} n a_n x^{n-1} - \sum_{n=1}^{+\infty} (n+1) a_n x^{n+1} - a_0 = 1,$$

$$-a_0 - 1 + a_1 + \sum_{n=1}^{+\infty} (n+2) a_{n+2} x^{n+1} - \sum_{n=1}^{+\infty} (n+1) a_n x^{n+1} = 0,$$

$$-a_0 - 1 + a_1 + \sum_{n=1}^{+\infty} \left[(n+2) a_{n+2} - (n+1) a_n\right] x^{n+1} = 0.$$

Comme g est impaire, tous les a_n pour n pair sont nuls. Par unicité du développement en série entière, on obtient

$$0 = -a_0 - 1 + a_1 = a_1 - 1 \quad \text{et} \quad a_{n+2} = \frac{n+1}{n+2} a_n \text{ si } n \geqslant 1.$$

D'où on tire $a_1 = 1$ et pour $n \geqslant 0$,

$$\begin{aligned} a_{2n+1} &= \frac{2n}{2n+1} a_{2n-1} = \frac{2n}{2n+1}\frac{2n-2}{2n-1} a_{2n-3} = \cdots \\ &= \frac{(2n)\cdot(2n-2)\ldots 4\cdot 2}{(2n+1)\cdot(2n-1)\ldots 5\cdot 3} a_1 \\ &= \frac{(2^n n!)^2}{(2n+1)!}. \end{aligned}$$

On a finalement, pour tout $x \in]-1,1[$,

$$\boxed{\frac{\arcsin x}{\sqrt{1-x^2}} = \sum_{n=0}^{+\infty} \frac{4^n (n!)^2}{(2n+1)!} x^{2n+1}}.$$

• Comme $\dfrac{\mathrm{d}(\arcsin x)^2}{\mathrm{d}x} = 2\dfrac{\arcsin x}{\sqrt{1-x^2}}$, on obtient le développement en série entière de $x \longmapsto (\arcsin x)^2$ en intégrant terme à terme celui de $2g$. Puisque $(\arcsin 0)^2 = 0$, on obtient, pour $|x| < 1$,

$$(\arcsin x)^2 = 2\sum_{n=0}^{+\infty} \frac{4^n (n!)^2}{(2n+1)!}\frac{x^{2n+2}}{2n+2} = \sum_{n=1}^{+\infty} \frac{4^{n-1}((n-1)!)^2}{(2n-1)!n} x^{2n}.$$

$$\boxed{(\arcsin x)^2 = \frac{1}{2}\sum_{n=1}^{+\infty} \frac{(2x)^{2n}}{n^2 \mathrm{C}_{2n}^n}}. \quad \triangleleft$$

En prenant $x = \dfrac{1}{2}$, il vient $\dfrac{\pi^2}{18} = \displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2 \mathrm{C}_{2n}^n}$. On trouvera aussi ce dernier développement par une autre méthode dans l'exercice 1.43.

Dans l'exercice suivant, on développe en série entière la composée d'une fonction développable en série entière et de la fonction exp. *On peut démontrer plus généralement, mais ce théorème n'est pas au programme, que la composée de deux fonctions développables en série entière est développable en série entière.*

3.20. Développement en série entière (3)

Soit $g(x) = \exp\left(\displaystyle\sum_{k=1}^{+\infty} \frac{x^k}{k^2}\right)$.

1. Démontrer que g est développable en série entière avec un rayon de convergence $\mathrm{R} \geqslant 1$. On note g_n les coefficients du développement de G.

2. Montrer que $g_n = o\left(\dfrac{\ln n}{n}\right)$.

(École polytechnique)

Solution.

1. Le rayon de convergence de la série entière $\sum \dfrac{x^n}{n^2}$ est 1. La fonction $f : x \longmapsto \sum_{n=1}^{+\infty} \dfrac{x^n}{n^2}$ est définie sur $[-1, 1]$. Elle est dérivable sur $]-1, 1[$ et $f'(x) = \sum_{n=1}^{+\infty} \dfrac{x^{n-1}}{n} = \sum_{n=0}^{+\infty} \dfrac{x^n}{n+1}$. On en déduit que g est définie sur $[-1, 1]$, dérivable sur $]-1, 1[$ et vérifie $g'(x) = g(x)f'(x)$. On a de plus $g(0) = 1$. La fonction f' étant continue sur $]-1, 1[$, l'équation différentielle linéaire $y' = f'y$ possède une solution maximale unique prenant en 0 la valeur 1 et qui est définie sur $]-1, 1[$ tout entier. Cette solution est g.

Montrons que l'équation différentielle $y' = yf'$ possède une solution développable en série entière G telle que $G(0) = 1$. On aura d'après ce qui précède $G = g$. Cherchons G sous la forme $G(x) = \sum_{n=0}^{+\infty} g_n x^n$ et supposons que cette série a un rayon de convergence non nul R. On a, pour $|x| < R$, $G'(x) = \sum_{n=1}^{+\infty} n g_n x^{n-1} = \sum_{n=0}^{+\infty} (n+1) g_{n+1} x^n$. La fonction fG est développable en série entière avec un rayon égal à $R' \geqslant \min(R, 1)$ le développement s'obtenant par un produit de Cauchy. L'égalité $G' = fG$ est vérifiée si on a, pour tout $n \in \mathbb{N}$,

$$(n+1) g_{n+1} = \sum_{k=0}^{n} \frac{g_k}{n-k+1}.$$

Le coefficient g_0 valant 1, cette relation de récurrence définit une suite unique. On montre facilement que, pour tout $n \in \mathbb{N}$, on a $0 \leqslant g_n \leqslant 1$. En effet $g_0 = 1$ et si on suppose la propriété vraie jusqu'au rang n alors

$$0 \leqslant g_{n+1} \leqslant \frac{1}{n+1} \sum_{k=0}^{n} \frac{g_k}{n-k+1} \leqslant \frac{1}{n+1} \sum_{k=0}^{n} g_k \leqslant 1.$$

On en déduit que le rayon de convergence de la série $\sum g_n x^n$ est supérieur ou égal à 1. L'équation différentielle possède une solution G telle que $G(0) = 1$, développable en série entière de rayon $R \geqslant 1$. Cette fonction est égale à g. La fonction g est donc développable en série entière sur $]-1, 1[$ [1].

2. On a, pour tout $n \in \mathbb{N}$, $0 \leqslant g_n \leqslant 1$. On en déduit que, pour tout $n \in \mathbb{N}^*$,

1. Notons que les résultats d'analyse complexe permettent d'obtenir sans calcul le

$$0 \leqslant g_n \leqslant \frac{1}{n}\sum_{k=0}^{n-1}\frac{1}{n-k} = \frac{1}{n}\sum_{k=1}^{n}\frac{1}{k} \underset{n\to\infty}{\sim} \frac{\ln n}{n}$$

et donc que $g_n = O\left(\dfrac{\ln n}{n}\right)$. Soit A tel que $g_n \leqslant \mathrm{A}\dfrac{\ln n}{n}$ pour tout $n \geqslant 2$. On obtient alors, pour $n \geqslant 2$,

$$\begin{aligned} 0 \leqslant g_n &\leqslant \frac{1}{n}\sum_{k=0}^{n-1}\frac{g_k}{n-k} \leqslant \frac{1}{n}\left(2 + \sum_{k=2}^{n-1}\frac{\mathrm{A}\ln k}{k(n-k)}\right) \\ &\leqslant \frac{1}{n}\left(2 + \mathrm{A}\ln n \sum_{k=2}^{n-1}\frac{1}{k(n-k)}\right) \\ &\leqslant \frac{1}{n}\left(2 + \mathrm{A}\frac{\ln n}{n}\sum_{k=2}^{n-1}\left(\frac{1}{k} + \frac{1}{n-k}\right)\right) \\ &\leqslant \frac{1}{n}\left(2 + 2\mathrm{A}\frac{\ln n}{n}\sum_{k=1}^{n}\frac{1}{k}\right) \underset{n\to\infty}{\sim} 2\mathrm{A}\frac{(\ln n)^2}{n^2}. \end{aligned}$$

Cela montre que $g_n = O\left(\dfrac{(\ln n)^2}{n^2}\right)$ et donc que $g_n \underset{n\to+\infty}{=} o\left(\dfrac{\ln n}{n}\right)$. ◁

Résumons la méthode utilisée dans cet exercice pour montrer que g est développable en série entière. On montre que g est solution d'une équation différentielle (E)*, puis que cette équation possède une solution développable en série entière h. En appliquant le théorème de Cauchy-Lipschitz, on montre que $h = g$ et donc que g est développable en série entière au voisinage de* 0.

3.21. Limite simple de polynômes à coefficients positifs

Soit $(\mathrm{P}_n)_{n\geqslant 0}$ une suite de polynômes à coefficients dans $\mathbb{R}_+$ qui converge simplement sur $[0,1]$ vers une fonction f. Montrer qu'il existe une suite $(b_n)_{n\geqslant 0}$ de réels positifs telle que

(*i*) le rayon de convergence R de la série $\sum b_n x^n$ est $\geqslant 1$;

(*ii*) $\forall x \in [0,1[,\ f(x) = \sum_{n=0}^{+\infty} b_n x^n$.

(**École normale supérieure**)

---

fait que la fonction $z \longmapsto \exp\sum_{n=1}^{+\infty}\frac{z^n}{n^2}$ est développable en série entière sur le disque unité ouvert de $\mathbb{C}$: c'est la composée de deux fonctions holomorphes.

Solution.

On pose, pour tout $n \in \mathbb{N}$, $P_n(x) = \sum_{k=0}^{+\infty} b_{n,k} x^k$, où $b_{n,k} = 0$ si k assez grand. Nous allons démontrer qu'il existe une application $\varphi : \mathbb{N} \longrightarrow \mathbb{R}$ strictement croissante telle que, pour tout $k \in \mathbb{N}$, $(b_{\varphi(n),k})$ converge. En appelant b_k la limite , on montrera qu'on obtient le résultat voulu.

Pour tout $n \in \mathbb{N}$, P_n est à coefficients positifs, donc pour tout $k \in \mathbb{N}$, on a $0 \leqslant b_{n,k} \leqslant P_n(1)$. La suite $(P_n(1))$ est convergente donc majorée par M. Ainsi, pour tout $(n,k) \in \mathbb{N}^2$, on a $0 \leqslant b_{n,k} \leqslant M$. Pour construire φ, on utilise le procédé diagonal de Cantor. La suite $(b_{n,0})$ est bornée donc, d'après le théorème de Bolzano-Weierstrass, on peut en extraire une sous-suite convergente. Soit donc φ_0 une extraction telle que la sous-suite $(b_{\varphi_0(n),0})$ soit convergente vers un réel b_0. La suite $(b_{\varphi_0(n),1})$ est bornée et on peut choisir une extraction φ_1 telle que la suite $(b_{\varphi_0\circ\varphi_1(n),1})$ converge vers un réel b_1. On réitère ce procédé pour construire par récurrence des extractions φ_k telles que $(b_{\varphi_0\circ\cdots\circ\varphi_k(n),k})$ converge vers b_k. On pose $\varphi(n) = \varphi_0 \circ \cdots \circ \varphi_n(n)$. Alors, pour tout $k \in \mathbb{N}$, $(b_{\varphi(n),k})_{n\geqslant k}$ est une suite extraite de $(b_{\varphi_0\circ\cdots\circ\varphi_k(n),k})_{n\geqslant k}$, qui converge vers b_k. En effet, on a $n-1 < n \leqslant \varphi(n)$, donc

$$\varphi(n-1) = \varphi_0 \circ \cdots \circ \varphi_{n-1}(n-1) < \varphi_0 \circ \cdots \circ \varphi_{n-1} \circ \varphi_n(n) = \varphi(n)$$

et $b_{\varphi(n),k} = b_{\varphi_0\circ\cdots\circ\varphi_k(\varphi_{k+1}\circ\cdots\circ\varphi_n)(n),k}$ pour $n \geqslant k$ donc $(b_{\varphi(n),k})$ converge vers b_k pour tout $k \in \mathbb{N}$.

Soit $x \in [0,1]$. Pour tout $n \in \mathbb{N}$, on a $0 \leqslant P_n(x) \leqslant P_n(1) \leqslant M$ et *a fortiori* $0 \leqslant P_{\varphi(n)}(x) \leqslant M$, *i.e.* $\sum_{k=0}^{+\infty} b_{\varphi(n),k} x^k \leqslant M$. Soit $N \in \mathbb{N}$. On a pour tout $n \in \mathbb{N}$, $\sum_{k=0}^{N} b_{\varphi(n),k} x^k \leqslant M$ et en faisant tendre n vers $+\infty$, on obtient $\sum_{k=0}^{N} b_k x^k \leqslant M$. Ainsi les sommes partielles de la série à termes positifs $\sum b_k x^k$ sont majorées donc cette série converge. Le rayon de convergence de $\sum b_k x^k$ est supérieur ou égal à 1. Posons $g(x) = \sum_{k=0}^{+\infty} b_k x^k$. Il reste à montrer que $g(x) = f(x)$ pour tout $x \in [0,1[$.

Soit $x \in [0,1[$ et $N \in \mathbb{N}$. On a pour tout $(n,k) \in \mathbb{N}^2$, $0 \leqslant b_{\varphi(n),k} \leqslant M$ et par passage à la limite $0 \leqslant b_k \leqslant M$. On en déduit

$$0 \leqslant \sum_{k=N}^{+\infty} b_{\varphi(n),k} x^k \leqslant \sum_{k=N}^{+\infty} M x^k \leqslant \frac{M x^N}{1-x},$$

et de même $0 \leqslant \sum_{k=N}^{+\infty} b_k x^k \leqslant \frac{M x^N}{1-x}$. On obtient ainsi

$$|P_{r(n)}(x) - g(x)| \leqslant \sum_{k=0}^{N-1} |b_{\rho(n),k} - b_k| x^k + \frac{2Mx^N}{1-x}.$$

En faisant tendre n vers $+\infty$, on en déduit

$$|f(x) - g(x)| \leqslant \frac{2Mx^N}{1-x}.$$

Quand N tend vers $+\infty$, $\dfrac{2Mx^N}{1-x}$ tend vers 0 donc $f(x) = g(x)$. ◁

En revanche, on n'a pas nécessairement $f(1) = g(1)$. Prenons par exemple, pour $n \in \mathbb{N}$, $P_n = \dfrac{1}{n+1} \sum_{k=0}^{n} x^l$. On a $P_n(x) = \dfrac{1-x^{n+1}}{(n+1)(1-x)}$ pour $x \in [0,1[$ et donc $f(x) = 0$. Comme $P_n(1) = 1$ pour tout $n \in \mathbb{N}$, $f(1) = 1$. Pour $(k,n) \in \mathbb{N}^2$, $b_{n,k} = \dfrac{1}{n+1}$ et donc $b_k = 0$ pour tout $k \in \mathbb{N}$. La fonction g est la fonction nulle et $f(1) \neq g(1)$.

Sont regroupés maintenant un ensemble d'exercices où il s'agit de trouver la limite ou un équivalent d'une fonction développable en série entière, en un point du cercle de convergence. Une première méthode, mise en œuvre dans l'exercice suivant, est la comparaison série-intégrale.

3.22. Recherche d'un équivalent

Trouver un équivalent quand x tend vers 1 de $f(x) = \sum_{n=0}^{+\infty} x^{2^n}$.

(École polytechnique)

▷ **Solution.**

Le rayon de convergence de la série entière $\sum x^{2^n}$ est 1 : la fonction f est donc définie sur $]-1,1[$. Pour déterminer un équivalent de f en 1, nous utiliserons la méthode de comparaison série-intégrale.

Soit $x \in [0,1[$; la fonction $t \longmapsto x^{2^t} = e^{2^t \ln x}$ est positive et décroissante sur $\mathbb{R}$, car $\ln x \leqslant 0$. On a donc, pour tout $k \in \mathbb{N}$,

$$x^{2^{k+1}} \leqslant \int_k^{k+1} x^{2^t}\,dt \leqslant x^{2^k}.$$

Pour $n \in \mathbb{N}$, en additionnant les inégalités obtenues en faisant varier k de 0 à n, on obtient

$$\sum_{k=1}^{n+1} x^{2^k} \leqslant \int_0^{n+1} x^{2^t}\,dt \leqslant \sum_{k=0}^{n} x^{2^k} \leqslant f(x).$$

On en déduit l'existence de $\int_{\mathbb{R}_+} x^{2^t}\mathrm{d}t$. En faisant tendre n vers $+\infty$, on obtient

$$f(x) - x \leqslant \int_0^{+\infty} x^{2^t}\mathrm{d}t \leqslant f(x).$$

c'est-à-dire

$$\int_0^{+\infty} x^{2^t}\mathrm{d}t \leqslant f(x) \leqslant \int_0^{+\infty} x^{2^t}\mathrm{d}t + x. \qquad (*)$$

On va chercher un équivalent de l'intégrale. Pour ce faire on va, en utilisant le changement de variable $u = -(\ln x)e^{t\ln 2}$, placer la variable x dans les bornes. On obtient

$$\int_0^{+\infty} x^{2^t}\mathrm{d}t = \int_0^{+\infty} e^{(\ln x)e^{t\ln 2}}\mathrm{d}t = \frac{1}{\ln 2}\int_{-\ln x}^{+\infty} \frac{e^{-u}}{u}\mathrm{d}u.$$

Il faut trouver un équivalent de $\int_{\ln x}^{+\infty} \frac{e^{-u}}{u}\mathrm{d}u$. Quand x tend vers 1, $\ln x$ tend vers 0. La fonction $u \longmapsto \dfrac{e^{-u}}{u}$ est intégrable sur $[1, +\infty[$, mais pas sur $]0, 1]$. De plus, on a $\dfrac{e^{-u}}{u} \underset{u\to 0}{\sim} \dfrac{1}{u}$. Ces fonctions étant de signe constant et pas intégrables sur $]0, 1]$, on en déduit que

$$\int_x^1 \frac{e^{-u}}{u}\mathrm{d}u \underset{x\to 0}{\sim} \int_x^1 \frac{1}{u}\mathrm{d}u = -\ln x \text{ et } \int_{-\ln x}^1 \frac{e^{-u}}{u}\mathrm{d}u \underset{x\to 1}{\sim} -\ln(-\ln x).$$

On obtient enfin $\displaystyle\int_{-\ln x}^{+\infty} \frac{e^{-u}}{u}\mathrm{d}u \underset{x\to 1}{\sim} -\ln(-\ln x)$ et donc

$$\int_0^{+\infty} x^{2^t}\mathrm{d}t \underset{x\to 1}{\sim} -\frac{\ln(-\ln x)}{\ln 2}.$$

Les inégalités $(*)$ montrent alors que $f(x) \underset{x\to 1}{\sim} -\dfrac{\ln(-\ln x)}{\ln 2}$. En écrivant, pour tout $x \in]0, 1]$,

$$\ln(-\ln x) = \ln\left(\frac{-\ln x}{1-x}\right) + \ln(1-x),$$

on en déduit que

$$\boxed{f(x) \underset{x\to 1}{\sim} -\frac{\ln(1-x)}{\ln 2}}. \quad \triangleleft$$

Les exercices suivants fournissent une autre méthode pour obtenir des équivalents de la somme d'une série entière. Ils affirment, sous certaines

conditions, que si $a_n \underset{n\to\infty}{\sim} b_n$, *alors* $\sum_{n=0}^{+\infty} a_n x^n \sim \sum_{n=0}^{+\infty} b_n x^n$, *l'équivalence ayant lieu au bord du domaine de convergence. L'énoncé suivant regroupe des questions provenant de plusieurs oraux.*

3.23. Étude asymptotique au bord du disque de convergence

Soient (a_n) et (b_n) deux suites de nombres complexes. On note R et R′ les rayons de convergence respectifs de $\sum a_n x^n$ et $\sum b_n x^n$ et, sous réserve de convergence, $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ et $g(x) = \sum_{n=0}^{+\infty} b_n x^n$. On suppose que (b_n) est à valeurs dans $\mathbb{R}_+^*$ et qu'il existe $\ell \in \mathbb{C}$ tel que $\lim_{n\to+\infty} \dfrac{a_n}{b_n} = \ell$.

1. Montrer que $R \geqslant R'$.
2. On suppose que $R' = 1$ et que $\sum b_n$ diverge. Montrer que
$$\lim_{x\to 1^-} \frac{f(x)}{g(x)} = \ell.$$
3. On suppose que $R' = +\infty$. Montrer que $\lim_{x\to+\infty} \dfrac{f(x)}{g(x)} = \ell$.
4. Soit $p \in \mathbb{N}$. Déterminer la limite lorsque x tend vers 1^- de
$$(1-x)^{p+1} \sum_{n=0}^{+\infty} n^p x^n.$$

(École polytechnique)

▷ **Solution.**

1. La suite $\left(\dfrac{a_n}{b_n}\right)$ converge donc elle est bornée. Soit M tel que, pour tout $n \in \mathbb{N}$, $\left|\dfrac{a_n}{b_n}\right| \leqslant M$. Pour tout $x \in \mathbb{C}$, on a $|a_n||x|^n \leqslant M|b_n||x|^n$. Si $|x| < R'$, la série $\sum b_n x^n$ converge, donc $\sum a_n x^n$ converge aussi et on a $R \geqslant R'$.

2. Les deux séries ont, d'après la première question, un rayon de convergence $\geqslant 1$. On remarque, pour commencer, que $\lim_{x\to 1^-} g(x) = +\infty$. En effet, pour tout $A > 0$, il existe $n_1 \in \mathbb{N}$ tel que $\sum_{n=0}^{n_1} b_n \geqslant 2A$, puisque $\sum b_n$ diverge et est à termes positifs. On a donc $\lim_{x\to 1^-} \sum_{n=0}^{n_1} b_n x^n \geqslant 2A$ et

il existe $x_0 \in]0,1[$ tel que, si $x \in [x_0, 1[$, on a $\sum_{n=0}^{n_1} b_n x^n \geqslant A$ et *a fortiori* $g(x) \geqslant A$ (car $b_n > 0$).

Soit $\varepsilon > 0$. Par hypothèse, il existe $n_0 \in \mathbb{N}$ tel que, pour $n \geqslant n_0$, on ait $\left|\frac{a_n}{b_n} - \ell\right| \leqslant \varepsilon$ et donc $|a_n - \ell b_n| \leqslant \varepsilon b_n$. On en déduit que, pour $x \in]0,1[$, on a

$$\left|\sum_{n=n_0}^{+\infty} a_n x^n - \ell \sum_{n=n_0}^{+\infty} b_n x^n\right| \leqslant \varepsilon \sum_{n=n_0}^{\infty} b_n x^n \leqslant \varepsilon g(x),$$

puis

$$|f(x) - \ell g(x)| \leqslant \sum_{n=0}^{n_0-1} |a_n - \ell b_n| x^n + \varepsilon g(x) \leqslant \sum_{n=0}^{n_0-1} |a_n - \ell b_n| + \varepsilon g(x).$$

Enfin, on obtient

$$\left|\frac{f(x)}{g(x)} - \ell\right| \leqslant \frac{\sum_{n=0}^{n_0-1} |a_n - \ell b_n|}{g(x)} + \varepsilon.$$

Comme $\lim_{x \to 1^-} \frac{\sum_{n=0}^{n_0-1} |a_n - \ell b_n|}{g(x)} + \varepsilon = \varepsilon$, il existe $x_1 \in]0,1[$ tel que, si $x \in [x_1, 1[$, alors $\left|\frac{f(x)}{g(x)} - \ell\right| \leqslant 2\varepsilon$, ce qui est le résultat voulu.

3. D'après la question 1. les deux séries entières ont un rayon de convergence infini. On choisit n_0 comme dans la question 2. La démonstration est la même. On obtient, pour tout réel $x \geqslant 1$,

$$\left|\frac{f(x)}{g(x)} - \ell\right| \leqslant \frac{\sum_{n=0}^{n_0-1} |a_n - \ell b_n| x^n}{g(x)} + \varepsilon \leqslant \frac{\left(\sum_{n=0}^{n_0-1} |a_n - \ell b_n|\right) x^{n_0-1}}{g(x)} + \varepsilon.$$

On remarque que $\lim_{x \to +\infty} \frac{x^{n_0-1}}{g(x)} = 0$, car $g(x) \geqslant b_{n_0} x^{n_0}$. Le majorant tend donc vers ε lorsque $x \to +\infty$, d'où l'on déduit l'existence de $x_1 > 0$ tel que $\left|\frac{f(x)}{g(x)} - \ell\right| \leqslant 2\varepsilon$, pour $x \geqslant x_1$.

4. On applique le résultat de la question 2. Le rayon de convergence de $\sum n^p x^n$ est 1. Si le calcul de $g(x) = \sum_{n=0}^{+\infty} n^p x^n$ n'est pas simple, on sait en revanche calculer $\sum_{n=0}^{+\infty} n(n-1)\ldots(n-p+1)x^{n-p}$, somme dans laquelle on reconnaît la dérivée p-ième de la série géométrique. En posant $h(x) = \sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$, on obtient donc, pour $p \in \mathbb{N}$ et $x \in]-1,1[$,

$$f(x) = \sum_{n=0}^{+\infty} n(n-1)\dots(n-p+1)x^n = x^p h^{(p)}(x) = x^p \frac{p!}{(1-x)^{p+1}}.$$

Si on pose $a_n = n(n-1)\dots(n-p+1)$ et $b_n = n^p$, la suite (b_n) est à termes strictement positifs, $\lim_{n\to+\infty} \dfrac{a_n}{b_n} = 1$ et $\sum b_n$ diverge. On en déduit que $\lim_{x\to1^-} \dfrac{f(x)}{g(x)} = 1$, c'est-à-dire que $f(x) \underset{x\to1^-}{\sim} g(x) \underset{x\to1^-}{\sim} \dfrac{p!}{(1-x)^{p+1}}$ et finalement

$$\boxed{\lim_{x\to1^-} (1-x)^{p+1} \sum_{n=0}^{+\infty} n^p x^n = p!}. \triangleleft$$

Avec les hypothèses de l'énoncé, il résulte en particulier de cet exercice que si $a_n \underset{n\to\infty}{\sim} b_n$ (resp. $a_n = o(b_n)$ en $+\infty$), on obtient en 1 ou en $+\infty$, $f \sim g$ (resp. $f = o(g)$).

Le résultat de la question 4 peut être obtenu également par comparaison à une intégrale, en opérant comme dans l'exercice 3.22. On trouve que $\sum_{n=0}^{+\infty} n^p x^n \underset{x\to1^-}{\sim} \int_{\mathbb{R}_+} t^p x^t\,dt$. On montre facilement, par récurrence sur p, que $\int_{\mathbb{R}_+} t^p x^t\,dt = \dfrac{p!}{(-\ln x)^{p+1}} \underset{x\to1^-}{\sim} \dfrac{p!}{(1-x)^{p+1}}$, ce qui conduit au résultat.

La première question de l'exercice suivant a été résolue d'une autre manière dans l'exercice 1.16 de notre tome 1 d'algèbre (il s'agissait de la démonstration même de Gauss). La suite de l'exercice est une application du résultat de l'exercice 3.23.

3.24. Un théorème de Gauss

On considère le plan $\mathbb{R}^2$ muni de sa norme euclidienne canonique $\|\ \|$. Pour $x \geqslant 0$, on pose

$$q(x) = \mathrm{Card}\{a \in \mathbb{Z}^2,\ \|a\| \leqslant x\} \quad \text{et} \quad q_+(x) = \mathrm{Card}\{a \in \mathbb{N}^2,\ \|a\| \leqslant x\}.$$

1. Trouver un équivalent de $q(x)$ et de $q_+(x)$ lorsque $x \to +\infty$.

2. On pose pour $t \in\ ]-1,1[$, $g(t) = \dfrac{1}{1-t}\left(\sum_{n=0}^{+\infty} t^{n^2}\right)^2$. Relier g et q_+ puis donner un équivalent de g en 1^-.

(**École normale supérieure**)

▷ **Solution.**

1. Pour tout $x \geqslant 0$, $q(x)$ représente le nombre de points à coordonnées entières dans le disque de centre 0 et de rayon x et $q_+(x)$ le nombre de points à coordonnées entières $\geqslant 0$ dans ce même disque. De manière intuitive, il y a environ 1 point à coordonnées entières par cm^2 (si on prend le centimètre comme unité de longueur) et il est donc assez raisonnable de penser que lorsque x tend vers $+\infty$, $q(x)$ va être équivalent à πx^2 (l'aire du disque euclidien de rayon x).

On commence par chercher un équivalent de $q_+(n)$ lorsque $n \in \mathbb{N}$ tend vers l'infini. L'idée est de compter le nombre de points de $\mathbb{Z}^2$ dans le quart de disque de rayon n en les comptant colonne par colonne.

Soit n un entier naturel. Pour (k,l) dans $\mathbb{N}^2$, l'inégalité $\|(k,l)\| \leqslant n$, qui s'écrit $k^2+l^2 \leqslant n^2$, équivaut à $0 \leqslant k \leqslant n$ et $0 \leqslant l \leqslant \sqrt{n^2-k^2}$. En comptant le nombre de couples (k,l) vérifiant cette inégalité pour k fixé, on obtient

$$q_+(n) = \sum_{k=0}^{n} \sum_{0 \leqslant l \leqslant \sqrt{n^2-k^2}} 1 = \sum_{k=0}^{n} \left(\mathrm{E}(\sqrt{n^2-k^2}) + 1 \right).$$

On a donc l'encadrement

$$\sum_{k=0}^{n} \sqrt{n^2-k^2} \leqslant q_+(n) \leqslant \sum_{k=0}^{n} \sqrt{n^2-k^2} + n + 1,$$

ce qui donne en divisant par n^2

$$\frac{1}{n} \sum_{k=0}^{n} \sqrt{1-\frac{k^2}{n^2}} \leqslant \frac{q_+(n)}{n^2} \leqslant \frac{1}{n} \sum_{k=0}^{n} \sqrt{1-\frac{k^2}{n^2}} + \frac{1}{n} + \frac{1}{n^2}.$$

Nous reconnaissons alors une somme de Riemann relative à la subdivision régulière du segment $[0,1]$ de pas $\dfrac{1}{n}$ pour la fonction $t \longmapsto \sqrt{1-t^2}$. On en déduit que

$$\frac{1}{n} \sum_{k=0}^{n} \sqrt{1-\frac{k^2}{n^2}} = \frac{1}{n} \sum_{k=0}^{n-1} \sqrt{1-\frac{k^2}{n^2}} \xrightarrow[n \to +\infty]{} \int_0^1 \sqrt{1-t^2}\,\mathrm{d}t.$$

Cette intégrale représente, comme prévu, l'aire d'un quart de disque de rayon 1 et vaut donc $\dfrac{\pi}{4}$. Par le calcul, cela se voit en posant $t = \sin\theta$:

$$\int_0^1 \sqrt{1-t^2}\,\mathrm{d}t = \int_0^{\pi/2} \sqrt{1-\sin^2\theta}\cos\theta\,\mathrm{d}\theta = \int_0^{\pi/2} \cos^2\theta\,\mathrm{d}\theta$$
$$= \int_0^{\pi/2} \frac{1+\cos(2\theta)}{2}\,\mathrm{d}\theta = \frac{\pi}{4}.$$

En vertu de l'inégalité établie et du théorème d'encadrement, on en déduit que

$$\lim_{n\to+\infty} \frac{q_+(n)}{n^2} = \frac{\pi}{4}.$$

Il est facile de voir que cet équivalent reste vrai pour un réel positif x quelconque : en effet, si x est un réel positif et $n = \mathrm{E}(x^2)$, on a $q_+(n) \leqslant q_+(x) \leqslant q_+(n+1)$ et donc, en divisant par x^2

$$\frac{q_+(n)}{x^2} \leqslant \frac{q_+(x)}{x^2} \leqslant \frac{q_+(n+1)}{x^2}, \text{ puis } \frac{q_+(n)}{(n+1)^2} \leqslant \frac{q_+(x)}{x^2} \leqslant \frac{q_+(n+1)}{n^2}.$$

Or, d'après ce qui précède, $\lim_{n\to+\infty} \dfrac{q_+(n)}{(n+1)^2} = \lim_{n\to+\infty} \dfrac{q_+(n+1)}{n^2} = \dfrac{\pi}{4}$.

Toujours d'après le théorème d'encadrement, on a $\lim_{n\to+\infty} \dfrac{q_+(x)}{x^2} = \dfrac{\pi}{4}$.

On conclut que

$$\boxed{q_+(x) \underset{x\to+\infty}{\sim} \frac{\pi x^2}{4}}.$$

On calcule $q(x)$ à partir de $q_+(x)$ en faisant attention aux points sur les axes qui sont comptés deux fois et au point $(0,0)$ compté 4 fois. On obtient $q(x) = 4q_+(x) - 4\mathrm{E}(x) - 3$ et

$$\boxed{q(x) \underset{x\to+\infty}{\sim} \pi x^2}.$$

2. Comme la série $\sum t^{n^2}$ est de rayon de convergence 1, g est définie sur $]-1,1[$. On va calculer g^2 en effectuant le produit de Cauchy de la série entière $\sum t^{n^2}$ par elle-même. Nous savons alors que l'on obtient une série entière de rayon au moins 1. Plus précisément, si on pose, pour $n \geqslant 0$, $a_n = \sum_{\substack{\mathrm{K}+\mathrm{L}=n \\ \mathrm{K,L\ carrés}}} 1 = \sum_{\substack{k^2+l^2=n \\ 0\leqslant k,l\leqslant n}} 1$, alors, on a, pour $t \in]-1,1[$.

$$\left(\sum_{n=0}^{+\infty} t^{n^2}\right)^2 = \sum_{n=0}^{+\infty} a_n t^n \quad \text{et} \quad g(t) = \frac{1}{1-t} \sum_{n=0}^{+\infty} a_n t^n.$$

Si $n \geqslant 0$, a_n est alors le nombre de couples $(k,l) \in \mathbb{N}^2$ tels que $k^2+l^2 = n$. Toujours d'après les règles du produit de Cauchy, on a, pour $t \in]-1,1[$,

$$\begin{aligned} g(t) &= \left(\sum_{n=0}^{+\infty} t^n\right)\left(\sum_{n=0}^{+\infty} a_n t^n\right) = \sum_{n=0}^{+\infty} (a_0 + a_1 + \cdots + a_n) t^n \\ &= \sum_{n=0}^{+\infty} q_+(\sqrt{n}) t^n. \end{aligned}$$

Pour obtenir un équivalent de $g(t)$ en 1^-, on va remplacer $q_+(\sqrt{n})$ par son équivalent trouvé en 1. Ceci est possible en vertu du lemme suivant :

Lemme. *Soient $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites de $\mathbb{R}_+$ équivalentes. On suppose que $r(t) = \sum_{n=0}^{+\infty} a_n t^n$ est de rayon 1 et que $\sum a_n$ diverge.*

Alors $\displaystyle\sum_{n=0}^{+\infty} b_n t^n \underset{t\to 1^-}{\sim} s(t) \underset{t\to 1^-}{\sim} r(t).$

Ce résultat classique fait l'objet de l'exercice 3.23. On obtient donc,

$$g(t) \underset{t\to 1^-}{\sim} \sum_{n=0}^{+\infty} \frac{\pi}{4} n t^n = \frac{\pi t}{4} \sum_{n=1}^{+\infty} n t^{n-1} = \frac{\pi t}{4} \frac{\mathrm{d}}{\mathrm{d}t} \frac{1}{1-t} = \frac{\pi t}{(1-t)^2}.$$

On conclut que

$$\boxed{g(t) \underset{t\to 1^-}{\sim} \frac{\pi}{1} \frac{1}{(1-t)^2}}. \quad \triangleleft$$

On en déduit que $\displaystyle\sum_{n=0}^{+\infty} t^{n^2} \underset{t\to 1^-}{\sim} \frac{\sqrt{\pi}}{2\sqrt{1-t}}$. *Notons que cela aurait aussi pu être établi, comme dans l'exercice 3.22, à l'aide de la méthode de comparaison série-intégrale en reliant* $\displaystyle\sum_{n=0}^{+\infty} t^{n^2}$ *et* $\int_0^{+\infty} t^{u^2}\,du$. *On trouve alors*

$$\sum_{n=0}^{+\infty} t^{n^2} \underset{t\to 1}{\sim} \frac{1}{\sqrt{-\ln t}} \int_0^{+\infty} e^{-v^2}\,dv \underset{t\to 1}{\sim} \frac{\sqrt{\pi}}{2\sqrt{1-t}}.$$

De cela on tire directement l'équivalent de g.

3.25. Série de Lambert

Soit $\alpha \in \mathbb{N}^*$ et $x \in [0,1[$.

1. Montrer que la série $\sum \frac{n^\alpha x^n}{1-x^n}$ converge.

2. Donner un équivalent de sa somme $\varphi(x) = \displaystyle\sum_{n=1}^{+\infty} \frac{n^\alpha x^n}{1-x^n}$, lorsque $x \to 1^-$ (on exprimera $\varphi(x)$ comme somme d'une série double et on considérera le développement en série entière de $\frac{\varphi(x)}{1-x}$).

(**École polytechnique**)

Solution.

1. Pour $x \in [0,1[$, le terme général $\frac{n^\alpha x^n}{1-x^n}$ de la série est équivalent à $n^\alpha x^n = o\left(\frac{1}{n^2}\right)$ lorsque n tend vers l'infini, donc la série converge, d'après le théorème de comparaison des séries à termes positifs.

2. On peut écrire, pour $n \geqslant 1$,

$$\frac{x^n}{1-x^n} = x^n \sum_{k=0}^{+\infty} (x^n)^k = \sum_{k=1}^{+\infty} x^{nk}.$$

D'après les théorèmes sur les familles sommables positives (le lecteur pourra se reporter à notre premier tome d'analyse p. 201), la famille $(n^\alpha x^{nk})_{n,k\geqslant 1}$ est sommable et sa somme est $\varphi(x)$. Le théorème d'associativité permet de rassembler les termes selon la valeur p du produit nk. On obtient

$$\varphi(x) = \sum_{p=1}^{+\infty} \left(\sum_{n|p} n^\alpha x^p \right) = \sum_{p=1}^{+\infty} \left(\sum_{n|p} n^\alpha \right) x^p.$$

Posons $a_p = \sum_{n|p} n^\alpha$ pour $p \geqslant 1$. Pour trouver un équivalent de $\varphi(x)$, on souhaiterait avoir un équivalent de a_p, mais ceci n'apparaît pas facile. Par exemple, pour $\alpha = 1$, a_p est la somme des diviseurs de p : c'est une suite qui varie de manière assez irrégulière. Souvent, ces fonctions arithmétiques prennent un comportement beaucoup plus régulier lorsqu'on les étudie en moyenne (le lecteur en trouvera un autre exemple dans l'exercice 3.26 du tome 1 d'analyse). C'est pour cela qu'on va, comme le préconise l'énoncé, faire le produit de Cauchy de $\varphi(x)$ avec $\dfrac{1}{1-x}$. D'après la règle sur le produit de Cauchy, on obtient, pour $x \in [0,1[$,

$$\frac{\varphi(x)}{1-x} = \sum_{p=1}^{+\infty} (a_1 + \cdots + a_p) x^p = \sum_{p=1}^{+\infty} b_p x^p$$

avec

$$b_p = a_1 + \cdots + a_p = \sum_{k=1}^{p} \sum_{n|k} n^\alpha = \sum_{n=1}^{p} n^\alpha \sum_{n|k} 1 = \sum_{n=1}^{p} \mathrm{E}\left(\frac{p}{n}\right) n^\alpha.$$

La suite b_p est beaucoup plus facile à estimer que la suite a_p. En effet,

$$\frac{b_p}{p^{\alpha+1}} = \frac{1}{p} \sum_{n=1}^{p} \mathrm{E}\left(\frac{p}{n}\right) \left(\frac{n}{p}\right)^\alpha$$

est une somme de Riemann associée à la fonction $x \longmapsto \mathrm{E}\left(\dfrac{1}{x}\right) x^\alpha$. Cette fonction admet en tout point de $[0,1]$ une limite à gauche et à droite (rappelons que $\alpha \geqslant 1$). Elle est donc réglée, c'est-à-dire limite uniforme de fonctions en escalier. La somme de Riemann ci-dessus converge donc vers l'intégrale de cette fonction entre 0 et 1. Calculons cette intégrale :

$$\begin{aligned}\int_0^1 \mathrm{E}\left(\frac{1}{x}\right) x^\alpha \mathrm{d}x &= \sum_{n=1}^{+\infty} \int_{\frac{1}{n+1}}^{\frac{1}{n}} \mathrm{E}\left(\frac{1}{x}\right) x^\alpha \mathrm{d}x \\ &= \sum_{n=1}^{+\infty} n \int_{\frac{1}{n+1}}^{\frac{1}{n}} x^\alpha \mathrm{d}x \\ &= \frac{1}{\alpha+1} \sum_{n=1}^{+\infty} n \left(\frac{1}{n^{\alpha+1}} - \frac{1}{(n+1)^{\alpha+1}}\right).\end{aligned}$$

Si $\mathrm{N} \geqslant 1$, on peut écrire

$$\begin{aligned}\sum_{n=1}^{\mathrm{N}} n \left(\frac{1}{n^{\alpha+1}} - \frac{1}{(n+1)^{\alpha+1}}\right) &= \sum_{n=1}^{\mathrm{N}} \frac{1}{n^\alpha} - \sum_{n=1}^{\mathrm{N}} \frac{n+1-1}{(n+1)^{\alpha+1}} \\ &= 1 - \frac{1}{(\mathrm{N}+1)^\alpha} + \sum_{n=1}^{\mathrm{N}} \frac{1}{(n+1)^{\alpha+1}} \\ &= \sum_{n=1}^{\mathrm{N}+1} \frac{1}{n^{\alpha+1}} - \frac{1}{(\mathrm{N}+1)^\alpha}.\end{aligned}$$

et en faisant tendre N vers $+\infty$, il vient finalement

$$\int_0^1 \mathrm{E}\left(\frac{1}{x}\right) x^\alpha \mathrm{d}x = \frac{1}{\alpha+1} \sum_{n=1}^{+\infty} \frac{1}{n^{\alpha+1}} = \frac{\zeta(\alpha+1)}{\alpha+1} \neq 0,$$

où ζ désigne la fonction zêta de Riemann définie pour tout $x > 1$ par $\zeta(x) = \sum_{n=1}^{+\infty} \frac{1}{n^x}$. On obtient donc $b_p \underset{p\to\infty}{\sim} \frac{\zeta(\alpha+1)}{\alpha+1} p^{\alpha+1}$, ce qui est encore équivalent à

$$\frac{\zeta(\alpha+1)}{\alpha+1}(p+\alpha+1)(p+\alpha)\dots(p+1).$$

D'après un résultat classique sur les équivalents de séries entières à coefficients positifs (on pourra se reporter à l'exercice 3.23), on en déduit que

$$\begin{aligned}\frac{\varphi(x)}{1-x} = \sum_{p=1}^{+\infty} b_p x^p &\underset{x\to 1^-}{\sim} \frac{\zeta(\alpha+1)}{\alpha+1} \sum_{p=0}^{+\infty} (p+\alpha+1)(p+\alpha)\dots(p+1)x^p \\ &\underset{x\to 1^-}{\sim} \frac{\zeta(\alpha+1)}{\alpha+1} \frac{(\alpha+1)!}{(1-x)^{\alpha+2}} \text{ (on dérive } \frac{1}{1-x} \ \alpha+1 \text{ fois)} \\ &\underset{x\to 1^-}{\sim} \frac{\alpha!\zeta(\alpha+1)}{(1-x)^{\alpha+2}}.\end{aligned}$$

On conclut

$$\boxed{\varphi(x) \underset{x\to 1^-}{\sim} \frac{\alpha!\zeta(\alpha+1)}{(1-x)^{\alpha+1}}}. \quad \triangleleft$$

Les théorèmes taubériens qui suivent énoncent des conditions sous lesquelles une réciproque du théorème d'Abel-Dirichlet (voir page 178) est valide. En effet, celle-ci est fausse en général. Si on considère la suite de terme général $a_n = (-1)^n$, la série $\sum a_n x^n$ a pour rayon de convergence 1, sa somme, égale à $\dfrac{1}{x+1}$, a pour limite $\dfrac{1}{2}$ quand x tend vers 1, sans que la série $\sum a_n$ converge. Nous avons regroupé ensemble plusieurs exercices posés à l'école polytechnique, avec des hypothèses variées. Notons que la même conclusion reste vérifiée avec l'hypothèse plus faible $a_n = O\left(\dfrac{1}{n}\right)$ (théorème taubérien de Hardy-Littlewood).

3.26. Théorèmes taubériens

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de $\mathbb{C}$ telle que $f(x) = \sum_{n=0}^{+\infty} a_n x^n$ existe pour tout $x \in]-1,1[$ et $\lim_{x\to 1^-} f(x) = \ell \in \mathbb{C}$. Démontrer que la série $\sum a_n$ converge et a pour somme ℓ dans les différents cas suivants :

1. $a_n \geqslant 0$ pour tout $n \in \mathbb{N}$;
2. $a_n = o\left(\dfrac{1}{n}\right)$ lorsque n tend vers l'infini ;
3. $\sum_{n=0}^{+\infty} n|a_n|^2$ converge (dans ce cas le résultat est dû à Fejér) ;
4. $\lim_{n\to\infty} \dfrac{a_1 + 2a_2 + \cdots + na_n}{n} = 0$.

(École polytechnique)

▷ **Solution.**

1. Dans ce cas, la fonction f est croissante sur $[0,1[$. On a donc, pour tout $x \in [0,1[$, $\sum_{n=0}^{+\infty} a_n x^n \leqslant \ell$ et, *a fortiori*, pour tout $N \in \mathbb{N}$, $\sum_{n=0}^{N} a_n x^n \leqslant \ell$. En faisant tendre x vers 1, on obtient $\sum_{n=0}^{N} a_n \leqslant \ell$. Ainsi $\sum a_n$ converge et sa somme est $\leqslant \ell$. Mais, d'autre part, $\sum_{n=0}^{+\infty} a_n$ majore f sur $[0,1]$, ce qui implique $\ell \leqslant \sum_{n=0}^{+\infty} a_n$ et l'égalité voulue.

2. Soit $\varepsilon > 0$. Il s'agit de prouver qu'on a, pour N assez grand, $\delta_N = \left|\ell - \sum_{n=0}^{N} a_n\right| \leqslant \varepsilon$. Pour ce faire, on effectue une découpe taubérienne ; on introduit $f(x)$ en choisissant x convenablement (en fonction de N) pour contrôler chaque terme. Soit donc $x \in [0,1[$ et

$$\delta_N = \left|\ell - f(x) + f(x) - \sum_{n=0}^{N} a_n\right|$$
$$= \left|\ell - f(x) + \sum_{n=N+1}^{+\infty} a_n x^n + \sum_{n=0}^{N} a_n(x^n - 1)\right|.$$

D'après l'inégalité triangulaire, la série entière convergeant absolument sur $[0, 1[$, on peut écrire

$$\begin{aligned}
\delta_N &\leqslant |\ell - f(x)| + \sum_{n=N+1}^{+\infty} |a_n| x^n + \sum_{n=0}^{N} |a_n|(1 - x^n) \\
&\leqslant |\ell - f(x)| + \sum_{n=N+1}^{+\infty} \frac{n|a_n|}{N} x^n + (1 - x) \sum_{n=0}^{N} |a_n|(1 + x + \cdots + x^{n-1}) \\
&\leqslant |\ell - f(x)| + \frac{1}{N} \sup_{n \geqslant N} |na_n| \sum_{n=N+1}^{+\infty} x^n + (1 - x) \sum_{n=0}^{N} n|a_n| \\
&\leqslant |\ell - f(x)| + \frac{1}{N} \sup_{n \geqslant N} |na_n| \frac{x^{N+1}}{1 - x} + (1 - x) \sum_{n=0}^{N} n|a_n| \\
&\leqslant |\ell - f(x)| + \frac{1}{N(1 - x)} \sup_{n \geqslant N} |na_n| + (1 - x)N \left(\frac{1}{N} \sum_{n=0}^{N} n|a_n|\right).
\end{aligned}$$

Par hypothèse, $\sup_{n \geqslant N} n|a_n| \xrightarrow[N \to +\infty]{} 0$ et, en vertu du théorème de Cesàro,

$$\frac{1}{N} \sum_{n=0}^{N} n|a_n| \xrightarrow[N \to +\infty]{} 0.$$

Il serait bon d'avoir $(1 - x)N = 1$. Rien ne nous en empêche : il suffit de prendre $x = 1 - \dfrac{1}{N}$. On a donc

$$\delta_N \leqslant \left|\ell - f\left(1 - \frac{1}{N}\right)\right| + \sup_{n \geqslant N} |na_n| + \frac{1}{N} \sum_{n=0}^{N} n|a_n|.$$

Comme $\lim_{x \to 1} f(x) = \ell$, on a, pour N assez grand, $\delta_N \leqslant \varepsilon$.

3. Nous allons remettre en place la même idée, mais avec ces conditions, la majoration est plus technique. Tout d'abord, pour utiliser l'hypothèse dans nos majorations, nous allons naturellement introduire le reste de la série dont on sait qu'il tend vers 0 : pour $N \in \mathbb{N}$, on pose $R_N = \sum_{n=N}^{+\infty} n|a_n|^2$. Un calcul immédiat donne, pour $n \geqslant 1$,

$$|a_n| = \sqrt{\frac{R_n - R_{n+1}}{n}}.$$

En reprenant les notations de la question précédente, on a toujours pour $N \in \mathbb{N}$ et $x \in [0,1[$.

$$\delta_N \leqslant |\ell - f(x)| + \sum_{n=N+1}^{+\infty} |a_n| x^n + \sum_{n=0}^{N} |a_n|(1-x^n).$$

Occupons-nous du second terme. On écrit, à l'aide de l'inégalité de Cauchy-Schwarz,

$$\left(\sum_{n=N+1}^{+\infty} |a_n| x^n\right)^2 = \left(\sum_{n=N+1}^{+\infty} |a_n| \sqrt{n} \frac{x^n}{\sqrt{n}}\right)^2$$
$$\leqslant \underbrace{\left(\sum_{n=N+1}^{+\infty} n|a_n|^2\right)}_{=R_{N+1}} \left(\sum_{n=N+1}^{+\infty} \frac{x^{2n}}{n}\right).$$

On a clairement

$$\sum_{n=N+1}^{+\infty} \frac{x^{2n}}{n} \leqslant \frac{1}{N} \sum_{n=0}^{+\infty} (x^2)^n = \frac{1}{N(1-x^2)} = \frac{1}{N(1-x)(1+x)} \leqslant \frac{1}{N(1-x)}.$$

On obtient donc

$$\sum_{n=N+1}^{+\infty} |a_n| x^n \leqslant \sqrt{\frac{R_{N+1}}{N(1-x)}}.$$

Venons-en au troisième terme. On a, comme dans la question précédente,

$$\sum_{n=0}^{N} |a_n|(1-x^n) \leqslant (1-x) \sum_{n=1}^{N} n|a_n|.$$

On en déduit que

$$\sum_{n=0}^{N} |a_n|(1-x^n) \leqslant (1-x) \sum_{n=1}^{N} \sqrt{nR_n - nR_{n+1}}$$
$$\leqslant (1-x) \left(\sum_{n=1}^{N} (nR_n - nR_{n+1})\right)^{1/2} \left(\sum_{n=1}^{N} 1^2\right)^{1/2},$$

toujours d'après l'inégalité de Cauchy-Schwarz. Un changement d'indice donne

$$\sum_{n=1}^{N} (nR_n - nR_{n+1}) = \sum_{n=1}^{N} nR_n - \sum_{n=2}^{N+1} (n-1)R_n$$
$$= \sum_{n=1}^{N} R_n - NR_{N+1} \leqslant \sum_{n=1}^{N} R_n.$$

Finalement, on obtient la majoration

$$\sum_{n=0}^{N} |a_n|(1-x^n) \leqslant (1-x)\left(\sum_{n=1}^{N} \mathrm{R}_n\right)^{1/2} \sqrt{\mathrm{N}} \leqslant (1-x)\mathrm{N}\left(\frac{1}{\mathrm{N}}\sum_{n=1}^{N} \mathrm{R}_n\right)^{1/2}.$$

Là encore, tout nous pousse à prendre $x = 1 - \dfrac{1}{\mathrm{N}}$ (*i.e.* $\mathrm{N}(1-x) = 1$). Il vient alors

$$\delta_{\mathrm{N}} \leqslant \left|\ell - f\left(1 - \frac{1}{\mathrm{N}}\right)\right| + \sqrt{\mathrm{R}_{\mathrm{N}+1}} + \sqrt{\frac{1}{\mathrm{N}}\sum_{n=1}^{N} \mathrm{R}_n}.$$

Ces trois termes tendent vers 0 lorsque N tend vers l'infini, le premier par hypothèse et le troisième en vertu du théorème de Cesàro. On a bien $\lim\limits_{\mathrm{N}\to+\infty} \delta_{\mathrm{N}} = 0$.

4. Cette fois-ci, il est naturel de poser pour $n > 0$, $\mathrm{S}_n = \dfrac{1}{n}\sum\limits_{k=1}^{n} ka_k$ et $\mathrm{S}_0 = 0$, ce qui conduit, pour $n \geqslant 1$, à

$$a_n = \frac{n\mathrm{S}_n - (n-1)\mathrm{S}_{n-1}}{n} = \mathrm{S}_n - \frac{(n-1)}{n}\mathrm{S}_{n-1}.$$

On en déduit que, pour $\mathrm{N} \in \mathbb{N}$ et $x \in \mathbb{R}$,

$$\begin{aligned}\sum_{n=1}^{N} a_n x^n &= \sum_{n=1}^{N}\left(\mathrm{S}_n x^n - \frac{n-1}{n}\mathrm{S}_{n-1}x^n\right)\\ &= \sum_{n=1}^{N} \mathrm{S}_n x^n - \sum_{n=1}^{N-1}\frac{n}{n+1}\mathrm{S}_n x^{n+1}\\ &= \mathrm{S}_{\mathrm{N}}x^{\mathrm{N}} + \sum_{n=1}^{N-1}\left(1 - \frac{nx}{n+1}\right)\mathrm{S}_n x^n.\end{aligned}$$

Pour $x = 1$, on obtient $\sum\limits_{n=1}^{N} a_n = \mathrm{S}_{\mathrm{N}} + \sum\limits_{n=1}^{N-1}\dfrac{1}{n+1}\mathrm{S}_n$. En faisant tendre N vers l'infini, x étant dans $[0,1[$, il vient, puisque $\lim\limits_{\mathrm{N}\to+\infty} \mathrm{S}_{\mathrm{N}} = 0$,

$$f(x) = a_0 + \sum_{n=1}^{+\infty}\left(1 - \frac{nx}{n+1}\right)\mathrm{S}_n x^n.$$

Pour simplifier, on pose, pour $n \in \mathrm{N}$, $u_n = \left(1 - \dfrac{nx}{n+1}\right)\mathrm{S}_n$. Avec les mêmes notations que précédemment, on a, pour $x \in [0,1[$,

$$\begin{aligned}\delta_N &\leqslant |l - f(x)| + \left| f(x) - \sum_{n=0}^{N} a_n \right| \\ &\leqslant |l - f(x)| + \left| a_0 + \sum_{n=1}^{+\infty} a_n x^n - a_0 - \sum_{n=1}^{N} a_n \right| \\ &\leqslant |l - f(x)| + \sum_{n=N+1}^{+\infty} |a_n| x^n + \left| \sum_{n=1}^{N} a_n x^n - \sum_{n=1}^{N} a_n \right|.\end{aligned}$$

Dans le second terme qui est égal à $\sum_{N+1}^{+\infty} \left(1 - \frac{nx}{n+1}\right) S_n x^n$, on majore chaque $|S_n|$ par $\sup_{n \geqslant N+1} |S_n|$ et on écrit

$$\begin{aligned}\sum_{n=N+1}^{+\infty} \left(1 - \frac{nx}{n+1}\right) x^n &= \sum_{n=N+1}^{+\infty} \left((1-x)x^n + \frac{x^{n+1}}{n+1} \right) \\ &\leqslant \sum_{n=N+1}^{+\infty} (1-x)x^n + \frac{1}{N} \sum_{n=N+1}^{+\infty} x^{n+1} \\ &\leqslant x^{N+1} + \frac{x^{N+2}}{N(1-x)} \leqslant 1 + \frac{1}{N(1-x)},\end{aligned}$$

ce qui conduit à

$$\sum_{n=N+1}^{+\infty} |a_n| x^n \leqslant \sup_{n \geqslant N+1} |S_n| \left(1 + \frac{1}{N(1-x)}\right).$$

On exprime le troisième terme en fonction des S_n. On obtient

$$\begin{aligned}\left| \sum_{n=1}^{N} a_n x^n - \sum_{n=1}^{N} a_n \right| &= \left| \sum_{n=1}^{N} a_n x^n - S_N - \sum_{n=1}^{N-1} \frac{1}{n+1} S_n \right| \\ &= \left| \sum_{n=1}^{N} a_n x^n - \frac{N}{N+1} S_N - \sum_{n=1}^{N} \frac{1}{n+1} S_n \right| \\ &= \left| -\frac{N}{N+1} S_N + \sum_{n=1}^{N} \left(a_n - \frac{1}{n+1} S_n \right) x^n + \sum_{n=1}^{N} \frac{1}{n+1} S_n (x^n - 1) \right| \\ &= \left| -\frac{N}{N+1} S_N + \sum_{n=1}^{N} \frac{n}{n+1} (1-x) S_n x^n + \sum_{n=1}^{N} \frac{1}{n+1} S_n (x^n - 1) \right|.\end{aligned}$$

En remarquant, de nouveau, que $|(x^n - 1)| \leqslant n(1-x)$, on en déduit que

$$\left| \sum_{n=1}^{N-1} a_n x^n - \sum_{n=1}^{N} a_n \right| \leqslant |S_N| + 2(1-x) \sum_{n=1}^{N} |S_n|.$$

Finalement, on a

$$\delta_N \leqslant |\ell - f(x)| + \sup_{n\geqslant N+1} |S_n|\left(1 + \frac{1}{N(1-x)}\right) + |S_N| + 2(1-x)\sum_{n=1}^{N} |S_n|.$$

On est invité à prendre à nouveau $N(1-x) = 1$, *i.e.* $x = 1 - \dfrac{1}{N}$ et dans ces conditions

$$\begin{aligned}\delta_N &\leqslant \left|\ell - f\left(1 - \frac{1}{N}\right)\right| + 2\sup_{n\geqslant N+1} |S_n| + |S_N| + \frac{2}{N}\sum_{n=1}^{N} |S_n| \\ &\leqslant \left|\ell - f\left(1 - \frac{1}{N}\right)\right| + 3\sup_{n\geqslant N} |S_n| + \frac{2}{N}\sum_{n=1}^{N} |S_n|.\end{aligned}$$

En vertu du théorème de Cesàro, $\dfrac{1}{N}\sum_{n=1}^{N} |S_n|$ tend vers 0 puisque S_N tend vers 0 et finalement δ_N tend vers 0. Ce qu'on voulait. ◁

On peut noter que les questions 2 et 3 ne sont que des cas particuliers de la question 4. En effet, si $a_n = o\left(\dfrac{1}{n}\right)$ la suite (na_n) tend vers 0 et le théorème de Cesàro permet de dire que $\lim\limits_{n\to\infty} \dfrac{a_1 + \cdots + na_n}{n} = 0$. La convergence de $\sum\limits_{n=0}^{+\infty} n|a_n|^2$ implique également ce résultat. Pour le voir on utilise l'inégalité de Cauchy-Schwarz. En écrivant $ka_k = \sqrt{k}\sqrt{k}a_k$, on a la majoration

$$\left|\sum_{k=p+1}^{n} ka_k\right| \leqslant \sqrt{\sum_{k=p+1}^{n} k}\sqrt{\sum_{k=p+1}^{n} k|a_k|^2}.$$

On en déduit que pour p fixé et $n \geqslant p$,

$$\frac{|a_1 + \cdots + na_n|}{n} \leqslant \frac{|a_1 + \cdots + pa_p|}{n} + \sqrt{R_p}$$

où $R_p = \sum\limits_{k=p+1}^{+\infty} k|a_k|^2$ est le reste de la série supposée convergente. Si $\varepsilon > 0$ est donné, on choisit p de sorte que $\sqrt{R_p} \leqslant \varepsilon$. Pour n assez grand on a alors $\dfrac{|a_1 + \cdots + na_n|}{n} \leqslant 2\varepsilon$. Cela prouve bien que $\lim\limits_{n\to\infty} \dfrac{a_1 + \cdots + na_n}{n} = 0$.

3.27. Condition suffisante d'injectivité d'une série entière

Soient $\mathrm{D} = \{z \in \mathbb{C}, |z| < 1\}$ et $(a_n)_{n\in\mathbb{N}}$ une suite complexe. On suppose la série $\sum na_n$ absolument convergente.

1. Vérifier que le rayon de convergence de $\sum_{n=0}^{+\infty} a_n z^n$ est supérieur ou égal à 1. Pour $z \in \mathrm{D}$, on pose $f(z) = \sum_{n=0}^{+\infty} a_n z^n$.

2. On suppose que $a_1 \neq 0$ et $\sum_{n=2}^{+\infty} n|a_n| \leqslant |a_1|$. Montrer que f est injective.

(École polytechnique)

▷ **Solution.**

1. Comme la série $\sum na_n$ est absolument convergente, on a en particulier $\lim_{n\to+\infty} na_n = 0$. Donc, pour n assez grand, $n|a_n| \leqslant 1$ *i.e.* $|a_n| \leqslant \dfrac{1}{n}$. Comme la série entière de coefficient $\dfrac{1}{n}$ est de rayon de convergence égal à 1, celui de $\sum_{n=0}^{+\infty} a_n z^n$ est supérieur ou égal à 1.

2. La fonction f est bien définie d'après la question précédente. Montrons que f est injective en raisonnant par l'absurde. Supposons qu'il existe z et z' distincts dans D tel que $f(z) = f(z')$. On a

$$0 = f(z') - f(z) = \sum_{n=0}^{+\infty} a_n(z'^n - z^n).$$

En isolant le terme d'indice $n = 1$ (celui correspondant à $n = 0$ est nul), on obtient

$$a_1(z - z') = \sum_{n=2}^{+\infty} a_n(z'^n - z^n) = (z' - z)\sum_{n=2}^{+\infty} a_n\left(\sum_{k=0}^{n-1} z'^k z^{n-1-k}\right). \quad (*)$$

On peut supposer qu'un des a_n pour $n \geqslant 2$ est non nul car sinon, $f : z \longmapsto a_1 z$ est injective, puisque $a_1 \neq 0$. Pour $n \geqslant 2$ et $a_n \neq 0$, on a

$$\left|a_n \sum_{k=0}^{n-1} z'^k z^{n-1-k}\right| \leqslant |a_n| \sum_{k=0}^{n-1} |z'|^k |z|^{n-1-k} < |a_n| \sum_{k=0}^{n-1} 1 = n|a_n|,$$

car $|a_n| > 0$ et z et z' sont dans D. En passant au module dans $(*)$, il vient donc

$$|a_1||z - z'| < |z - z'| \sum_{n=2}^{+\infty} n|a_n|.$$

L'inégalité est bien justifiée puisque la série $\sum na_n$ est absolument convergente. Elle est bien stricte puisque $|z-z'|>0$. Comme $|z-z'|>0$, il vient $|a_1|<\sum_{n=2}^{+\infty} n|a_n|$, ce qui contredit l'hypothèse.

Conclusion. La fonction f est injective. ◁

L'exercice suivant étudie un peu le problème inverse : si une fonction développable en série entière sur le disque D *est injective, que peut-on dire de la taille de ses coefficients ?*

3.28. Théorème de Bieberbach dans le cas réel

Soit $f(z)=z+\sum_{n=2}^{+\infty} a_n z^n$ la somme d'une série entière de rayon de convergence $\geqslant 1$. On suppose que tous les a_n sont réels et que f est injective sur $\mathrm{D}=\{z\in\mathbb{C},\ |z|<1\}$.

1. Soit $z\in\mathrm{D}$. Montrer que $f(z)\in\mathbb{R}$ si et seulement si $z\in\mathbb{R}$. En déduire que si $\operatorname{Im} z\geqslant 0$ alors $\operatorname{Im} f(z)\geqslant 0$.

2. Calculer, pour $0<r<1$ et $n\in\mathbb{N}^*$, $\int_0^\pi \operatorname{Im} f(re^{i\theta})\sin n\theta\,\mathrm{d}\theta$. En déduire que $|a_n|\leqslant n$ pour tout $n\geqslant 1$.

3. Montrer que la majoration obtenue est optimale.

(École polytechnique)

▷ **Solution.**

1. Soit $z\in\mathrm{D}$. Bien entendu, si z est réel, $f(z)$ aussi car les coefficients a_n sont tous réels. Supposons inversement que $f(z)\in\mathbb{R}$. On a donc $\overline{f(z)}=f(z)$. Or, $\overline{f(z)}=f(\overline{z})$. L'injectivité de f permet de conclure que $z=\overline{z}$ c'est-à-dire que $z\in\mathbb{R}$.

Notons alors D^+ l'ensemble des éléments z de D tels que $\operatorname{Im} z>0$. La fonction $z\mapsto\operatorname{Im} f(z)$ est continue et ne s'annule pas sur D^+ d'après ce qu'on vient de voir. Comme D^+ est un ensemble connexe, cette fonction garde donc un signe constant. Il ne reste plus qu'à déterminer ce signe. Pour cela, on prend des valeurs proches de 0. Lorsque $t>0$ tend vers 0, $f(it)$ est équivalent à it et on a donc $\operatorname{Im} f(it)>0$ pour t assez petit. Par conséquent $\operatorname{Im} f(z)>0$ pour tout z de D^+. Cela montre bien que pour tout z de D tel que $\operatorname{Im} z\geqslant 0$ on a $\operatorname{Im} f(z)\geqslant 0$.

2. Soit r dans $]0,1[$ et $n\in\mathbb{N}^*$. On a $\operatorname{Im} f(re^{i\theta})=\sum_{k=1}^{+\infty} a_k r^k\sin k\theta$, où l'on pose $a_1=1$. Cette série converge uniformément pour $\theta\in[0,\pi]$. L'interversion de l'intégration et de la sommation qui suit est donc légitime :

$$\int_0^\pi \operatorname{Im} f(re^{i\theta}) \sin n\theta \mathrm{d}\theta = \sum_{k=1}^{+\infty} a_k r^k \int_0^\pi \sin k\theta \sin n\theta \mathrm{d}\theta = \frac{\pi}{2} a_n r^n$$

car toutes les intégrales sont nulles sauf celle qui correspond à $k = n$.

On utilise cette expression intégrale de a_n pour le majorer. Comme la partie imaginaire est toujours positive d'après la question précédente, on a

$$\frac{\pi}{2}|a_n|r^n \leqslant \int_0^\pi \operatorname{Im} f(re^{i\theta})|\sin n\theta|\mathrm{d}\theta.$$

On va maintenant majorer le terme $|\sin n\theta|$. Une majoration brutale par 1 n'est pas suffisante. De même la majoration $|\sin n\theta| \leqslant n\theta$ ne convient pas. En fait, on a l'inégalité $|\sin n\theta| \leqslant n|\sin\theta|$ pour tout n et tout réel θ. Elle se démontre très facilement par récurrence sur n : c'est clair au rang $n = 1$, et pour passer au rang $n + 1$ il suffit d'utiliser la formule d'addition $\sin(n + 1)\theta = \cos\theta \sin n\theta + \cos n\theta \sin\theta$. On obtient alors

$$\frac{\pi}{2}|a_n|r^n \leqslant n \int_0^\pi \operatorname{Im} f(re^{i\theta}) \sin\theta \mathrm{d}\theta = n\frac{\pi}{2} a_1 r.$$

Comme $a_1 = 1$, on a après simplification $|a_n|r^{n-1} \leqslant n$. Cela vaut pour tout réel r de $]0, 1[$. En faisant tendre r vers 1 on obtient le résultat.

3. Considérons la série entière $\sum_{n=1}^{+\infty} nz^n$. Son rayon de convergence est égal à 1 et pour $|z| < 1$ sa somme vaut $f(z) = \dfrac{z}{(1-z)^2}$. Pour voir que la majoration précédente est optimale, il suffit de vérifier que f est injective sur D. Soient z et z' deux points de D tels que $f(z) = f(z')$. On a $z(1 - z')^2 = z'(1 - z)^2$ ce qui équivaut à $(z' - z)(zz' - 1) = 0$. Comme $|zz'| < 1$ on a forcément $z = z'$ et f est donc bien injective. ◁

En fait Bieberbach avait conjecturé en 1916 que pour toute série entière complexe $\sum_{n=1}^{+\infty} a_n z^n$ *de rayon de convergence* $\geqslant 1$ *dont la somme est injective sur* D, *on a* $|a_n| \leqslant n|a_1|$ *pour tout* $n \geqslant 1$. *Ce résultat n'a été prouvé qu'en 1984 dans le cas général. L'étude du cas réel de l'exercice est due à Dieudonné (1931).*

Les deux exercices suivants sont des applications du principe des zéros isolés dont nous rappelons l'énoncé. Soit f *la somme de la série entière* $\sum a_n z^n$, *définie sur son disque de convergence. S'il existe une suite* (z_p) *de nombres complexes non nuls tendant vers* 0 *tels que* $f(z_p) = 0$, *pour tout* p, *alors* $a_n = 0$, *pour tout entier* n.

3.29. Série de Laurent à valeurs entières

Soit $F(z) = b_m z^m + \cdots + b_1 z + b_0 + \sum_{i \geqslant 1} b_{-i} z^{-i}$, la suite $(b_i)_{-\infty < i \leqslant m}$ étant une suite de réels ; on suppose qu'il existe $A \in \mathbb{R}$ tel que $F(z)$ existe pour $|z| \geqslant A$.

1. On suppose que $b_1, b_2, \ldots, b_m$ appartiennent à $\mathbb{Q}$ et qu'il existe $N \in \mathbb{N}$ tel que pour p entier, $p \geqslant N$ implique $F(p) \in \mathbb{Z}$. Montrer que $b_0 \in \mathbb{Q}$ et que, pour tout $i \geqslant 1$, $b_{-i} = 0$.

2. Même question en ne supposant plus $b_1, \ldots, b_m$ rationnels.

(**École normale supérieure**)

▷ **Solution.**

1. Posons, pour $z \in \mathbb{C}$, $P(z) = \sum_{i=1}^{m} b_i z^i$ et $G(z) = \sum_{i \geqslant 1} b_{-i} z^i$, quand cette série converge. Par hypothèse, $G\left(\frac{1}{z}\right)$ existe pour $|z| \geqslant A$. On en déduit que le rayon R de la série entière définissant $G(z)$ est supérieur ou égal à $\frac{1}{A}$ (et $R = +\infty$ si $A = 0$). On a, pour tout $|z| \geqslant A$, l'égalité

$$F(z) = P(z) + b_0 + G\left(\frac{1}{z}\right).$$

On suppose que $b_1, \ldots, b_m$ sont dans $\mathbb{Q}$. On considère un dénominateur commun q à $b_1, \ldots, b_m$. Pour $p \in \mathbb{N}$, $p \geqslant N$, $qP(p)$ et $qF(p) = qP(p) + b_0 q + qG\left(\frac{1}{p}\right)$ appartiennent à $\mathbb{Z}$. On en déduit que $b_0 q + qG\left(\frac{1}{p}\right) \in \mathbb{Z}$. Mais nous savons que $\lim_{p \to +\infty} b_0 q + qG\left(\frac{1}{p}\right) = b_0 q + qG(0) = b_0 q$ (car G, somme d'une série entière est continue sur $\left]-\frac{1}{A}, \frac{1}{A}\right[$). Une suite convergente d'entiers est stationnaire. Il existe donc un entier N' tel que, pour $p \geqslant N', p \in \mathbb{N}$, on ait $G\left(\frac{1}{p}\right) = 0$. Ceci montre que $b_0 q \in \mathbb{Z}$ et donc que $b_0 \in \mathbb{Q}$. On observe par ailleurs que la suite $\left(\frac{1}{p}\right)_{p \geqslant N'}$ converge vers 0 et que, pour $p \geqslant N'$, $G\left(\frac{1}{p}\right) = 0$. D'après le principe des zéros isolés, la série entière qui définit G a tous ses coefficients nuls : pour tout $i \geqslant 1$, $b_{-i} = 0$. Ceci achève la démonstration.

On a démontré qu'en fait, F est un polynôme. Il prend des valeurs entières sur $\mathbb{Z} \cap [p, +\infty[$. Nous avons démontré dans notre tome

1 d'algèbre (exercice 5.24) qu'un tel polynôme est combinaison linéaire à coefficients entiers des polynômes de Hilbert.

2. Essayons d'aboutir à la même conclusion sans supposer que les b_i sont dans $\mathbb{Q}$. Pour cela, nous montrerons que l'hypothèse $F(p) \in \mathbb{Z}$ pour $p \geqslant N$ implique que, pour tout $i \in [\![1, m]\!]$ $b_i \in \mathbb{Q}$, ce qui permettra de conclure grâce à la question 1.

Posons, pour simplifier, $\alpha_p = G\left(\frac{1}{p}\right)$, pour $p \in \mathbb{N}, p \geqslant A$. Considérons le polynôme $Q = P + b_0$. Nous allons exprimer ses coefficients en fonction des α_p. Le polynôme Q est de degré inférieur ou égal à m et est donc déterminé, pour $p \geqslant N$ par ses valeurs en $p, p+1, \ldots, p+m$. Il s'exprime en fonction des polynômes $L_p, L_{p+1}, \ldots, L_{p+m}$, polynômes interpolateurs de base relatifs au $(m+1)$-uplet $(p, p+1, \ldots, p+m)$. Pour tout $z \in \mathbb{C}$, on obtient

$$Q(z) = \sum_{k=p}^{p+m} Q(k) L_k(z), \quad \text{où } L_k(X) = \frac{\prod\limits_{\substack{p \leqslant j \leqslant p+m \\ j \neq k}} (X - j)}{\prod\limits_{\substack{p \leqslant j \leqslant p+m \\ j \neq k}} (k - j)}.$$

Pour commencer, intéressons-nous à b_m, le coefficient de X^m dans cette expression. On trouve

$$\begin{aligned} b_m &= \sum_{k=p}^{p+m} \frac{Q(k)}{\prod\limits_{\substack{p \leqslant j \leqslant p+m \\ j \neq k}} (k - j)} \\ &= \sum_{k=p}^{p+m} \frac{Q(k)}{(k-p)!(-1)^{p+m-k}(p+m-k)!} \\ &= \sum_{j=0}^{m} \frac{Q(p+j)(-1)^{m-j}}{j!(m-j)!}. \end{aligned}$$

On en déduit que

$$\begin{aligned} m!\, b_m &= \sum_{j=0}^{m} Q(p+j)(-1)^{m-j} C_m^j \\ &= \sum_{j=0}^{m} F(p+j)(-1)^{m-j} C_m^j - \sum_{j=0}^{m} \alpha_{p+j}(-1)^{m-j} C_m^j. \end{aligned}$$

Par hypothèse, $u_p = \sum\limits_{j=0}^{m} F(p+j)(-1)^{m-j} C_m^j$ est dans $\mathbb{Z}$ pour $p \geqslant N$; la suite $(\alpha_p)_{p \geqslant N}$ converge vers 0, donc $(u_p)_{p \geqslant N}$ converge vers $m!\, b_m$. C'est une suite d'entiers ; elle est donc stationnaire. On obtient pour $p \geqslant N'$, $m!\, b_m = u_p \in \mathbb{Z}$. Nous avons démontré que $b_m \in \mathbb{Q}$, et plus précisément que $m!\, b_m \in \mathbb{Z}$.

Nous sommes maintenant en mesure de démontrer que les b_i (pour $1 \leqslant i \leqslant m$) sont dans $\mathbb{Q}$, par récurrence sur $m \in \mathbb{N}^*$.

Si $m = 1$, la démonstration qui précède montre que $b_1 \in \mathbb{Q}$; c'est terminé. Supposons la propriété vérifiée au rang $m-1$, $m \geqslant 2$. Ce qui précède montre que $m!\,b_m$ est dans $\mathbb{Q}$. Posons, pour $x \geqslant \mathrm{A}$,

$$\mathrm{F}'(x) = m!\,\mathrm{F}(x) - m!\,b_m x^m = \sum_{i=0}^{m-1} b_i m!\,x^i + \sum_{i \geqslant 1} b_i m!\,x^i.$$

On a, par hypothèse, pour $p \in \mathbb{N}, p \geqslant \mathrm{N}$, $\mathrm{F}(p) \in \mathbb{Z}$, donc $\mathrm{F}'(p) \in \mathbb{Z}$. La fonction F' vérifie l'hypothèse de récurrence. On en déduit que, pour $i \in [\![1, m-1]\!]$, $b_i m! \in \mathbb{Q}$ et donc $b_i \in \mathbb{Q}$. La récurrence est terminée.

On a donc, pour tout $i \in [\![1, m]\!]$, $b_i \in \mathbb{Q}$ et on peut conclure comme dans la question 1. $\lhd$

Les trois exercices suivants concernent les fonctions analytiques réelles. Rappelons qu'une fonction f de classe $\mathcal{C}^\infty$ sur un intervalle I est dite analytique si pour tout $x_0 \in \mathrm{I}$ il existe un voisinage V de x_0 sur lequel f est égale à la somme de sa série de Taylor en x_0.

3.30. Caractérisation des fonctions réelles analytiques

Soit I un intervalle ouvert et $f : \mathrm{I} \to \mathbb{R}$ une fonction de classe $\mathcal{C}^\infty$. Montrer qu'il y a équivalence entre :

(*i*) f est analytique ;

(*ii*) pour tout segment $\mathrm{J} \subset \mathrm{I}$, il existe des constantes C, r strictement positives telles que

$$\forall n \in \mathbb{N},\ \forall x \in \mathrm{J},\ |f^n(x)| \leqslant \mathrm{C} r^n n!.$$

(École polytechnique)

▷ **Solution.**

• Montrons tout d'abord que $(ii) \Longrightarrow (i)$. Soit $x_0 \in \mathrm{I}$ et $\eta > 0$ tel que $\mathrm{J} = [x_0 - \eta, x_0 + \eta] \subset \mathrm{I}$. On fixe des constantes C et r telles que

$$\forall n \in \mathbb{N},\ \forall x \in \mathrm{J},\ |f^n(x)| \leqslant \mathrm{C} r^n n!.$$

On utilise la formule de Taylor avec reste intégral pour évaluer l'écart entre $f(x)$ et la somme partielle de la série de Taylor en x_0, pour x proche de x_0. On a, pour $x \in \mathrm{J}$ et $n \in \mathbb{N}$,

$$f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + \mathrm{R}_n(x),$$

avec

$$R_n(x) = \int_{x_0}^{x} \frac{(x-t)^n}{n!} f^{(n+1)}(t)\,dt.$$

L'hypothèse permet de majorer le reste de la manière suivante : pour $x \geqslant x_0$,

$$|R_n(x)| \leqslant \int_{x_0}^{x} \frac{(x-t)^n}{n!} C r^{n+1}(n+1)!dt = C r^{n+1}|x-x_0|^{n+1}.$$

On obtient la même majoration pour $x \leqslant x_0$.

Ainsi, lorsque $|x - x_0| < \min\left(\frac{1}{r}, \eta\right)$, $R_n(x)$ tend vers 0. Par conséquent, f est somme de sa série de Taylor en x_0 autour de x_0. Cela vaut pour tout point x_0 de I et f est donc analytique.

• Montrons maintenant que $(i) \Longrightarrow (ii)$. Cherchons d'abord à majorer les dérivées successives de f au voisinage d'un point x_0 quelconque de I. Par hypothèse il existe $\eta > 0$ tel que $f(x) = \sum_{n=0}^{+\infty} \frac{f^{(n)}(x_0)}{n!}(x-x_0)^n$ lorsque $|x - x_0| \leqslant \eta$. En particulier, la suite $\frac{f^{(n)}(x_0)}{n!}\eta^n$ tend vers 0 et on peut donc trouver $C > 0$ tel que $|f^{(n)}(x_0)| \leqslant Cn!r^n$ pour tout n, où $r = \frac{1}{\eta}$. On va en déduire une inégalité du même type valable sur tout un voisinage de x_0 en augmentant les constantes C et r. Les dérivées de la fonction f sur $]x_0 - \eta, x_0 + \eta[$ s'obtiennent en dérivant terme terme le développement en série de f. On a donc, pour $x \in]x_0 - \eta, x_0 + \eta[$ et $n \in \mathbb{N}$,

$$f^{(n)}(x) = \sum_{k=n}^{+\infty} \frac{f^{(k)}(x_0)}{k!} k(k-1)\dots(k-n+1)(x-x_0)^{k-n}.$$

On en déduit que

$$\begin{aligned}\left|f^{(n)}(x)\right| &\leqslant C \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1) r^k |x-x_0|^{k-n} \\ &= C r^n \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)(r|x-x_0|)^{k-n}.\end{aligned}$$

Prenons alors $|x - x_0| \leqslant \frac{1}{2r} = \frac{\eta}{2}$. La série ci-dessus est majorée par $\sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)\frac{1}{2^{k-n}}$. Or, cette somme se calcule facilement. Le développement en série entière de $\frac{1}{1-x}$ sur $]-1,1[$ est $\sum_{k=0}^{+\infty} x^k$. Si on le dérive n fois on obtient,

$$\frac{n!}{(1-x)^{n+1}} = \sum_{k=n}^{+\infty} k(k-1)\dots(k-n+1)x^{k-n},$$

et il suffit d'évaluer cela en $\frac{1}{2}$. On obtient donc pour tout x tel que $|x-x_0| \leqslant \frac{\eta}{2}$ et tout entier n,

$$\left|f^{(n)}(x)\right| \leqslant \mathrm{C}r^n n! 2^{n+1} = 2\mathrm{C}(2r)^n n!.$$

On peut alors prouver la proposition (ii). Soit J un segment inclus dans I. Ce qui précède montre que pour chaque point y de J on peut trouver $\eta_y > 0$, $\mathrm{C}_y > 0$ et $r_y > 0$ tels que pour tout n et tout x dans $]y-\eta_y, y+\eta_y[$ on ait $\left|f^{(n)}(x)\right| \leqslant \mathrm{C}_y(r_y)^n n!$. Les intervalles $]y-\eta_y, y+\eta_y[$ pour y variant dans J forment un recouvrement ouvert de J. Comme J est compact, on peut en extraire un sous-recouvrement fini par la propriété de Borel-Lebesgue. Si on note $y_1, \dots, y_s$ les centres des intervalles retenus, il suffit de poser $\mathrm{C} = \max_{1\leqslant i\leqslant s} \mathrm{C}_{y_i}$ et $r = \max_{1\leqslant i\leqslant s} r_{y_i}$. On a alors, pour tout x de J et tout entier n, $\left|f^{(n)}(x)\right| \leqslant \mathrm{C}r^n n!$, ce qui prouve (ii). ◁

L'exercice suivant étudie à quelle condition une fonction définie comme somme d'une séries de fonctions est analytique. Cela passe par une estimation, assez délicate, des dérivées successives de la fonction.

3.31. Étude d'analycité

Pour $a > 0$ et $x \in \mathbb{R}$ on pose $f(x) = \sum_{n=1}^{+\infty} \sin(nx)\exp(-n^a)$.

1. Montrer que f est bien définie et de classe $\mathcal{C}^\infty$ sur $\mathbb{R}$.

2. Pour quelles valeurs de a la fonction f est-elle développable en série entière au voisinage de tout point ?

(École normale supérieure)

▷ **Solution.**

1. On pose $u_n(x) = \sin(nx)e^{-n^a}$ pour tout $n \geqslant 1$ et tout $x \in \mathbb{R}$. La fonction u_n est de classe $\mathcal{C}^\infty$ et pour tout $k \in \mathbb{N}$,

$$u_n^{(k)}(x) = n^k \sin\left(nx + \frac{k\pi}{2}\right)\exp(-n^a).$$

On en déduit que, pour tout $x \in \mathbb{R}$, $|u_n^{(k)}(x)| \leqslant n^k \exp(-n^a)$. Pour tout $k \in \mathbb{N}$, on a $n^k \exp(-n^a) = o\left(\frac{1}{n^2}\right)$ donc la série de fonctions $\sum u_n^{(k)}$ converge normalement sur $\mathbb{R}$. On en déduit que f est définie sur $\mathbb{R}$ et,

en appliquant de manière réitérée le théorème de dérivation des séries de fonctions, que f est de classe $\mathcal{C}^\infty$ avec pour tout $k \in \mathbb{N}$, $f^{(k)} = \sum_{n=1}^{+\infty} u_n^{(k)}$.

2. Soit x_0 un réel quelconque. D'après la première question on a, pour tout $k \in \mathbb{N}$,

$$|f^{(k)}(x_0)| \leqslant \sum_{n=1}^{+\infty} n^k \exp(-n^a).$$

Le majorant est d'autant plus petit que a est grand. Pour avoir une idée de son ordre de grandeur, commençons par regarder le cas particulier $a = 1$ à l'aide d'une comparaison série-intégrale. On a

$$\begin{aligned}\sum_{n=1}^{+\infty} n^k \exp(-n) &\leqslant \sum_{n=1}^{+\infty} \int_n^{n+1} t^k e^{-(t-1)}\,\mathrm{d}t \leqslant e^{-1} \int_1^{+\infty} t^k e^{-t}\,\mathrm{d}t \\ &\leqslant e^{-1} \int_0^{+\infty} t^k e^{-t}\,\mathrm{d}t \leqslant e^{-1} k!.\end{aligned}$$

On a donc $\left|\dfrac{f^{(k)}(x_0)}{k!}\right| \leqslant e^{-1}$ et on en déduit que la série de Taylor en x_0 $\sum \dfrac{f^{(k)}(x_0)}{k!}(x-x_0)^k$ converge pour $|x-x_0| < 1$. Montrons que sa somme est $f(x)$. Pour $|x - x_0| < 1$ et $\mathrm{N} \in \mathbb{N}$, la formule de Taylor-Lagrange donne l'existence d'un réel c entre x_0 et x tel que

$$f(x) = \sum_{k=0}^{\mathrm{N}} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k + \frac{f^{(\mathrm{N}+1)}(c)}{(\mathrm{N}+1)!}(x-x_0)^{\mathrm{N}+1}.$$

D'après ce qui précède, on a $\left|\dfrac{f^{(\mathrm{N}+1)}(c)}{(\mathrm{N}+1)!}(x-x_0)^{\mathrm{N}+1}\right| \leqslant e^{-1}|x-x_0|^{\mathrm{N}+1}$. On en déduit que $\lim\limits_{\mathrm{N}\to\infty} \dfrac{f^{(\mathrm{N}+1)}(c)}{(\mathrm{N}+1)!}(x-x_0)^{\mathrm{N}+1} = 0$ et donc que

$$f(x) = \sum_{k=0}^{+\infty} \frac{f^{(k)}(x_0)}{k!}(x-x_0)^k.$$

Comme x_0 était quelconque on en déduit que f est analytique sur $\mathbb{R}$. Tout cela reste *a fortiori* valable lorsque $a \geqslant 1$.

Supposons maintenant que $a < 1$. On va voir dans ce cas que la série de Taylor de f en 0 a un rayon de convergence nul. On a pour tout $k \in \mathbb{N}$,

$$f^{(4k+1)}(0) = \sum_{n=1}^{+\infty} n^{4k+1} e^{-n^a}.$$

De nouveau, on compare à une intégrale, mais cette fois en minorant. On obtient

$$f^{(4k+1)}(0) \geqslant \sum_{n=1}^{+\infty} \int_{n-1}^{n} t^{4k+1} e^{-(t+1)^a}\, dt \geqslant \int_0^{\infty} t^{4k+1} e^{-(t+1)^a}\, dt.$$

Essayons à nouveau de faire apparaître la fonction Γ d'Euler. Pour $t \geqslant 0$, il existe $c \in]t, t+1[$ tel que $(t+1)^a - t^a = ac^{a-1}$. On en déduit que $(t+1)^a - t^a \leqslant a$ si $t \geqslant 1$, car $a - 1 < 0$. On a donc

$$f^{(4k+1)}(0) \geqslant \int_1^{+\infty} t^{4k+1} e^{-t^a - a}\, dt \geqslant e^{-a} \int_1^{+\infty} u^{\frac{4k+2}{a}-1} e^{-u} du$$
$$\geqslant e^{-a} \int_1^{+\infty} u^{\frac{4k}{a}} e^{-u} du,$$

avec le changement de variable $u = t^a$. Comme l'intégrande est inférieure à 1 sur $[0,1]$, on a

$$f^{(4k+1)}(0) \geqslant e^{-a} \int_0^{\infty} u^{\frac{4k}{a}} e^{-u} du - e^{-a} \geqslant e^{-a} \Gamma\left(\frac{4k}{a} + 1\right) - e^{-a}.$$

On a

$$\frac{f^{(4k+1)}(0)}{(4k+1)!} \geqslant e^{-a} \frac{\Gamma\left(\frac{4k}{a}+1\right) - 1}{(4k+1)!} \sim e^{-a} \frac{\Gamma\left(\frac{4k}{a}+1\right)}{(4k+1)!} = u_k.$$

Déterminons le rayon de convergence de la série entière $\sum u_k x^k$. On utilise la formule de Stirling

$$\Gamma(x+1) \underset{x \to +\infty}{\sim} \sqrt{2\pi x}\, x^x e^{-x}.$$

On en déduit, quand k tend vers $+\infty$,

$$\frac{u_{k+1}}{u_k} \sim \frac{(4k+1)! \sqrt{2\pi \frac{4k+4}{a}} \left(\frac{4k+4}{a}\right)^{\frac{4k+4}{a}} e^{-\frac{4k+4}{a}}}{(4k+5)! \sqrt{2\pi \frac{4k}{a}} \left(\frac{4k}{a}\right)^{\frac{4k}{a}} e^{-\frac{4k}{a}}} \sim \frac{(4k+4)^{\frac{4k+4}{a}} (ae)^{-\frac{4}{a}}}{(4k)^4 (4k)^{\frac{4k}{a}}}$$
$$\sim (4k)^{\frac{4}{a}-4} \left(\frac{k+1}{k}\right)^{\frac{4k}{a}} (ae)^{-\frac{4}{a}} \sim a^{-\frac{4}{a}} (4k)^{\frac{4}{a}-4}.$$

Comme $\frac{4}{a} - 4 > 0$, on a donc $\lim\limits_{k \to +\infty} \frac{u_{k+1}}{u_k} = +\infty$. Le rayon de convergence de la série entière $\sum u_k x^k$ est donc nul. Il en est de même *a fortiori* de la série entière $\sum \frac{f^{(4k+1)}(0)}{(4k+1)!} x^k$ et donc de $\sum \frac{f^{(4k+1)}(0)}{(4k+1)!} x^{4k+1}$. Pour tout $x \in \mathbb{R}$, la suite $\left(\frac{f^{(4k+1)}(0)}{(4k+1)!} x^{4k+1}\right)$ n'est pas bornée et *a fortiori* la suite $\left(\frac{f^{(k)}(0)}{k!} x^k\right)$ n'est pas bornée. Le rayon de convergence de

$\sum \frac{f^{(k)}(0)}{k!} x^k$ est nul. La série de Taylor de f en 0 ne converge sur aucun voisinage de 0. Donc f n'est pas développable en série entière au voisinage de 0.

Conclusion. La fonction f est développable en série entière au voisinage de tout point si et seulement si $a \geqslant 1$. ◁

Le résultat de l'exercice suivant est très classique. Il donne une condition suffisante, due à Bernstein, pour qu'une fonction de classe $\mathcal{C}^\infty$ soit analytique.

3.32. Un théorème de Bernstein

Soit $a < b$ deux réels et $f :]a, b[\to \mathbb{R}$ une fonction de classe $\mathcal{C}^\infty$. On dit que f est *absolument monotone* sur $]a, b[$ si f et toutes ses dérivées sont positives sur $]a, b[$.

1. Donner des exemples de fonctions absolument monotones.

2. Montrer qu'une fonction absolument monotone est analytique.

(École normale supérieure, école polytechnique)

▷ **Solution.**

1. La fonction exponentielle est positive et comme $\exp' = \exp$, toutes ses dérivées aussi. Elle est donc absolument monotone sur n'importe quel intervalle de $\mathbb{R}$.

La fonction tangente est absolument monotone sur $I = [0, \frac{\pi}{2}[$. En effet, on a $\tan' = 1 + \tan^2 > 0$ et on montre alors facilement par récurrence que toutes les dérivées de tan sont positives sur I. Il suffit pour cela de dériver n fois cette égalité en utilisant la formule de Leibniz :

$$\tan^{(n+1)} = \sum_{k=0}^{n} \mathrm{C}_n^k \tan^{(k)} \tan^{(n-k)} .$$

2. Soit $f :]a, b[\to \mathbb{R}$ une fonction absolument monotone et $x_0 \in]a, b[$. On va prouver que sur un voisinage de x_0, f est égale à la somme de sa série de Taylor en x_0. Quitte à effectuer une translation de la variable (c'est-à-dire remplacer f par $f(x + x_0)$) on peut supposer $x_0 = 0$ et donc $a < 0$ et $b > 0$. On pose alors, pour $x \in]a, b[$

$$\mathrm{R}_n(x) = f(x) - \sum_{k=0}^{n} \frac{f^{(k)}(0)}{k!} x^k$$

le reste d'ordre n. La formule de Taylor avec reste intégral donne l'expression suivante de ce reste :

$$\mathrm{R}_n(x) = \int_0^x \frac{(x-t)^n}{n!} f^{(n+1)}(t)\mathrm{d}t = \frac{x^{n+1}}{n!} \int_0^1 (1-u)^n f^{(n+1)}(xu)\mathrm{d}u.$$

Notons que ce reste est positif pour $x \geqslant 0$. Prenons $\eta > 0$ tel que $[-\eta, \eta] \subset]a, b[$ et $x \in [-\eta, \eta]$. Comme $f^{(n+1)}$ est croissante (car $f^{(n+2)}$ est positive), on peut majorer $f^{(n+1)}(xu)$ par $f^{(n+1)}(\eta u)$ pour tout $u \in [0, 1]$ et comme ces termes sont positifs on a

$$\int_0^1 (1-u)^n f^{(n+1)}(xu)\mathrm{d}u \leqslant \int_0^1 (1-u)^n f^{(n+1)}(\eta u)\mathrm{d}u.$$

Il en résulte donc que,

$$|\mathrm{R}_n(x)| \leqslant \frac{|x|^{n+1}}{\eta^{n+1}} \mathrm{R}_n(\eta).$$

Il suffit maintenant de majorer $\mathrm{R}_n(\eta)$. Comme pour tout n, $\mathrm{R}_n(\eta) \geqslant 0$, les sommes partielles de la série à termes positifs $\sum \frac{f^{(k)}(0)}{k!} \eta^k$ sont toutes majorées par $f(\eta)$. Ainsi, cette série converge et a une somme $\leqslant f(\eta)$. Par suite on a $\mathrm{R}_n(\eta) \leqslant f(\eta)$ pour tout n. L'inégalité

$$|\mathrm{R}_n(x)| \leqslant \frac{|x|^{n+1}}{\eta^{n+1}} f(\eta)$$

montre alors que $\mathrm{R}_n(x)$ tend vers 0 pour tout x dans $]-\eta, \eta[$. D'où le résultat. $\lhd$

L'exercice suivant démontre le principe des zéros isolés pour les fonctions analytiques réelles.

3.33. Principe des zéros isolés

Soit $a < b$ deux réels et $f :]a, b[\to \mathbb{R}$ une fonction analytique. On suppose que la fonction f admet une infinité de zéros dans un segment $[c, d]$ de $]a, b[$. Montrer que f est identiquement nulle.

(École polytechnique)

$\rhd$ **Solution.**

Par hypothèse on peut se donner une suite $(u_n)_{n\geqslant 0}$ formée de zéros deux à deux distincts de f et, quitte à la remplacer par une sous-suite, on peut très bien supposer que cette suite converge vers un point x_0 du segment $[c, d]$. Par continuité de f on a $f(x_0) = 0$. Montrons que nécessairement toutes les dérivées de f sont nulles en x_0. En effet, si ce n'est pas le cas, en notant p le plus petit entier tel que $f^{(p)}(x_0) \neq 0$, on

a $f(x) = \dfrac{f^{(p)}(x_0)}{p!}(x-x_0)^p + o(x-x_0)^p$ par un simple développement limité à l'ordre p. Ainsi, $\dfrac{f(x)}{(x-x_0)^p}$ tend vers une limite non nulle en x_0 et le théorème de localisation permet de dire qu'il existe un voisinage V de x_0 tel que $f(x) \neq 0$ pour $x \in \mathrm{V} \setminus \{x_0\}$. Cela est absurde puisque les zéros u_n sont tous dans V pour n assez grand.

Posons alors $\mathrm{E} = \{x \in]a,b[,\ \forall n \in \mathbb{N}, f^{(n)}(x) = 0\}$. On vient de voir que E n'est pas vide. Comme f et toutes ses dérivées sont continues, il est clair que E est un ensemble fermé (de $]a,b[$). Enfin, f étant analytique, si $x \in \mathrm{E}$, f est identiquement nulle sur un voisinage de x et par conséquent E est aussi ouvert. Comme l'intervalle $]a,b[$ est connexe, on a forcément $\mathrm{E} =]a,b[$ et f est donc identiquement nulle. ◁

Si $\sum a_n z^n$ est une série entière de rayon de convergence $\mathrm{R} > 0$ et de somme f, pour tout $r \in]0,1[$, la série $\sum a_n r^n e^{in\theta}$, dont la somme est $f(re^{i\theta})$, converge uniformément par rapport à θ. Ceci permet de justifier l'échange de l'intégration et de la sommation qui conduit à $a_n r^n = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(re^{i\theta})e^{-in\theta}\,d\theta$. Ce résultat très important sera supposé connu dans tous les exercices qui suivent. C'est un instrument puissant. Il permet de traduire une information sur f en une information sur les coefficients a_n. Plus précisément, de la connaissance de f sur le cercle de centre 0 et de rayon r, on en déduit les coefficients a_n et donc la connaissance de la fonction f sur tout le disque de convergence. On obtient en particulier les inégalités de Cauchy : si m_r est le maximum de $|f(z)|$, pour $|z| = r$, on a $|a_n| r^n \leqslant m_r$.

Notons de plus que, pour $r \in]0,\mathrm{R}[$, la fonction $f_r : \theta \longmapsto f(re^{i\theta})$ est continue et 2π-périodique. Ses coefficients de Fourier exponentiels sont d'après ce qui précède les $a_n r^n$ pour $n \geqslant 0$ et 0 sinon. La formule de Parseval appliquée à f_r donne donc $\displaystyle\sum_{n=0}^{+\infty} |a_n|^2 r^{2n} = \frac{1}{2\pi}\int_0^{2\pi} |f(r^{i\theta})|^2\,d\theta$.

Outre la formule donnant les coefficients a_n sous forme intégrale, l'exercice suivant utilise aussi le principe des zéros isolés.

3.34. Nullité sur un arc du cercle de convergence

On pose $\mathrm{D} = \{z \in \mathbb{C},\ |z| < 1\}$.

1. Soit $f : \overline{\mathrm{D}} \to \mathbb{C}$ continue sur $\overline{\mathrm{D}}$, développable en série entière sur D et nulle sur le cercle unité S^1. Montrer que f est nulle.

2. Même question en supposant seulement f nulle sur un arc du cercle unité de longueur $\alpha > 0$.

(École normale supérieure)

▷ **Solution.**

1. Posons $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ pour $|z| < 1$. On a pour tout $n \in \mathbb{N}$ et $0 < r < 1$,

$$a_n r^n = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} d\theta.$$

Comme $(r, \theta) \longmapsto f(re^{i\theta})e^{-in\theta}$ est continue sur $[0,1] \times \mathbb{R}$, l'application $r \longmapsto \int_0^{2\pi} f(re^{i\theta})e^{-in\theta} d\theta$ est continue sur $[0,1]$. On fait tendre r vers 1 dans l'égalité précédente et on obtient $a_n = 0$. Il en résulte que $f = 0$.

2. L'idée est de construire à partir de f une fonction g vérifiant l'hypothèse de la question 1 en faisant « tourner » l'arc de longueur α pour recouvrir tout le cercle S^1.

Posons $g(z) = \prod_{k=0}^{n} f(e^{ik\alpha} z)$, où n est choisi de sorte que $n\alpha > 2\pi$. La fonction g est continue sur $\overline{\mathrm{D}}$, développable en série entière sur D, car chaque application $f_k : z \longmapsto f(e^{ik\alpha} z)$ l'est, et s'annule sur S^1 (car si $u \in \mathrm{S}^1$ il existe k tel que $f_k(u) = 0$). D'après la question 1, la fonction g est nulle. Pour tout $p \in \mathbb{N}^*$, on a $g\left(\frac{1}{p}\right) = 0$ et donc il existe un entier $k \in [\![0, n-1]\!]$ tel que $f\left(\frac{1}{p} e^{ik\alpha}\right) = 0$ pour une infinité de valeurs de p. Le principe des zéros isolés permet d'affirmer que la fonction $z \longmapsto f(ze^{ik\alpha})$ est nulle sur D, c'est-à-dire que f est nulle sur D, puisque $ze^{ik\alpha}$ décrit D quand z décrit D. Par continuité, f est nulle sur $\overline{\mathrm{D}}$. ◁

3.35. Série entière à coefficients entiers, bornée sur son disque de convergence

Soit (a_n) une suite de $\mathbb{Z}$ telle que la série $\sum a_n z^n$ ait un rayon de convergence $\mathrm{R} \geqslant 1$. On pose $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ pour $|z| < 1$ et on suppose que f est bornée sur le disque ouvert $\mathrm{D} = \{z \in \mathbb{C},\ |z| < 1\}$. Montrer que f est un polynôme.

(École normale supérieure)

▷ **Solution.**

Soit $r \in]0,1[$. L'égalité de Parseval, appliquée à la fonction f_r définie par $f_r(\theta) = f(re^{i\theta})$, s'écrit

$$\sum_{p=0}^{+\infty} |a_p|^2 r^{2p} = \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})|^2 \mathrm{d}\theta.$$

Soit M un majorant de $|f|$ sur le disque ouvert unité. On a la majoration $\sum_{p=0}^{+\infty} |a_p|^2 r^{2p} \leqslant M^2$ et donc, pour tout $N \in \mathbb{N}$, $\sum_{p=0}^{N} |a_p|^2 r^{2p} \leqslant M^2$. Ceci étant vrai pour tout $r \in]0,1[$, on en déduit, en faisant tendre r vers 1, que $\sum_{p=0}^{N} |a_p|^2 \leqslant M^2$, pour tout $N \in \mathbb{N}$. Ainsi la série $\sum |a_n|^2$ converge et donc la suite (a_n) converge vers 0. Les a_n étant des entiers, ceci implique que tous les a_n sont nuls à partir d'un certain rang. La fonction f est donc un polynôme. ◁

Soit f une fonction entière (c'est-à-dire développable en série entière avec un rayon de convergence infini) et bornée sur $\mathbb{C}$. Des formules de Cauchy, on déduit que, pour tout $n \in \mathbb{N}$ et tout $r > 0$,

$$|a_n| \leqslant \frac{2\pi}{r^n} \left| \int_0^{2\pi} f(re^{i\theta})\, d\theta \right| \leqslant \frac{1}{r^n} \|f\|_\infty.$$

En faisant tendre r vers l'infini, on obtient que $a_n = 0$ pour $n > 0$ et donc que f est constante. On obtient le théorème de Liouville : une fonction entière et bornée sur $\mathbb{C}$ est constante.

L'exercice suivant montre que cela reste vrai si on suppose seulement la partie réelle de f bornée.

3.36. Fonctions entières de partie réelle bornée

Soit $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ la somme d'une série entière de rayon de convergence $+\infty$. On pose $m_f(r) = \sup_{|z|=r} |f(z)|$ et $a_f(r) = \sup_{|z|=r} |\operatorname{Re} f(z)|$ pour $r > 0$.

1. Soit $0 < r_1 < r_2$; montrer que

$$m_f(r_1) \leqslant \frac{2r_1}{r_2 - r_1} a_f(r_2) + \frac{r_2 + r_1}{r_2 - r_1} |f(0)|.$$

On pourra se ramener au cas $f(0) = 0$.

2. On suppose qu'il existe M tel que, pour tout $z \in \mathbb{C}$, $|\operatorname{Re}(f(z))| \leqslant M$. Montrer que f est constante.

(École normale supérieure)

▷ **Solution.**

1. On va commencer par chercher une majoration des coefficients a_p en fonction de la quantité $a_f(r)$. Pour cela, on va utiliser une expression intégrale de a_p faisant intervenir la fonction $\operatorname{Re} f$.

Pour $r > 0$ et $\theta \in \mathbb{R}$, on a $f(re^{i\theta}) = \sum_{n=0}^{+\infty} a_n r^n e^{in\theta}$ et donc

$$\operatorname{Re} f(re^{i\theta}) = \frac{1}{2} \sum_{n=0}^{+\infty} r^n (a_n e^{in\theta} + \overline{a}_n e^{-in\theta}).$$

On en déduit que, pour tout $p \in \mathbb{N}$,

$$e^{-ip\theta} \operatorname{Re} f(re^{i\theta}) = \frac{1}{2} \sum_{n=0}^{+\infty} r^n (a_n e^{i(n-p)\theta} + \overline{a}_n e^{-i(n+p)\theta}).$$

La série $\sum |a_n| r^n$ converge, donc la série qui définit $e^{-ip\theta} \operatorname{Re} f(re^{i\theta})$ converge uniformément par rapport à θ, ce qui justifie l'interversion de l'intégration et de la sommation dans ce qui suit. Pour $p \in \mathbb{N}^*$, on a

$$\begin{aligned} \int_0^{2\pi} e^{-ip\theta} \operatorname{Re} f(re^{i\theta}) \mathrm{d}\theta &= \frac{1}{2} \sum_{n=0}^{+\infty} r^n \int_0^{2\pi} (a_n e^{i(n-p)\theta} + \overline{a}_n e^{-i(n+p)\theta}) \mathrm{d}\theta \\ &= \pi r^p a_p, \end{aligned}$$

car $\int_0^{2\pi} e^{ik\theta} \mathrm{d}\theta = 0$, pour tout $k \in \mathbb{N}^*$; il y a donc un seul terme non nul dans cette somme, celui qui correspond à $n = p$.

On en déduit, pour tout $p \in \mathbb{N}^*$, la majoration suivante :

$$\pi r^p |a_p| \leqslant \int_0^{2\pi} |\operatorname{Re} f(re^{i\theta})| \mathrm{d}\theta \leqslant 2\pi a_f(r),$$

c'est-à-dire $|a_p| \leqslant \dfrac{2a_f(r)}{r^p}$.

Supposons maintenant que $f(0) = 0$, c'est-à-dire que $a_0 = 0$. Soit $0 < r_1 < r_2$. On applique les inégalités précédentes avec $r = r_2$. La série géométrique de terme général $\left(\dfrac{r_1}{r_2}\right)^n$ converge, donc on peut écrire, si $|z| = r_1$,

$$\begin{aligned} |f(z)| &\leqslant \sum_{n=1}^{+\infty} r_1^n |a_p| \leqslant 2a_f(r_2) \sum_{n=1}^{+\infty} \left(\frac{r_1}{r_2}\right)^n, \\ |f(z)| &\leqslant 2a_f(r_2) \frac{\dfrac{r_1}{r_2}}{1 - \dfrac{r_1}{r_2}} = a_f(r_2) \frac{2r_1}{r_2 - r_1}. \end{aligned}$$

Cela étant vrai pour tout z de module r_1, on conclut que

$$m_f(r_1) \leqslant a_f(r_2) \frac{2r_1}{r_2 - r1}.$$

Dans le cas général, où on ne suppose plus $f(0) = 0$, on considère la fonction g définie par $g(z) = f(z) - f(0)$. Elle vérifie l'inégalité précédente. De plus, on a, pour tout z de module r_1, $|f(z)| \leqslant |g(z)| +$

$|f(0)|$ et donc $m_f(r_1) \leqslant m_g(r_1) + |f(0)|$. On obtient de même, pour $|z| = r_2$, $\mathrm{Re}(g(z)) = \mathrm{Re}(f(z)) + \mathrm{Re}(f(0))$, d'où l'on tire $|a_g(r_2)| \leqslant |a_f(r_2)| + |\mathrm{Re}(f(0))| \leqslant |a_f(r_2)| + |f(0)|$. On obtient alors, à partir de ces différentes inégalités

$$m_f(r_1) \leqslant m_g(r_1) + |f(0)| \leqslant \frac{2r_1}{r_2 - r_1} a_g(r_2) + |f(0)|$$
$$m_f(r_1) \leqslant \frac{2r_1}{r_2 - r_1} a_f(r_2) + \frac{r_2 + r_1}{r_2 - r_1} |f(0)|,$$

ce qui est l'inégalité demandée.

2. On se ramène à une fonction nulle en 0, en considérant à nouveau la fonction g définie par $g(z) = f(z) - f(0)$. On a, pour tout $z \in \mathbb{C}$, $|\mathrm{Re}(g(z))| \leqslant \mathrm{M}'$, où $\mathrm{M}' = \mathrm{M} + |\mathrm{Re}(f(0))|$ et pour tout $r > 0$, $a_g(r) \leqslant \mathrm{M}'$. Si $0 < r_1 < r_2$, on a, d'après la question précédente,

$$m_g(r_1) \leqslant \frac{2r_1}{r_2 - r_1} \mathrm{M}'.$$

En faisant tendre r_2 vers $+\infty$, on obtient, pour tout $r_1 > 0$, $m_g(r_1) \leqslant 0$ et donc $m_g(r_1) = 0$. La fonction g est nulle sur tous les cercles de la forme $|z| = r_1$ et elle est nulle en 0. C'est la fonction nulle. La fonction f est donc constante. ◁

Si f est une fonction entière, $f(z) = \sum_{n=0}^{+\infty} a_n z^n$, un contrôle des coefficients a_n implique assez facilement un contrôle de f. Par exemple si $|a_n| \leqslant \dfrac{\mathrm{K}^n}{n!}$ pour tout n, on a tout de suite $|f(z)| \leqslant e^{\mathrm{K}|z|}$ pour tout z. L'exercice suivant aborde le problème inverse : comment déduire une estimation des coefficients a_n à partir d'une information sur $|f(z)|$?

3.37. Fonction entière dominée par une exponentielle

Soit $f : \mathbb{C} \longrightarrow \mathbb{C}$ développable en série entière sur $\mathbb{C}$, $f(z) = \sum_{n=0}^{+\infty} c_n z^n$. On suppose qu'il existe $\mathrm{K} > 0$ tel que pour tout $\mathrm{K}_1 > \mathrm{K}$, il existe $\mathrm{A}_{\mathrm{K}_1} \geqslant 0$, vérifiant, pour tout $z \in \mathbb{C}$,

$$|z| \geqslant \mathrm{A}_{\mathrm{K}_1} \Longrightarrow |f(z)| \leqslant e^{\mathrm{K}_1 |z|}.$$

Soit $\mathrm{K}_2 > \mathrm{K}$. Montrer qu'il existe $n_0 \in \mathbb{N}$ tel que, pour tout $n \in \mathbb{N}$,

$$n \geqslant n_0 \Longrightarrow |c_n| \leqslant \frac{\mathrm{K}_2^n}{n!}.$$

(École polytechnique)

▷ **Solution.**

On a, pour $r>0$ et $n\in\mathbb{N}$, $c_n r^n = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f(re^{i\theta})e^{-in\theta}\mathrm{d}\theta$.

Soit $K_2 > K$ et $K_1 \in]K, K_2[$. Par hypothèse, si $r \geqslant A_{K_1}$ on a pour tout θ, $|f(re^{i\theta})| \leqslant e^{K_1 r}$. On en déduit que

$$|c_n|r^n \leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(re^{i\theta})|\mathrm{d}\theta \leqslant \frac{1}{2\pi}\int_0^{2\pi} e^{K_1 r}\mathrm{d}\theta \leqslant e^{K_1 r}$$

et donc $|c_n| \leqslant \dfrac{1}{r^n}e^{K_1 r}$.

La fonction $r \longmapsto \dfrac{1}{r^n}e^{K_1 r}$ est minimale pour $r = \dfrac{n}{K_1}$ et ce minimum est égal à $\dfrac{e^n K_1^n}{n^n}$. On peut prendre la valeur $r = \dfrac{n}{K_1}$ dans l'inégalité précédente si $\dfrac{n}{K_1} \geqslant A_{K_1}$, soit $n \geqslant K_1 A_{K_1}$. Posons $n_1 = E(K_1 A_{K_1}) + 1$. Pour $n \geqslant n_1$, on a donc $|c_n| \leqslant \dfrac{e^n K_1^n}{n^n}$.

Il faut comparer cette dernière expression à $\dfrac{K_2^n}{n!}$. Étudions le quotient des deux $\dfrac{e^n K_1^n}{n^n} \Big/ \dfrac{K_2^n}{n!} = \dfrac{n! e^n K_1^n}{n^n K_2^n}$. D'après la formule de Stirling, on sait que $n! \underset{n\to\infty}{\sim} \left(\dfrac{n}{e}\right)^n \sqrt{2\pi n}$ et donc $\dfrac{n! e^n K_1^n}{n^n K_2^n} \underset{n\to\infty}{\sim} \left(\dfrac{K_1}{K_2}\right)^n \sqrt{2\pi n}$. Puisque $\dfrac{K_1}{K_2} < 1$, on en déduit que $\lim\limits_{n\to+\infty} \dfrac{n! e^n K_1^n}{n^n K_2^n} = 0$. Il existe un entier $n_0 \geqslant n_1$ tel que $\dfrac{n! e^n K_1^n}{n^n K_2^n} \leqslant 1$ pour $n \geqslant n_0$. On a alors, pour $n \geqslant n_0$, $|c_n| \leqslant \dfrac{K_2^n}{n!}$. ◁

3.38. Relation entre $\sqrt[n]{|a_n|}$ et $\dfrac{\ln|f(z)|}{|z|}$

Soit $(a_n)_{n\in\mathbb{N}}$ une suite de nombres complexes. On note R le rayon de convergence de la série entière $\sum \dfrac{a_n}{n!}z^n$ et $f(z)$ sa somme. Soit l un réel strictement positif. Montrer l'équivalence entre :

(i) pour tout $l' > l$, il existe $n_0 \in \mathbb{N}^*$ tel que $\sqrt[n]{|a_n|} \leqslant l'$ pour $n \geqslant n_0$;

(ii) $R = +\infty$ et pour tout $l' > l$, il existe $r_0 > 0$ tel que $\dfrac{\ln|f(z)|}{|z|} \leqslant l'$ pour $|z| \geqslant r_0$.

(École polytechnique)

▷ **Solution.**

• Supposons que (i) est vérifiée. La suite $(\sqrt[n]{|a_n|})$ est majorée par M.

On a, pour tout $z \in \mathbb{C}$ et tout $n \in \mathbb{N}^*$, $\left|\frac{a_n}{n!}z^n\right| \leqslant \frac{|\mathrm{M}z|^n}{n!}$ donc $\sum \frac{a_n}{n!}z^n$ converge absolument. Le rayon de convergence de la séric est donc $+\infty$.

Étant donné $l' > l$, choisissons $l'' \in]l, l'[$. Soit n_0 tel que $\sqrt[n]{|a_n|} \leqslant l''$ pour $n \geqslant n_0$. On a alors, pour tout $z \in \mathbb{C}$,

$$\begin{aligned}|f(z)| &\leqslant \sum_{n=0}^{+\infty} \frac{|a_n|}{n!}|z|^n \leqslant \sum_{n=0}^{n_0-1} \frac{|a_n|}{n!}|z|^n + \sum_{n=n_0}^{+\infty} \frac{|l''z|^n}{n!} \\ &= \sum_{n=0}^{n_0-1} \left(\frac{|a_n|}{n!} - \frac{l''^n}{n!}\right)|z|^n + e^{l''|z|} = \mathrm{P}(|z|) + e^{l''|z|},\end{aligned}$$

où P est un polynôme. On en déduit, que pour tout $z \in \mathbb{C}$, on a

$$\begin{aligned}\ln|f(z)| &\leqslant \ln\left(\mathrm{P}(|z|) + e^{l''|z|}\right) \\ &\leqslant l''|z| + \ln\left(1 + \mathrm{P}(|z|)e^{-(l'')|z|}\right).\end{aligned}$$

$$\frac{\ln|f(z)|}{|z|} \leqslant l'' + \frac{\ln\left(1 + \mathrm{P}(|z|)e^{-l''|z|}\right)}{|z|}.$$

Sachant que $\lim\limits_{x\to+\infty} l'' + \dfrac{\ln\left(1 + \mathrm{P}(x)e^{-l''x}\right)}{x} = l'' < l'$, il existe $r_0 > 0$ tel que l'on ait, pour $|z| \geqslant r_0$, $\dfrac{\ln|f(z)|}{|z|} \leqslant l'$.

• Supposons que (ii) est vérifiée. On exprime classiquement les coefficients a_n sous forme intégrale. On a, pour $r > 0$ et $n \in \mathbb{N}$,

$$|a_n| = \frac{n!r^{-n}}{2\pi}\left|\int_0^{2\pi} f(re^{i\theta})e^{-in\theta}\mathrm{d}\theta\right| \leqslant \frac{n!r^{-n}}{2\pi}\int_0^{2\pi}|f(re^{i\theta})|\mathrm{d}\theta.$$

Soit $l' > l$. Considérons $l'' \in]l, l'[$. Il existe $r_0 > 0$ tel que, si $|z| \geqslant r_0$, alors $\dfrac{\ln|f(z)|}{|z|} \leqslant l''$ et donc $|f(z) \leqslant e^{l''|z|}$. De ce qui précède, on déduit que, pour $r \geqslant r_0$, on a, pour tout $n \in \mathbb{N}$,

$$|a_n| \leqslant \frac{n!r^{-n}}{2\pi}\int_0^{2\pi} e^{l''r}\mathrm{d}\theta \leqslant n!r^{-n}e^{l''r}.$$

Le minimum de la fonction $r \longmapsto r^{-n}e^{l''r}$ est obtenu pour $r = \dfrac{n}{l''}$. Il vaut $\left(\dfrac{l''}{n}\right)^n e^n$. On a donc, pour $n \geqslant r_0 l''$, $|a_n| \leqslant n!\left(\dfrac{l''e}{n}\right)^n$ et donc $\sqrt[n]{|a_n|} \leqslant \sqrt[n]{n!}\left(\dfrac{l''e}{n}\right)$. Cherchons un équivalent du membre de droite. La formule de Stirling nous donne $n! \underset{n\to\infty}{\sim} \left(\dfrac{n}{e}\right)^n\sqrt{2\pi n}$. On en déduit que $\sqrt[n]{n!} \underset{n\to\infty}{\sim} \dfrac{n}{e}(2\pi n)^{\frac{1}{2n}} \underset{n\to\infty}{\sim} \dfrac{n}{e}$, puis que

$$\lim_{n\to+\infty} \sqrt[n]{n!}\left(\frac{l''e}{n}\right) = l'' < l'.$$

En particulier, il existe un entier n_0 tel que $\sqrt[n]{n!}\,\dfrac{l''e}{n} \leqslant l'$ pour $n \geqslant n_0$. On en déduit que $\sqrt[n]{|a_n|} \leqslant l'$ pour $n \geqslant n_0$. ◁

Dans l'exercice 3.12 on a vu qu'une fraction rationnelle complexe dont 0 n'est pas pôle est développable en série entière en 0 avec un rayon de convergence égal au plus petit module des pôles de f. Il est aisé d'en déduire la première question de l'énoncé suivant, à savoir qu'une fraction rationnelle f est analytique sur son ouvert de définition, c'est-à-dire développable en série entière au voisinage de tout point qui n'est pas un pôle. La suite de l'exercice démontre le principe du maximum pour une telle fonction.

3.39. Principe du maximum pour une fraction rationnelle

Soit f une fonction rationnelle qui n'a pas de pôle dans le disque fermé $\overline{\mathrm{D}}(c,\mathrm{R})$ de centre c et de rayon R et $z_0 \in \mathrm{D}(c,\mathrm{R})$.

1. Montrer que f est développable en série entière au voisinage de z_0 ; quel est son rayon de convergence ?

2. Soit $(a_n)_{n\in\mathbb{N}}$ les coefficients du développement en série entière ; calculer en fonction des a_n, pour r assez petit,

$$\frac{1}{2\pi}\int_0^{2\pi} |f(z_0 + re^{i\theta})|^2 \mathrm{d}\theta.$$

3. Prouver que $|f|$ atteint son maximum sur $\overline{\mathrm{D}}(c,\mathrm{R})$ en un point de la frontière.

(**École polytechnique**)

▷ **Solution.**

1. Comme $z_0 \in \mathrm{D}(c,\mathrm{R})$, il n'est pas pôle de f. Considérons la fonction g définie par $g(z) = f(z+z_0)$. C'est encore une fraction rationnelle et 0 n'en est pas pôle. On sait alors que g est développable en série entière en 0 avec un rayon de convergence égal au plus petit module de ses pôles. Cela revient exactement à dire que f est développable en série entière en z_0 avec un rayon de convergence R' égal à $\min|z_0 - \alpha|$ où α parcourt les pôles de f.

2. On écrit, pour $|z-z_0|<\mathrm{R}'$, $f(z)=\sum_{n=0}^{+\infty} a_n(z-z_0)^n$. Si $r\in\,]0,\mathrm{R}'[$, on obtient $f(z_0+re^{i\theta})=\sum_{n=0}^{+\infty} a_n r^n e^{in\theta}$. La fonction $\theta\longmapsto f(z_0+re^{i\theta})$ est continue et 2π-périodique. Ses coefficients de Fourier sont obtenus, pour $n\in\mathbb{Z}$ par

$$c_n=\frac{1}{2\pi}\int_0^{2\pi} f(z_0+re^{i\theta})e^{-in\theta}\mathrm{d}\theta=\frac{1}{2\pi}\int_0^{2\pi}\left(\sum_{k=0}^{+\infty} a_k r^k e^{i(k-n)\theta}\right)\mathrm{d}\theta$$

$$=\frac{1}{2\pi}\sum_{k=0}^{+\infty}\int_0^{2\pi} a_k r^k e^{i(k-n)\theta}\mathrm{d}\theta=a_n r^n,$$

l'interversion de l'intégration et de la sommation étant justifiée par la convergence normale de la série. La formule de Parseval donne alors

$$\frac{1}{2\pi}\int_0^{2\pi}|f(z_0+re^{i\theta})|^2\mathrm{d}\theta=\sum_{n=0}^{+\infty}|a_n|^2r^{2n}.$$

3. Le disque fermé $\overline{\mathrm{D}}(c,\mathrm{R})$ étant compact, la fonction continue $|f|$ y atteint son maximum. Supposons que ce maximum est atteint en un point du disque ouvert $\mathrm{D}(c,\mathrm{R})$. On prend pour z_0 un tel point et on applique ce qui précède. Pour $r>0$ assez petit, on a donc,

$$\frac{1}{2\pi}\int_0^{2\pi}|f(z_0+re^{i\theta})|^2\mathrm{d}\theta\leqslant\frac{1}{2\pi}\int_0^{2\pi}|f(z_0)|^2\mathrm{d}\theta=|f(z_0)|^2=|a_0|^2.$$

On obtient donc

$$\sum_{n=0}^{+\infty} r^{2n}|a_n|^2\leqslant|a_0|^2,$$

ce qui implique $a_n=0$ pour $n\geqslant 1$. On a donc $f(z)=a_0$, pour tout z tel que $|z-z_0|<\mathrm{R}'$. La fraction rationnelle f, qui prend une infinité de fois la valeur a_0, est donc constante. Mais alors $|f|$ atteint son maximum en tout point et donc en particulier sur la frontière de $\overline{\mathrm{D}}(c,\mathrm{R})$. ◁

Le lecteur notera que le même raisonnement s'appliquerait à toute fonction f continue sur $\overline{\mathrm{D}}$ et analytique sur D. Le maximum de $|f|$ sur le bord de $\overline{\mathrm{D}}$ est égal au maximum de $|f|$ sur $\overline{\mathrm{D}}$. Ce résultat est une des formulations possibles du principe du maximum.

L'exercice suivant, qui étudie diverses normes sur l'espace des fonctions développables en série entière avec un rayon de convergence R, *démontre le principe du maximum dans le cas général par une méthode différente.*

3.40. Principe du maximum

Soit $f(z) = \sum_{n=0}^{+\infty} \alpha_n z^n$ une série entière de rayon de convergence $R > 0$. Pour $\rho \in]0, R[$, on pose $N_{1,\rho}(f) = \sum_{n=0}^{+\infty} |\alpha_n| \rho^n$,

$$N_{2,\rho}(f) = \left(\sum_{n=0}^{+\infty} |\alpha_n|^2 \rho^{2n} \right)^{1/2}, \text{ et } N_{\infty,\rho}(f) = \sup_{|z|=\rho} |f(z)|.$$

1. Montrer les inégalités suivantes :

$$N_{\infty,\rho}(f) \leqslant N_{1,\rho}(f), \quad N_{2,\rho}(f) \leqslant N_{\infty,\rho}(f),$$

$$\text{et pour } 0 < \delta < R - \rho, \quad N_{1,\rho}(f) \leqslant N_{2,\rho+\delta} \frac{\rho + \delta}{\sqrt{\delta(2\rho + \delta)}}.$$

2. On fixe $\rho \in]0, R[$ et on pose $u_n = \left(N_{1,\rho}(f^n) \right)^{1/n}$ ainsi que $v_n = \left(N_{2,\rho}(f^n) \right)^{1/n}$. Déterminer les limites des suites $(u_n)_{n \geqslant 1}$ et $(v_n)_{n \geqslant 1}$. Montrer que $N_{\infty,\rho}(f)$ est une fonction croissante de ρ.

(**École polytechnique**)

▷ **Solution.**

1. La série définissant f converge normalement sur le cercle de centre 0 et de rayon ρ, ce qui entraîne l'existence de $N_{1,\rho}(f)$ et *a fortiori* celle de $N_{2,\rho}(f)$, puisque $|\alpha_n|^2 \rho^{2n} = o(|\alpha_n| \rho^n)$. L'existence de $N_{\infty,\rho}(f)$ résulte de la continuité de f et de la compacité du cercle $\{z \in \mathbb{C}, |z| = \rho\}$.

Pour tout z tel que $|z| = \rho$, on a $|f(z)| \leqslant \sum_{n=0}^{+\infty} |\alpha_n| \rho^n \leqslant N_{1,\rho}(f)$, ce qui montre que $N_{\infty,\rho}(f) \leqslant N_{1,\rho}(f)$.

La fonction $\theta \longmapsto f(\rho e^{i\theta})$ est 2π-périodique et continue. La série définissant $f(\rho e^{i\theta})$ (c'est-à-dire $\sum_{n=0}^{+\infty} \alpha_n \rho^n e^{in\theta}$) étant uniformément convergente par rapport à θ, on en déduit que les coefficients de Fourier de cette fonction sont les $\alpha_n \rho^n$, pour $n \geqslant 0$, et sont nuls pour $n < 0$. La formule de Parseval permet d'écrire

$$\begin{aligned} (N_{2,\rho}(f))^2 &= \sum_{n=0}^{+\infty} |\alpha_n|^2 \rho^{2n} = \frac{1}{2\pi} \int_0^{2\pi} |f(\rho e^{i\theta})|^2 \mathrm{d}\theta \\ &\leqslant \frac{1}{2\pi} \int_0^{2\pi} (N_{\infty,\rho}(f))^2 \mathrm{d}\theta = (N_{\infty,\rho}(f))^2, \end{aligned}$$

et donc

$$N_{2,\rho}(f) \leqslant N_{\infty,\rho}(f).$$

Enfin, pour $0 < \delta < R-\rho$, on écrit $N_{1,\rho}(f) = \sum_{n=0}^{+\infty} |\alpha_n|(\rho+\delta)^n \times \frac{\rho^n}{(\rho+\delta)^n}$ et on applique l'inégalité de Cauchy-Schwarz. On obtient

$$\begin{aligned} N_{1,\rho}(f) &\leqslant \left(\sum_{n=0}^{+\infty} |\alpha_n|^2(\rho+\delta)^{2n}\right)^{\frac{1}{2}} \left(\sum_{n=0}^{+\infty} \frac{\rho^{2n}}{(\rho+\delta)^{2n}}\right)^{\frac{1}{2}} \\ &\leqslant N_{2,\rho+\delta}(f) \left(\frac{1}{1-\dfrac{\rho^2}{(\rho+\delta)^2}}\right)^{\frac{1}{2}} \\ &\leqslant N_{2,\rho+\delta}(f) \frac{\rho+\delta}{\sqrt{\delta(\delta+2\rho)}}. \end{aligned}$$

2. Montrons que la suite (u_n) converge vers $N_{\infty,\rho}(f)$. On va utiliser les trois inégalités démontrées dans la question 1. On a d'une part $N_{\infty,\rho}(f) = (N_{\infty,\rho}(f^n))^{\frac{1}{n}} \leqslant u_n$, et d'autre part, pour $\delta \in]0, R-\rho[$,

$$\begin{aligned} u_n &\leqslant \left(N_{2,\rho+\delta}(f^n)\right)^{\frac{1}{n}} \left(\frac{\rho+\delta}{\sqrt{\delta(2\rho+\delta)}}\right)^{\frac{1}{n}} \leqslant (N_{\infty,\rho+\delta}(f^n))^{\frac{1}{n}} \left(\frac{\rho+\delta}{\sqrt{\delta(2\rho+\delta)}}\right)^{\frac{1}{n}} \\ &\leqslant N_{\infty,\rho+\delta}(f) \left(\frac{\rho+\delta}{\sqrt{\delta(2\rho+\delta)}}\right)^{\frac{1}{n}}. \end{aligned}$$

Montrons que la fonction $\rho \longmapsto N_{\infty,\rho}(f)$ est continue sur $]0, R[$. Fixons $\delta_0 \in]0, R-\rho[$. La fonction f est uniformément continue sur le disque fermé D de centre 0 et de rayon $\rho+\delta_0$. Pour tout $\varepsilon > 0$, il existe $\eta \in]0, \delta_0[$, tel que, pour tout $(z, z') \in D^2$, $|z-z'| \leqslant \eta$ implique $|f(z) - f(z')| \leqslant \varepsilon$. Si $\rho' \in]0, \rho+\delta_0[$ est tel que $|\rho'-\rho| \leqslant \eta$, alors, pour tout $\theta \in \mathbb{R}$, on a $|f(\rho' e^{i\theta}) - f(\rho e^{i\theta})| \leqslant \varepsilon$ et donc

$$|f(\rho' e^{i\theta})| \leqslant |f(\rho e^{i\theta})| + \varepsilon \leqslant N_{\infty,\rho}(f) + \varepsilon.$$

Ceci étant vrai pour tout θ, on en déduit que que $N_{\infty,\rho'}(f) \leqslant N_{\infty,\rho}(f) + \varepsilon$. Pour des raisons de symétries, on a $N_{\infty,\rho}(f) \leqslant N_{\infty,\rho'}(f) + \varepsilon$ et donc $|N_{\infty,\rho'}(f) - N_{\infty,\rho'}(f)| \leqslant \varepsilon$, ce qui démontre le résultat voulu.

On a donc $\lim\limits_{\delta \to 0^+} N_{\infty,\rho+\delta}(f) = N_{\infty,\rho}(f)$. Soit $\varepsilon > 0$. On choisit un réel $\delta \in]0, R-\rho[$ tel que $N_{\infty,\rho+\delta}(f) \leqslant N_{\infty,\rho}(f) + \varepsilon$, si bien que

$$u_n \leqslant (N_{\infty,\rho}(f) + \varepsilon) \left(\frac{\rho+\delta}{\sqrt{\delta(2\rho+\delta)}}\right)^{\frac{1}{n}},$$

pour tout $n \in \mathbb{N}$. Le membre de droite de cette inégalité tend vers $\mathrm{N}_{\infty,\rho}(f)+\varepsilon$, quand n tend vers $+\infty$. Il existe donc $n_0 \in \mathbb{N}$ tel que, pour $n \geqslant n_0$,

$$(\mathrm{N}_{\infty,\rho}(f)+\varepsilon)\left(\frac{\rho+\delta}{\sqrt{\delta(2\rho+\delta)}}\right)^{\frac{1}{n}} \leqslant \mathrm{N}_{\infty,\rho}(f)+2\varepsilon.$$

On a alors $\mathrm{N}_{\infty,\rho}(f) \leqslant u_n \leqslant \mathrm{N}_{\infty,\rho}(f)+2\varepsilon$ si $n \geqslant n_0$. Ceci démontre que

$$\lim_{n\to+\infty} u_n = \mathrm{N}_{\infty,\rho}(f).$$

Montrons que (v_n) converge vers la même limite. La méthode est la même que pour (u_n). En remplaçant ρ par $\rho-\delta$ dans la dernière inégalité de la question 1. on obtient

$$\mathrm{N}_{\infty,\rho-\delta}(f^n) \leqslant \mathrm{N}_{1,\rho-\delta}(f^n) \leqslant \mathrm{N}_{2,\rho}(f^n)\frac{\rho}{\sqrt{\delta(\rho+\delta)}},$$

$$\mathrm{N}_{\infty,\rho-\delta}(f^n)\frac{\sqrt{\delta(\rho+\delta)}}{\rho} \leqslant \mathrm{N}_{2,\rho}(f^n) \leqslant \mathrm{N}_{\infty,\rho}(f^n),$$

$$\mathrm{N}_{\infty,\rho-\delta}(f)\left(\frac{\sqrt{\delta(\rho+\delta)}}{\rho}\right)^{\frac{1}{n}} \leqslant v_n \leqslant \mathrm{N}_{\infty,\rho}.$$

Soit $\varepsilon>0$ et $\delta \in]0,\mathrm{R}-\rho[$ tel que $\mathrm{N}_{\infty,\rho-\delta}(f) \geqslant \mathrm{N}_{\infty,\rho}(f)-\varepsilon$. Sachant que $\lim_{n\to+\infty}\left(\frac{\sqrt{\delta(\rho+\delta)}}{\rho}\right)^{\frac{1}{n}}=1$, on aura

$$\mathrm{N}_{\infty,\rho-\delta}(f)\left(\frac{\sqrt{\delta(\rho+\delta)}}{\rho}\right)^{\frac{1}{n}} \geqslant \mathrm{N}_{\infty,\rho}(f)-2\varepsilon$$

et donc $\mathrm{N}_{\infty,\rho}(f)-2\varepsilon \leqslant v_n \leqslant \mathrm{N}_{\infty,\rho}(f)$, pour n assez grand, ce qui démontre que (v_n) converge vers $\mathrm{N}_{\infty,\rho}(f)$.

D'après la question 1, on a, pour $\delta \in]0,\mathrm{R}-\rho[$,

$$\mathrm{N}_{1,\rho}(f^n) \leqslant \mathrm{N}_{2,\rho+\delta}(f^n)\frac{\rho+\delta}{\sqrt{\delta(\delta+2\rho)}}$$

et donc

$$(\mathrm{N}_{1,\rho}(f^n))^{\frac{1}{n}} \leqslant (\mathrm{N}_{2,\rho+\delta}(f^n))^{\frac{1}{n}}\left(\frac{\rho+\delta}{\sqrt{\delta(\delta+2\rho)}}\right)^{\frac{1}{n}}.$$

D'après ce qui vient d'être démontré, le membre de gauche de l'inégalité, qui est égal à u_n, tend vers $\mathrm{N}_{\infty,\rho}(f)$ quand n tend vers $+\infty$. L'expression $(\mathrm{N}_{2,\rho+\delta}(f^n))^{\frac{1}{n}}$, qui est la valeur de v_n obtenue quand on remplace ρ par $\rho+\delta$, tend vers $\mathrm{N}_{\infty,\rho+\delta}(f)$. Enfin, $\left(\frac{\rho+\delta}{\sqrt{\delta(\delta+2\rho)}}\right)^{\frac{1}{n}}$ tend vers 1. On en

déduit, en faisant tendre n vers $+\infty$, que $N_{\infty,\rho}(f) \leqslant N_{\infty,\rho+\delta}(f)$: la fonction $\rho \longmapsto N_{\infty,\rho}(f)$ est croissante. $\lhd$

De la croissance de la fonction $\rho \longmapsto N_{\infty,\rho}(f)$, découle directement le principe du maximum : pour $\rho \in]0,R[$,

$$\max_{|z|\leqslant\rho} |f(z)| = \max_{|z|=\rho} |f(z)|.$$

On retrouve notamment le résultat de l'exercice précédent.

Voici encore un exercice sur le même thème.

3.41. Extrema d'une fonction entière

1. Que peut-on dire d'un polynôme $P \in \mathbb{C}[X]$ tel que $|P|$ admette un extremum local en un point de $\mathbb{C}$?

2. Même question pour une fonction somme d'une série entière de rayon de convergence infini.

(École normale supérieure)

▷ **Solution.**

1. Soit $P \in \mathbb{C}[X]$. Supposons que $|P|$ possède un extremum local en z_0. Si $P(z_0) = 0$, c'est-à-dire si z_0 est une racine de P, alors $|P|$ possède un minimum global en z_0. On suppose désormais que $P(z_0) \neq 0$ et on va montrer qu'alors le polynôme P est constant. En considérant le polynôme $Q = P(z_0 + X)$ on se ramène au cas où $z_0 = 0$. On a $P = \sum_{j=0}^{d} a_j X^j$, avec $a_0 \neq 0$. Supposons que P ne soit pas un polynôme constant, c'est-à-dire qu'il existe $j \geqslant 1$ tel que $a_j \neq 0$. On note k le plus petit indice supérieur ou égal à 1 tel que $a_k \neq 0$ et on pose $\dfrac{a_k}{a_0} = b$. Il existe un polynôme R tel que $R(0) = 0$ et $P = a_0(1 + bX^k + X^k R)$, puis $\rho > 0$ tel que $|z| \leqslant \rho$ implique $|R(z)| \leqslant \dfrac{1}{2}|b|$. On écrit, pour tout $z \in \mathbb{C}$, $\dfrac{|P(z)|}{|a_0|} = |1 + bz^k + z^k R(z)|$. On va choisir z tel que bz^k soit réel. Soit θ un argument de b. Pour tout $0 < r < \rho$ et $z = re^{-i\frac{\theta}{k}}$, on obtient

$$\frac{|P(z)|}{|a_0|} = \left|1 + |b|r^k + z^k R(z))\right| \quad \text{et} \quad |z^k P(z)| \leqslant \frac{1}{2} r^k,$$

d'où l'on déduit

$$\frac{|P(z)|}{|a_0|} \geqslant 1 + \frac{1}{2}|b|r^k > 1.$$

On a des z aussi proches que l'on veut de 0 pour lesquels $|P(z)| > |a_0|$. La fonction $|P|$ ne présente pas de maximum local en 0. Mais si on prend

$0 < r < \rho$ et $z = re^{-i\frac{\theta+\pi}{k}}$, on obtient

$$\frac{|P(z)|}{|a_0|} = \left|1 - |b|r^k + z^k R(z)\right| \quad \text{et} \quad |z^k P(z)| \leqslant \frac{1}{2}r^k,$$

d'où l'on déduit

$$\frac{|P(z)|}{|a_0|} \leqslant 1 - \frac{1}{2}|b|r^k < 1.$$

On a des z aussi proches que l'on veut de 0 pour lesquels $|P(z)| < |a_0|$. La fonction $|P|$ ne présente pas de minimum local en 0.

On obtient donc une contradiction et on conclut. Si P possède un extremum local en z_0 et si $P(z_0) \neq 0$, le polynôme P est constant. *Ce qui précède fournit une démonstration du théorème de d'Alembert-Gauss. Soit* P *un polynôme non constant. On pose* $m = \inf |P|$. *Comme* $\lim_{|z|\to+\infty} |P(z)| = +\infty$, *il existe* $r > 0$ *tel que* $|z| \geqslant r$ *implique* $|P(z)| \geqslant m+1$. *On a donc* $m = \inf_{|z|\leqslant r} |P(z)|$ *et comme le disque fermé de centre* 0 *et de rayon* r *est compact, il existe* z_0 *tel que* $|P(z_0)| = m$. *La fonction* $|P|$ *possède donc un minimum en* z_0 *et comme* P *n'est pas constant, on a* $P(z_0) = 0$.

2. Soit f la somme d'une série entière de rayon de convergence infini. On a, pour tout $z \in \mathbb{C}$, $f(z) = \sum_{n=0}^{+\infty} a_n z^n$. Montrons que f est développable en série entière au voisinage de tout point avec un rayon de convergence infini. Soit $z_0 \in \mathbb{C}$. On a pour tout $z \in \mathbb{C}$,

$$f(z) = \sum_{n=0}^{+\infty} a_n (z - z_0 + z_0)^n = \sum_{n=0}^{+\infty} a_n \sum_{k=0}^{n} \binom{n}{k} (z - z_0)^k z_0^{n-k}.$$

Pour tout $(n,k) \in \mathbb{N}^2$, on pose

$$u_{n,k} = \begin{cases} a_n \binom{n}{k} (z - z_0)^k z_0^{n-k} & \text{si} \quad k \leqslant n \\ 0 & \text{sinon.} \end{cases}$$

On a, pour tout $n \in \mathbb{N}$,

$$\sum_{k=0}^{+\infty} |u_{n,k}| = \sum_{k=0}^{n} |a_n| \binom{n}{k} |z - z_0|^k |z_0|^{n-k} = |a_n| (|z - z_0| + |z_0|)^n.$$

La série entière $\sum |a_n| z^n$ a un rayon de convergence infini comme $\sum a_n z^n$, car pour tout $z \in \mathbb{C}$, la suite $(|a_n| z^n)$ est bornée. On en déduit que $\sum |a_n| (|z - z_0| + |z_0|)^n$ converge. Ainsi la série double $\sum u_{n,k}$ est absolument convergente et sa somme est

$$\sum_{n=0}^{+\infty} \sum_{k=0}^{+\infty} u_{n,k} = \sum_{n=0}^{+\infty} a_n z^n = f(z).$$

Or, en échangeant l'ordre de sommation on a

$$\sum_{k=0}^{+\infty}\sum_{n=0}^{+\infty} u_{n,k} = \sum_{k=0}^{+\infty}\sum_{n=k}^{+\infty}\binom{n}{k}(z-z_0)^k z_0^{n-k} = \sum_{k=0}^{+\infty}(z-z_0)^k \sum_{n=k}^{+\infty}\binom{n}{k} z_0^{n-k}.$$

En posant $b_k = \sum_{n=k}^{+\infty}\binom{n}{k} z_0^{n-k}$, on a donc $f(z) = \sum_{k=0}^{+\infty} b_k(z-z_0)^k$ pour tout $z \in \mathbb{C}$.

Supposons que $|f|$ possède un extremum local en z_0. On montre comme dans la première question que si $f(z_0) \neq 0$, *i.e.* $b_0 \neq 0$, la fonction f est constante. Si f n'est pas constante, il existe $j \geqslant 1$ tel que $b_j \neq 0$. On note k le plus petit indice supérieur ou égal à 1 tel que $b_k \neq 0$ et on pose $b = \dfrac{b_k}{b_0}$. Il existe une fonction g développable en série entière au voisinage de z_0 telle que $g(z_0) = 0$ et

$$f(z) = b_0\left(1 + b(z-z_0)^k + (z-z_0)^k g(z)\right),$$

où $g(z) = \dfrac{1}{b_0}\sum_{n=k+1}^{+\infty} b_n(z-z_0)^{n-k}$. Par continuité de g, il existe $\rho > 0$ tel que $|g(z)| \leqslant \dfrac{1}{2}|b|$. En notant θ un argument de b, on obtient $|\mathrm{P}(z)| < |b_0|$ si $0 < r < \rho$ et $z = re^{-i\frac{\theta}{k}}$ et $|\mathrm{P}(z)| > |b_0|$ si $0 < r < \rho$ et $z = re^{-i\frac{\theta+\pi}{k}}$. On trouve dans tout voisinage de z_0 des points où $|\mathrm{P}(z)| < |b_0|$ et d'autres où $|\mathrm{P}(z)| > |b_0|$, ce qui contredit l'hypothèse.

Conclusion. Si $|f|$ possède un extremum local en z_0 et si $f(z_0) \neq 0$, la fonction f est constante. ◁

Les deux exercices suivants traitent de la convergence des suites de fonctions développables en série entière sur le disque de centre 0 et de rayon 1. Le premier montre que si les coefficients des différentes fonctions sont bornés uniformément, la convergence simple d'une suite (f_n) vers f équivaut à la convergence, pour tout k, des suites $(a_k(f_n))_{n\in\mathbb{N}}$ vers $a_k(f)$ où $a_k(f)$ désigne le coefficient de z^k dans le développement de f.

3.42. Critère de convergence simple d'une suite de séries entières

Soit $\mathrm{M} > 0$ et $\mathcal{S}_\mathrm{M}$ l'ensemble des fonctions g de la variable réelle à valeurs dans $\mathbb{C}$, développables en série entière sur $]-1,1[$ et telles que si, pour tout $x \in]-1,1[$, $g(x) = \sum_{k=0}^{+\infty} a_k(g)x^k$, on ait, pour tout $k \in \mathbb{N}$, $|a_k(g)| \leqslant \mathrm{M}$.

On considère $g \in \mathcal{S}_M$ et une suite $(g_n)_{n\in\mathbb{N}}$ de fonctions appartenant à $\mathcal{S}_M$. Montrer que chaque suite $(a_k(g_n))_{n\in\mathbb{N}}$ $(k \in \mathbb{N})$ converge vers $a_k(g)$ si et seulement si la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement vers g. Y a-t-il alors convergence uniforme ?

(École normale supérieure)

Solution.

• Supposons que, pour tout $k \in \mathbb{N}$, la suite $(a_k(g_n))_{n\in\mathbb{N}}$ converge vers $a_k(g)$. Alors, si $x \in]-1,1[$ et $n \in \mathbb{N}$,

$$\begin{aligned}|g(x) - g_n(x)| &\leqslant \sum_{k=0}^{+\infty} |a_k(g_n) - a_k(g)||x|^k \\ &\leqslant \sum_{k=0}^{K} |a_k(g_n) - a_k(g)| + \sum_{k\geqslant K+1} 2M|x|^k \\ &\leqslant \sum_{k=0}^{K} |a_k(g_n) - a_k(g)| + \frac{2M|x|^{K+1}}{1-|x|}.\end{aligned}$$

Soit $\varepsilon > 0$. Fixons un entier K tel que $\dfrac{2M|x|^{K+1}}{1-|x|} < \varepsilon$. Comme on a $\lim\limits_{n\to+\infty} \sum\limits_{k=0}^{K} |a_k(g_n) - a_k(g)| = 0$, il existe un entier n_0 tel que, pour tout $n \geqslant n_0$, $\sum\limits_{k=0}^{K} |a_k(g_n) - a_k(g)| < \varepsilon$ et donc $|g_n(x) - g(x)| < 2\varepsilon$. Cela montre que la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement vers g.

On remarque, de plus, que la convergence est uniforme sur $[-a,a]$, pour tout $a \in [0,1[$. En effet, on a pour tout $x \in [-a,a]$, la majoration

$$|g_n(x) - g(x)| \leqslant \sum_{k=0}^{K} |a_k(g_n) - a_k(g)| + \frac{2Ma^{N+1}}{1-a},$$

ce qui permet de trouver un n_0 indépendant de x.

En revanche, la convergence n'est pas nécessairement uniforme sur l'intervalle $]-1,1[$. En effet si on prend $g(x) = \dfrac{1}{1-x} = \sum\limits_{k=0}^{+\infty} x^k$ et pour tout $n \in \mathbb{N}$, $g_n(x) = \dfrac{n}{n+1} g(x) = \sum\limits_{k=0}^{+\infty} \dfrac{n}{n+1} x^k$, alors, pour tout $k \in \mathbb{N}$, $a_k(g_n) = \dfrac{n}{n+1} \xrightarrow[n\to\infty]{} 1 = a_k(g)$. Mais pour tout $n \in \mathbb{N}^*$, $\left|g_n\left(1-\frac{1}{n}\right) - g\left(1-\frac{1}{n}\right)\right| = \dfrac{n}{n+1} \xrightarrow[n\to\infty]{} 1$ et il n'y a donc pas convergence uniforme de $(g_n)_{n\in\mathbb{N}}$ vers g sur $]-1,1[$.

• Supposons réciproquement que la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement vers g. Il n'est pas possible d'utiliser ici l'expression intégrale des coefficients $a_k(g_n) = \dfrac{1}{2\pi r^k}\displaystyle\int_0^{2\pi} g_n(re^{i\theta})e^{-ik\theta}\mathrm{d}\theta$ et de passer à la limite à l'aide du théorème de convergence dominée, car on sait seulement que la suite $(g_n)_{n\in\mathbb{N}}$ converge simplement vers g sur l'intervalle $]-1,1[$.

On va prouver que $a_k(g_n)$ converge vers $a_k(g)$ par récurrence sur k. Pour $k=0$ c'est facile, car $a_0(g_n) = g_n(0)$ converge par hypothèse vers $g(0) = a_0(g)$. Pour passer au rang 1, on a envie de considérer la suite des fonctions $x \mapsto \dfrac{g_n(x) - a_0(g_n)}{x}$. Cette suite converge simplement vers $\dfrac{g(x) - a_0(g)}{x}$ mais a priori seulement pour $x \neq 0$. On ne peut donc pas utiliser la valeur 0 comme avant pour obtenir que $a_1(g_k)$ converge vers $a_1(g)$. En fait, on va montrer que dans le cas $k=0$, la convergence simple sur l'intervalle ouvert $]0,1[$ suffit amplement. En effet, soit $\varepsilon > 0$. On a, pour tout $x \in]0,1[$,

$$\begin{aligned}|a_0(g_n) - a_0(g)| &= \left|g_n(x) - g(x) + \sum_{k=1}^{+\infty}(a_k(g) - a_k(g_n))x^k\right| \\ &\leqslant |g_n(x) - g(x)| + \frac{2\mathrm{M}x}{1-x}.\end{aligned}$$

On choisit d'abord $x \in]0,1[$ tel que le second terme soit majoré par ε. Pour cette valeur de x, il existe un rang N tel que pour $n \geqslant \mathrm{N}$, $|g_n(x) - g(x)| \leqslant \varepsilon$ et donc $|a_0(g_n) - a_0(g)| \leqslant 2\varepsilon$. Cela montre que $a_0(g_n)$ converge vers $a_0(g)$.

À partir de ce résultat la récurrence se déroule sans problème. Supposons que $a_i(g_n)$ converge vers $a_i(g)$ pour tout $i \leqslant k$. On considère les fonctions

$$x \mapsto \frac{g_n(x) - a_0(g_n) - a_1(g_n)x - \cdots - a_k(g_n)x^k}{x^{k+1}}.$$

Il est clair que cette suite de fonctions converge simplement sur $]0,1[$ vers la fonction $x \mapsto \dfrac{g(x) - a_0(g) - a_1(g)x - \cdots - a_k(g)x^k}{x^{k+1}}$, et qu'elles sont toujours dans $\mathcal{S}_\mathrm{M}$. Le résultat qu'on vient de prouver montre alors que $a_{k+1}(g_n)$ converge vers $a_{k+1}(g)$ ce qui termine la récurrence. ◁

Si on note comme d'habitude D *le disque ouvert de centre* 0 *et de rayon* 1, *l'ensemble des fonctions continues sur* $\overline{\mathrm{D}}$ *et développables en séries entière sur* D *est complet pour la norme* $\|\ \|_\infty$. *La preuve de ce résultat fait l'objet de l'exercice suivant.*

3.43. Complétude de l'espace des fonctions continues sur $\overline{\mathrm{D}}$, développables en série entière sur D

Soit $\mathrm{D} = \{z \in \mathbb{C}, |z| < 1\}$ et E l'ensemble des fonctions continues sur $\overline{\mathrm{D}}$ et développables en série entière dans D. Pour $f \in \mathrm{E}$ et $z \in \mathrm{D}$, on écrit $f(z) = \sum_{n=0}^{+\infty} a_n(f) z^n$.

1. Montrer que $a_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} \mathrm{d}\theta$ pour tout $f \in \mathrm{E}$ et tout $n \in \mathbb{N}$.

2. Montrer que si une suite de E converge uniformément sur $\overline{\mathrm{D}}$, sa limite est encore dans E.

3. Montrer que E est complet pour la norme de la convergence uniforme.

4. Montrer que le sous-espace des polynômes est dense dans E.

(École polytechnique)

Solution.

1. Soit $0 < r < 1$. La série entière $\sum_{n=0}^{+\infty} a_n(f) z^n$ converge normalement sur le cercle de rayon r centré en 0, si bien que pour n entier naturel

$$\int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} \mathrm{d}\theta = \int_0^{2\pi} \sum_{p=0}^{+\infty} a_p(f) r^p e^{i(p-n)\theta} \mathrm{d}\theta$$

$$= \sum_{p=0}^{\infty} a_p(f) r^p \int_0^{2\pi} e^{i(p-n)\theta} \mathrm{d}\theta = 2\pi a_n(f) r^n.$$

La fonction $(r, \theta) \longmapsto f(re^{i\theta}) e^{-in\theta}$ est continue sur $[0,1] \times [0, 2\pi]$, ce qui assure la continuité de l'intégrale à paramètre $r \longmapsto \int_0^{2\pi} f(re^{i\theta}) e^{-in\theta} \mathrm{d}\theta$ sur $[0,1]$. En faisant tendre r vers 1, il vient

$$\boxed{a_n(f) = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} \mathrm{d}\theta}.$$

2. Soit $(f_k)_{k \geqslant 0}$ une suite de E qui converge uniformément sur $\overline{\mathrm{D}}$ vers une fonction f. Les f_k étant continues, f est continue sur $\overline{\mathrm{D}}$. Il reste à prouver que f est développable en série entière dans D. On pose naturellement $\alpha_n = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) e^{-in\theta} \mathrm{d}\theta$ et on va montrer que pour tout $z \in \mathrm{D}$, $f(z) = \sum_{n=0}^{+\infty} \alpha_n z^n$. Soit donc z un point fixé de D. On a $|\alpha_n| \leqslant \|f\|_\infty$ pour tout n, donc la série $\sum_{n=0}^{+\infty} \alpha_n z^n$ converge. On a pour

tout entier k,

$$\left|f(z)-\sum_{n=0}^{+\infty}\alpha_n z^n\right| \leqslant |f(z)-f_k(z)| + \left|\sum_{n=0}^{+\infty}(a_n(f_k)-\alpha_n)z^n\right|.$$

Or, d'après la question 1 on a, pour tout n, $|a_n(f_k)-\alpha_n| \leqslant \|f-f_k\|_\infty$. Il vient finalement, pour tout k,

$$\left|f(z)-\sum_{n=0}^{+\infty}\alpha_n z^n\right| \leqslant \|f-f_k\|_\infty\left(1+\frac{1}{1-|z|}\right).$$

Il suffit de faire tendre k vers l'infini pour conclure.

Conclusion. f est bien développable en série entière sur D et continue sur $\overline{\mathrm{D}}$: f est dans E.

3. Considérons $(f_k)_{k\geqslant 0}$ une suite de E de Cauchy pour la norme de la convergence uniforme. L'espace $\mathcal{C}(\overline{\mathrm{D}},\mathbb{C})$ des fonctions continues sur $\overline{\mathrm{D}}$ est complet pour la norme $\|\ \|_\infty$. Il existe donc f continue sur $\overline{\mathrm{D}}$ telle que la suite (f_k) converge uniformément vers f. D'après la question précédente, f est dans E. Ainsi la suite (f_k) converge vers f dans $(\mathrm{E},\|\ \|_\infty)$.

Conclusion. L'espace E muni de la norme de la convergence uniforme $\|\ \|_\infty$ est complet.

4. Soit $f\in\mathrm{E}$. Il s'agit d'exhiber une suite de polynômes qui converge uniformément vers f sur $\overline{\mathrm{D}}$. Il n'est pas question de prendre les sommes partielles du développement en série entière de f : la convergence est certes uniforme sur tout compact inclus dans D, mais cela est loin d'être suffisant (cette série peut même ne pas converger en un point du bord de D). L'idée est d'introduire une autre fonction g, développable en série entière avec un rayon >1, et telle que $|f(z)-g(z)|<\varepsilon$ en tout point z, puis de prendre une somme partielle de la série entière de g pour approcher g uniformément à ε près sur $\overline{\mathrm{D}}$. Pour construire g, on utilise le fait que f est uniformément continue sur le compact $\overline{\mathrm{D}}$. Soit $\varepsilon>0$ fixé et $\eta>0$ un ε-module d'uniforme continuité de f sur $\overline{\mathrm{D}}$. Posons alors $g(z)=f\left(\dfrac{z}{1+\eta}\right)$ pour $|z|\leqslant 1+\eta$. Pour tout z de $\overline{\mathrm{D}}$ on a $|f(z)-g(z)|\leqslant\varepsilon$ car $\left|z-\dfrac{z}{1+\eta}\right|\leqslant\eta$. De plus, il est clair que g est développable en série entière sur le disque de centre 0 et de rayon $1+\eta$. Comme $\overline{\mathrm{D}}$ est un compact inclus dans ce disque ouvert, la série entière de g converge uniformément vers g sur $\overline{\mathrm{D}}$. En prenant une somme partielle, on peut donc trouver un polynôme P tel que $|\mathrm{P}(z)-g(z)|\leqslant\varepsilon$ pour tout $z\in\overline{\mathrm{D}}$. Par inégalité triangulaire, on a $|f(z)-\mathrm{P}(z)|\leqslant 2\varepsilon$ pour tout z de $\overline{\mathrm{D}}$. Cela implique la densité de l'espace des polynômes dans E. ◁

Le lecteur pourra noter que l'espace des polynômes n'est pas dense dans l'espace de toutes les fonctions continues sur $\overline{\mathrm{D}}$: par exemple la fonction $z\mapsto\bar{z}$ n'est pas limite uniforme sur $\overline{\mathrm{D}}$ d'une suite de polynômes

(raisonner par l'absurde en supposant l'existence d'une telle suite (P_n) *et regarder les intégrales* $\int_0^{2\pi} \mathrm{P}_n(e^{i\theta})e^{i\theta}\,d\theta$*). Il y a donc une différence sensible avec le théorème de Weierstrass réel étudié en cours.*

Les exercices qui terminent ce chapitre sont très difficiles et atteignent les limites de ce qu'on peut faire sans se lancer dans une véritable étude de l'analyse complexe. Donnons donc, en préambule, quelques résultats sur la dérivabilité au sens complexe.

Les fonctions définies par des séries entières sont dérivables par rapport à la variable complexe sur leur disque de convergence, c'est-à-dire que, pour tout $z_0 \in \mathrm{D}$*, le rapport* $\dfrac{f(z)-f(z_0)}{z-z_0}$ *a une limite quand* $z \in \mathbb{C}$ *tend vers* z_0*. On obtient* f'*, comme dans* $\mathbb{R}$*, en dérivant terme à terme la série entière définissant* f*. Démontrons ce résultat.*

Posons $f(z) = \sum_{n=0}^{+\infty} a_n z^n$*, pour* $|z| < \mathrm{R}$*, rayon de convergence de la série. Le rayon de convergence de la série dérivée est encore égal à* R*. On pose, pour* $|z| < \mathrm{R}$*,* $g(z) = \sum_{n=1}^{+\infty} n a_n z^{n-1}$*. Soit* $z_0 \in \mathrm{D}$ *et* r *tel que* $|z_0| < r < \mathrm{R}$*. On a, pour* $|z| < r$*,* $z \neq z_0$*.*

$$\frac{f(z)-f(z_0)}{z-z_0} = \sum_{n=1}^{+\infty} a_n \left(\sum_{k=0}^{n-1} z_0^k z^{n-1-k} \right).$$

De la majoration $\left| \sum_{k=0}^{n-1} z_0^k z^{n-1-k} \right| \leqslant n r^{n-1}$*, on déduit que, pour* $\mathrm{N} \in \mathbb{N}^*$*,*

$$\left| \frac{f(z)-f(z_0)}{z-z_0} - g(z_0) \right| \leqslant \sum_{n=1}^{\mathrm{N}-1} |a_n| \left| \sum_{k=0}^{n-1} z_0^k z^{n-1-k} - n z_0^{n-1} \right| + 2 \sum_{n=\mathrm{N}+1}^{+\infty} |a_n| n r^{n-1}.$$

On choisit N *tel que que le deuxième terme soit* $\leqslant \varepsilon$ *(c'est possible car la série converge) puis, le premier terme tendant vers* 0 *quand* z *tend vers* z_0*, on choisit* $\eta > 0$ *tel que* $\left| \dfrac{f(z)-f(z_0)}{z-z_0} - g(z_0) \right| \leqslant 2\varepsilon$ *pour* $|z-z_0| \leqslant \eta$*. C'est le résultat voulu.*

Notons que la condition de dérivabilité par rapport à la variable complexe est très forte. On peut démontrer que si une fonction f *est continûment dérivable sur* $\mathrm{D} = \{z \in \mathbb{C}, |z| < \mathrm{R}\}$*, alors* f *est développable en série entière sur* D*.*

3.44. Logarithme d'une fonction entière qui ne s'annule pas

Soit $(a_n)_{n\geqslant 0}$ une suite de nombres complexes telle que $\sum a_n z^n$ ait un rayon de convergence infini. On pose $f(z) = \sum_{n=0}^{+\infty} a_n z^n$ et on suppose que, pour tout $z \in \mathbb{C}$, $f(z) \neq 0$. Existe-t-il une suite $(b_n)_{n\geqslant 0}$ telle que,

$$\forall z \in \mathbb{C},\ \exp\left(\sum_{n=0}^{+\infty} b_n z^n\right) = f(z)\ ?$$

(École polytechnique)

▷ **Solution.**

La réponse est positive. L'application $\exp : \mathbb{C} \longrightarrow \mathbb{C}^*$ est surjective. Soit b_0 telle que

$$\exp(b_0) = f(0) = a_0.$$

On cherche une fonction g développable en série entière sur $\mathbb{C}$, telle que $g(0) = b_0$ et, pour tout $z \in \mathbb{C}$, $e^{g(z)} = f(z)$.

• *Analyse.* Comme f est développable en série entière avec un rayon de convergence infini, elle est dérivable par rapport à la variable complexe sur $\mathbb{C}$. Si g est une fonction développable en série entière telle que $e^g = f$ alors $f' = g' e^g = g' f$ et donc $g' = \dfrac{f'}{f}$.

• *Synthèse.* Supposons que l'on ait démontré que $h = \dfrac{f'}{f}$ est développable en série entière avec un rayon de convergence infini. Soit g la primitive de h telle que $g(0) = b_0$, c'est-à-dire que si $h(z) = \sum_{n=0}^{+\infty} d_n z^n$, alors $g(z) = b_0 + \sum_{n=0}^{+\infty} \dfrac{d_n}{n+1} z^{n+1}$. La série qui définit g a aussi un rayon de convergence infini. On a alors $(fe^{-g})' = (f' - g'f)e^{-g} = (f' - hf)e^{-g} = 0$. Montrons que la fonction $k = fe^{-g}$ est constante. On se ramène à la variable réelle. Soit $z \in \mathbb{C}$. Pour $t \in [0,1]$, on pose $k_z(t) = k(zt)$. Pour tout $t \in [0,1]$, $k_z'(t) = zk'(zt) = 0$ donc

$$k(z) = k_z(1) = k_z(0) = k(0) = f(0)e^{-g(0)} = a_0 e^{-b(0)} = 1,$$

par le choix de b_0. On en déduit que $f = e^g$. Ainsi on a déterminé une fonction g développable en série entière sur $\mathbb{C}$ telle que $e^g = f$.

Pour finir il reste donc à démontrer que $\dfrac{f'}{f}$ est développable en série entière sur $\mathbb{C}$. Pour cela il suffit de montrer que la fonction $\varphi = \dfrac{1}{f}$ l'est puisqu'on sait déjà que f' est développable en série entière avec

un rayon de convergence infini. On cherche une suite $(d_n)_{n\geqslant 0}$ telle que $\sum d_n z^n$ ait un rayon de convergence infini, et telle que pour tout $z\in\mathbb{C}$, $\varphi(z)=\sum_{n=0}^{+\infty} d_n z^n$. En écrivant que le produit de Cauchy $f\varphi$ est égal à la fonction constante 1 on obtient $a_0d_0=1$ et pour tout $n\geqslant 1$, $\sum_{k=0}^{n} a_k d_{n-k}=0$, c'est-à-dire

$$d_0=\frac{1}{a_0}\quad\text{et}\quad\forall n\geqslant 1,\ d_n=-\frac{1}{a_0}\sum_{k=1}^{n}a_k d_{n-k}.$$

Ces relations définissent une suite $(d_n)_{n\geqslant 0}$ unique (car $a_0=f(0)\neq 0$).

Montrons déjà que le rayon de convergence de la série entière $\sum d_n z^n$ n'est pas nul. Le rayon de convergence de la série entière $\sum |a_n|z^n$ est infini comme celui de $\sum a_n z^n$. Soit $R>0$ telle que $\sum_{n=1}^{+\infty}|a_n|R^n\leqslant |a_0|$ (un tel R existe car $\lim_{z\to 0}\sum_{n=1}^{+\infty}|a_n|z^n=0$). Montrons qu'il existe $K>0$ tel que $|d_n|\leqslant \frac{K}{R^n}$ pour tout $n\in\mathbb{N}$. On choisit $K\geqslant |d_0|$. Supposons que la propriété soit vérifiée jusqu'au rang $n-1$. On a alors

$$\begin{aligned}|d_n|&\leqslant\frac{1}{|a_0|}\sum_{k=1}^{n}|a_k|\,|d_{n-k}|\leqslant\frac{1}{|a_0|}\sum_{k=1}^{n}|a_k|\frac{K}{R^{n-k}}\\&\leqslant\frac{K}{|a_0|R^n}\sum_{k=1}^{n}|a_k|R^k\leqslant\frac{K}{R^n}.\end{aligned}$$

Le rayon de convergence de la série $\sum d_n z^n$ est donc supérieur ou égal à R et, par construction, on a $\frac{1}{f(z)}=\varphi(z)=\sum_{n=0}^{+\infty}d_n z^n$ pour $|z|<R$.

On va maintenant montrer que cette égalité reste valable pour tout $z\in\mathbb{C}$. Pour $r>0$, considérons la fonction $\varphi_r: t\longmapsto\varphi(re^{it})$. La fonction φ_r est 2π-périodique et de classe $\mathcal{C}^1$ (car f ne s'annule pas et car il est clair que $t\longmapsto f(re^{it})$ est de classe $\mathcal{C}^1$). Elle est donc développable en série de Fourier et

$$\varphi(re^{it})=\varphi_r(t)=\sum_{n=-\infty}^{+\infty}c_n(\varphi_r)e^{int}.$$

Or, lorsque $0\leqslant r<R$, on a

$$\varphi_r(t)=\varphi(re^{it})=\sum_{k=0}^{+\infty}d_k r^k e^{ikt}.$$

Comme la série est uniformément convergente par rapport à t, on peut échanger somme et intégrale. On obtient, pour $0\leqslant r<R$ et $n\in\mathbb{Z}$,

$$c_n(\varphi_r) = \frac{1}{2\pi}\sum_{k=0}^{+\infty}\int_0^{2\pi} d_k r^k e^{i(k-n)t}\,\mathrm{d}t,$$

et donc,

$$c_n(\varphi_r) = \begin{cases} d_n r^n & \text{si } n \geqslant 0 \\ 0 & \text{sinon.} \end{cases}$$

On va montrer que ces valeurs des coefficients de Fourier restent valides pour $r \geqslant \mathrm{R}$. Posons pour $r > 0$, $\psi(r) = r^{-n}\dfrac{1}{2\pi}\displaystyle\int_0^{2\pi}\varphi(re^{it})e^{-int}\,\mathrm{d}t$. Comme φ est continûment dérivable, la fonction $(r,t) \longmapsto \varphi(re^{it})e^{-int}\,\mathrm{d}t$ est de classe $\mathcal{C}^1$ sur $\mathbb{R}_+^* \times [0, 2\pi]$, et on déduit du théorème de dérivation des intégrales à paramètre que ψ est de classe $\mathcal{C}^1$ sur $\mathbb{R}_+^*$. On a pour tout $r > 0$,

$$\begin{aligned}
\psi'(r) &= -nr^{-n-1}\frac{1}{2\pi}\int_0^{2\pi}\varphi(re^{it})e^{-int}\,\mathrm{d}t + r^{-n}\frac{1}{2\pi}\int_0^{2\pi} e^{it}\varphi'(re^{it})e^{-int}\,\mathrm{d}t \\
&= \frac{r^{-n-1}}{2\pi}\int_0^{2\pi}(-n\varphi(re^{it})e^{-int} + re^{it}\varphi'(re^{it})e^{-int})\,\mathrm{d}t \\
&= \frac{r^{-n-1}}{2i\pi}\int_0^{2\pi}\frac{\mathrm{d}\left(\varphi(re^{it})e^{-int}\right)}{\mathrm{d}t}\,\mathrm{d}t = \frac{r^{-n-1}}{2i\pi}[\varphi(re^{it})e^{-int}]_0^{2\pi} = 0.
\end{aligned}$$

La fonction ψ est donc constante. Compte tenu de la valeur de $\psi(r)$ pour $r < \mathrm{R}$, on obtient, pour tout $r \geqslant 0$,

$$c_n(\varphi_r) = r^n\psi(r) = \begin{cases} b_n r^n & \text{si } n \geqslant 0 \\ 0 & \text{sinon.} \end{cases}$$

Finalement, pour tout $r > 0$ et tout $t \in \mathbb{R}$,

$$\varphi(re^{it}) = g_r(t) = \sum_{n=-\infty}^{+\infty} c_n(\varphi_r)e^{int} = \sum_{n=0}^{+\infty} d_n r^n e^{int} = \sum_{n=0}^{+\infty} d_n (re^{it})^n$$

et $\varphi = \dfrac{1}{f}$ est développable en série entière avec un rayon de convergence infini. D'où le résultat. $\lhd$

3.45. Un théorème de Fejér.

Soient $\mathrm{D} = \{z \in \mathbb{C},\ |z| < 1\}$, $f : \overline{\mathrm{D}} \longrightarrow \mathbb{C}$ continue. On suppose f développable en série entière sur D : si $z \in \mathrm{D}$, $f(z) = \displaystyle\sum_{n=0}^{+\infty} a_n z^n$. On suppose de plus f injective.

1. On pose $\mathrm{P} = \mathrm{Re}(f)$ et $\mathrm{Q} = \mathrm{Im}(f)$. On identifie $\mathbb{C}$ à $\mathbb{R}^2$ et D à un ouvert de $\mathbb{R}^2$. Montrer que les fonctions P et Q sont de classe $\mathcal{C}^1$ et vérifient les *conditions de Cauchy* : pour tout $(x, y) \in \mathrm{D}$,

$$\frac{\partial \mathrm{P}}{\partial x}(x,y) = \frac{\partial \mathrm{Q}}{\partial y}(x,y) \quad \text{et} \quad \frac{\partial \mathrm{P}}{\partial y}(x,y) = -\frac{\partial \mathrm{Q}}{\partial x}(x,y).$$

2. On note, pour tout $r \in [0,1[$, $\overline{\mathrm{D}}_r = \{z \in \mathbb{C},\ |z| \leqslant r\}$ et A_r l'aire de $f(\overline{\mathrm{D}}_r)$ et A l'aire de $f(\overline{\mathrm{D}})$. Montrer que, pour tout $r \in [0,1[$,

$$\mathrm{A}_r = \iint_{(x,y)\in\overline{\mathrm{D}}_r} |f'(x+iy)|^2 \mathrm{d}x\mathrm{d}y.$$

En déduire que $\sum n|a_n|^2$ converge et que $\sum_{n=0}^{+\infty} n|a_n|^2 \leqslant \dfrac{\mathrm{A}}{\pi}$.

3. Montrer que $\sum a_n z^n$ converge uniformément sur $\overline{\mathrm{D}}$.

(École normale supérieure)

▷ **Solution.**

1. Nous allons montrer que P et Q sont de classe $\mathcal{C}^1$ et exprimer leurs dérivées partielles en fonction de f'. Soit $(x_0, y_0) \in \mathbb{R}^2$, $z_0 = x_0 + y_0$. On a, pour $x \neq x_0$, en posant $z = x + iy_0$,

$$\begin{aligned}\frac{f(z)-f(z_0)}{z-z_0} &= \frac{\mathrm{P}(x,y_0)+i\mathrm{Q}(x,y_0)-\mathrm{P}(x_0,y_0)-i\mathrm{Q}(x_0,y_0)}{x-x_0}\\ &= \frac{\mathrm{P}(x,y_0)-\mathrm{P}(x_0,y_0)}{x-x_0} + i\frac{\mathrm{Q}(x,y_0)-\mathrm{Q}(x_0,y_0)}{x-x_0}.\end{aligned}$$

Sachant que $\dfrac{f(z)-f(z_0)}{z-z_0}$ a pour limite $f'(z_0)$ quand z tend vers z_0, on en déduit que $\dfrac{\mathrm{P}(x,y_0)-\mathrm{P}(x_0,y_0)}{x-x_0}$ et $\dfrac{\mathrm{Q}(x,y_0)-\mathrm{Q}(x_0,y_0)}{x-x_0}$ ont des limites quand x tend vers x_0. Autrement $\dfrac{\partial \mathrm{P}}{\partial x}(x_0,y_0)$ et $\dfrac{\partial \mathrm{Q}}{\partial x}(x_0,y_0)$ existent et vérifient, de plus

$$\frac{\partial \mathrm{P}}{\partial x}(x_0,y_0) + i\frac{\partial \mathrm{Q}}{\partial x}(x_0,y_0) = f'(z_0).$$

En écrivant de même, pour $y \neq y_0$ et $z = x_0 + iy$,

$$\begin{aligned}\frac{f(z)-f(z_0)}{z-z_0} &= \frac{\mathrm{P}(x_0,y)+i\mathrm{Q}(x_0,y)-\mathrm{P}(x_0,y_0)-i\mathrm{Q}(x_0,y_0)}{i(y-y_0)}\\ &= \frac{\mathrm{Q}(x_0,y)-\mathrm{Q}(x_0,y_0)}{y-y_0} - i\frac{\mathrm{P}(x_0,y)-\mathrm{P}(x_0,y_0)}{y-y_0},\end{aligned}$$

on prouve l'existence de dérivées partielles par rapport à y vérifiant

$$\frac{\partial Q}{\partial y}(x_0,y_0)-i\frac{\partial P}{\partial y}(x_0,y_0)=f'(z_0).$$

On en déduit les conditions de Cauchy

$$\frac{\partial P}{\partial x}(x_0,y_0)=\frac{\partial Q}{\partial y}(x_0,y_0)\quad\text{et}\quad\frac{\partial P}{\partial y}(x_0,y_0)=-\frac{\partial Q}{\partial x}(x_0,y_0).$$

On note de plus que, ces dérivées partielles sont continues, puisque f' est continue (elle est développable en série entière en 0 avec un rayon de convergence $\geqslant 1$ et donc en particulier continue sur D).

2. Pour tout $r\in[0,1[$, on note $\overline{\mathrm{D}}_r=\{z\in\mathbb{C},\ |z|\leqslant r\}$, $\mathrm{K}_r=f(\overline{\mathrm{D}}_r)$ et $\mathrm{K}=f(\overline{\mathrm{D}})$. Les ensembles K_r et K sont compacts en tant qu'images de compact par une fonction continue, ce qui justifie l'existence de A et A_r. Nous allons calculer $\mathrm{A}_r=\iint_{(u,v)\in\mathrm{K}_r}\mathrm{d}u\,\mathrm{d}v$ par changement de variables.

La fonction f est continue et injective sur $\overline{\mathrm{D}}$ et établit donc une bijection de $\overline{\mathrm{D}}$ sur K. Cette bijection est même un homéomorphisme, puisque $\overline{\mathrm{D}}$ est compact (tout fermé F de $\overline{\mathrm{D}}$ est compact, donc $f(\mathrm{F})$ est compact et en particulier fermé dans K : l'image réciproque d'un fermé par f^{-1} étant un fermé, f^{-1} est continue). Elle est de classe $\mathcal{C}^1$ sur D et son jacobien en (x,y) est

$$\mathrm{Jac}(f)(x,y)=\begin{vmatrix}\dfrac{\partial P}{\partial x}&\dfrac{\partial P}{\partial y}\\[2mm]\dfrac{\partial Q}{\partial x}&\dfrac{\partial Q}{\partial y}\end{vmatrix}(x,y)=\left(\frac{\partial P}{\partial x}\right)^2+\left(\frac{\partial Q}{\partial x}\right)^2=|f'(x+iy)|^2,$$

en vertu des conditions de Cauchy.

Le changement de variables $(x,y)\longmapsto(u,v)=f(x,y)$ dans l'intégrale double nous donne donc, pour $r\in[0,1[$,

$$\mathrm{A}_r=\iint_{(x,y)\in\overline{\mathrm{D}}_r}|f'(x+iy)|^2\mathrm{d}x\mathrm{d}y.$$

On va calculer cette intégrale en passant en coordonnées polaires et en appliquant le théorème de Fubini. On obtient

$$\mathrm{A}_r=\iint_{\substack{0\leqslant s\leqslant r\\0\leqslant\theta\leqslant 2\pi}}|f'(se^{i\theta})|^2 s\,\mathrm{d}s\,\mathrm{d}\theta=\int_0^r\left(\int_0^{2\pi}|f'(se^{i\theta})|^2\mathrm{d}\theta\right)s\,\mathrm{d}s.$$

L'intégrale sur θ va s'exprimer en fonction des coefficients a_n à l'aide de la formule de Parseval (voir page 238). Pour $z\in\mathrm{D}$, on a

$$f'(z)=\sum_{n=1}^{+\infty}na_nz^n=\sum_{n=0}^{+\infty}(n+1)a_nz^n.$$

Prenons $s\in[0,r]$ et considérons la fonction $\theta\longmapsto g_s(\theta)=f'(se^{i\theta})$. Elle est continue et 2π-périodique et vérifie, pour tout $\in[0,2\pi]$,

$$g_s(\theta) = \sum_{n=0}^{+\infty} (n+1)a_{n+1}s^n e^{in\theta}.$$

Les coefficients de Fourier $c_n(g)$ de g sont donnés pour $n \geqslant 0$ par $c_n(g) = (n+1)a_{n+1}s^n$ puisque la convergence de la série est normale. Le théorème de Parseval donne donc

$$\frac{1}{2\pi}\int_0^{2\pi} |f'(se^{i\theta})|^2 \mathrm{d}\theta = \sum_{n=0}^{+\infty} (n+1)^2 |a_{n+1}|^2 s^{2n} = \sum_{n=1}^{+\infty} n^2 |a_n|^2 s^{2n-2}.$$

On obtient donc

$$\mathrm{A}_r = \int_0^r 2\pi \left(\sum_{n=1}^{+\infty} n^2 |a_n|^2 s^{2n-2} \right) s\mathrm{d}s = 2\pi \int_0^r \sum_{n=1}^{+\infty} \underbrace{n^2 |a_n|^2 s^{2n-1}}_{\geqslant 0} \mathrm{d}s.$$

D'après le théorème de convergence monotone, on peut intervertir sommation et intégration puisque les fonctions sont positives, ce qui donne

$$\mathrm{A}_r = 2\pi \sum_{n=1}^{+\infty} |a_n|^2 n^2 \int_0^r s^{2n-1}\mathrm{d}s = 2\pi \sum_{n=1}^{+\infty} |a_n|^2 n^2 \frac{r^{2n}}{2n} = \pi \sum_{n=1}^{+\infty} n|a_n|^2 r^{2n}.$$

Comme $\overline{\mathrm{D}}_r \subset \overline{\mathrm{D}}$, on a $\mathrm{K}_r \subset \mathrm{K}$ et donc $\mathrm{A}_r \leqslant \mathrm{A}$. Ainsi pour tout $r \in [0,1[$,

$$\sum_{n=1}^{+\infty} n|a_n|^2 r^{2n} \leqslant \frac{\mathrm{A}}{\pi} \quad \text{et en particulier} \quad \sum_{n=1}^{\mathrm{N}} n|a_n|^2 r^{2n} \leqslant \frac{\mathrm{A}}{\pi}$$

pour tout $\mathrm{N} \in \mathbb{N}^*$. En faisant tendre r vers 1, on obtient $\sum_{n=1}^{\mathrm{N}} n|a_n|^2 \leqslant \frac{\mathrm{A}}{\pi}$. La série à termes positifs $\sum n|a_n|^2$ converge donc et sa somme est majorée par $\frac{\mathrm{A}}{\pi}$.

En fait, on a l'égalité $\frac{\mathrm{A}}{\pi} = \sum_{n=1}^{+\infty} n|a_n|^2$. *En effet, on peut écrire, pour tout* $r \in]0,1[$, $\sum_{n=1}^{+\infty} n|a_n|^2 \geqslant \sum_{n=1}^{+\infty} n|a_n|^2 r^{2n} \geqslant \frac{\mathrm{A}_r}{\pi}$. *On remarque que*

$$\mathrm{A}_r = \int_0^r \left(\int_0^{2\pi} f(se^{i\theta})d\theta \right) ds \xrightarrow[r\to 1]{} \int_0^1 \left(\int_0^{2\pi} f(se^{i\theta})d\theta \right) ds = \mathrm{A}.$$

En faisant tendre r *vers* 1, *on obtient* $\sum_{n=1}^{+\infty} n|a_n|^2 \geqslant \frac{\mathrm{A}}{\pi}$ *et donc l'égalité annoncée.*

3. En vertu d'un théorème d'Abel dont on trouvera la démonstration dans l'exercice 3.6, il suffit de prouver la convergence uniforme de la série entière sur $\mathrm{U} = \{z \in \mathbb{C}, |z| = 1\}$ *i.e.* de prouver que la série $\sum_{n=0}^{+\infty} a_n e^{in\theta}$ converge uniformément pour $\theta \in \mathbb{R}$. Fixons $\varepsilon > 0$.

Soient $N \in \mathbb{N}$, $\theta \in \mathbb{R}$ et $r \in [0,1[$. Posons $R_N = \sum_{n=N+1}^{+\infty} n|a_n|^2$. On a

$$\left|f(e^{i\theta}) - \sum_{n=0}^{N} a_n e^{in\theta}\right| \leqslant |f(e^{i\theta}) - f(re^{i\theta})| + \left|\sum_{n=0}^{N} a_n e^{in\theta}(1-r^n)\right| + \left|\sum_{n=N+1}^{+\infty} a_n r^n e^{in\theta}\right|.$$

On majore chaque terme.

- D'après l'inégalité de Cauchy-Schwarz, on a

$$\begin{aligned}\left|\sum_{n=N+1}^{+\infty} a_n r^n e^{in\theta}\right| &\leqslant \sum_{n=N+1}^{+\infty} |a_n| r^n = \sum_{n=N+1}^{+\infty} |a_n|\sqrt{n}\frac{r^n}{\sqrt{n}} \\ &\leqslant \sqrt{\left(\sum_{n=N+1}^{+\infty} n|a_n|^2\right)\left(\sum_{n=N+1}^{+\infty} \frac{r^{2n}}{n}\right)} \\ &\leqslant \sqrt{\frac{R_N}{N+1}\sum_{n=0}^{+\infty} r^n} \leqslant \sqrt{\frac{R_N}{(N+1)(1-r)}}.\end{aligned}$$

- Pour le second terme, on remarque que

$$\begin{aligned}\left|\sum_{n=0}^{N} a_n e^{in\theta}(1-r^n)\right| &\leqslant \sum_{n=0}^{N} |a_n|(1-r)(1+r+r^2+\cdots+r^{n-1}) \\ &\leqslant (1-r)\sum_{n=0}^{N} n|a_n|.\end{aligned}$$

L'inégalité de Cauchy-Schwarz permet d'écrire

$$\sum_{n=0}^{N} n|a_n| \leqslant \sqrt{\left(\sum_{n=0}^{N} n|a_n|^2\right)\left(\sum_{n=0}^{N} n\right)} \leqslant \sqrt{\frac{A}{\pi}\cdot\frac{N(N+1)}{2}} \leqslant K(N+1).$$

où $K = \sqrt{\dfrac{A}{2\pi}}$. Il en résulte $\left|\sum_{n=0}^{N} a_n e^{in\theta}(1-r^n)\right| \leqslant K(N+1)(1-r)$.

- La fonction f étant continue sur le compact $\overline{D}$, y est uniformément continue d'après le théorème de Heine. Soit donc $\varepsilon > 0$ et η un module d'uniforme continuité de f pour ε. Si $|1-r| \leqslant \eta$, on a $|e^{i\theta} - re^{i\theta}| \leqslant \eta$ et $|f(e^{i\theta}) - f(re^{i\theta})| \leqslant \varepsilon$.

- Posons $r_N = 1 - \dfrac{\varepsilon}{K(N+1)}$ (et donc $K(N+1)(1-r_N) = \varepsilon$). Pour N assez grand, $r_N \in [0,1[$ et même à partir d'un certain rang n_0, $1-\eta \leqslant r_N \leqslant 1$. On a donc pour $N \geqslant n_0$ et θ dans $\mathbb{R}$.

$$\left|f(e^{i\theta}) - \sum_{n=0}^{N} a_n e^{in\theta}\right| \leqslant \varepsilon + \sqrt{\frac{R_N}{(N+1)(1-r_N)}} + \varepsilon \leqslant \varepsilon + \sqrt{\frac{R_N K}{\varepsilon}} + \varepsilon \leqslant 3\varepsilon$$

pour N assez grand puisque $R_N \xrightarrow[N\to+\infty]{} 0$.

La convergence est donc uniforme sur U et par conséquent sur $\overline{D}$. ◁

La démonstration faite dans cette dernière question reprend exactement la démarche utilisée dans la question 3 de l'exercice 3.26. On va donner maintenant un autre éclairage de ce résultat à l'aide du théorème de Fejér sur les séries de Fourier que le lecteur trouvera à l'exercice 4.31. Posons $g(\theta) = f(e^{i\theta})$. La fonction g est continue, 2π-périodique, et comme cela est fait dans la question 1 de l'exercice 3.43, on montre que ses coefficients de Fourier $c_n(g)$ sont donnés par $c_n(g) = a_n$ si $n \geqslant 0$ et $c_n(g) = 0$ sinon. Le problème est donc de démontrer que la série de Fourier de g converge uniformément vers g. Notons $S_n(\theta) = \sum_{k=0}^{n} a_k e^{ik\theta}$ la somme partielle de cette série de Fourier. Le théorème de Fejér prouvé dans l'exercice 4.31 affirme que la moyenne de Cesàro des sommes partielles, à savoir $F_n = \dfrac{S_0 + S_1 + \cdots + S_n}{n}$ possède cette propriété : elle converge uniformément vers g sur $\mathbb{R}$. Un petit calcul montre que

$$F_n(\theta) = \sum_{k=0}^{n} \left(1 - \frac{k}{n}\right) a_k e^{ik\theta}.$$

On écrit $F_n(\theta) = S_n(\theta) - \dfrac{1}{n}\sum_{k=0}^{n} k a_k e^{ik\theta}$. Il suffit de voir que le second terme converge uniformément vers 0 pour conclure. Or,

$$\frac{1}{n}\left|\sum_{k=0}^{n} k a_k e^{ik\theta}\right| \leqslant \frac{|a_1| + 2|a_2| + \cdots + n|a_n|}{n}$$

et cette dernière suite tend vers 0 car la série $\sum n|a_n|^2$ converge (comme cela est démontré dans la remarque qui suit l'exercice 3.26).

Le dernier exercice de ce chapitre recherche une fonction développable en série entière sur $\mathbb{C}$ et prenant des valeurs données pour $z = 2^n$ ($n \in \mathbb{N}$). L'étude de l'unicité est l'occasion de se poser la question de l'existence d'une fonction s'annulant en des points donnés. Signalons que si (z_n) est une suite de nombres complexes non nuls telle que $(|z_n|)$ tende vers l'infini, il existe une fonction développable en série entière sur $\mathbb{C}$, s'annulant en z_n pour tout n et n'ayant pas d'autres zéros sur $\mathbb{C}$.

3.46. Fonction entière telle que $f(2^n) = (-1)^n$

1. Démontrer qu'il existe une suite $(a_n)_{n\in\mathbb{N}}$ de nombres réels telle que, pour tout $n \in \mathbb{N}$,

$$\sum_{k=0}^{\infty} 2^{nk} a_k = (-1)^n.$$

2. Y a-t-il unicité de la suite (a_n) ?

(École normale supérieure, école polytechnique)

▷ **Solution.**

1. Il s'agit de déterminer une suite réelle (a_n) telle que, si on pose $f(z) = \sum_{k=0}^{+\infty} a_k z^k$, on ait, pour tout $n \in \mathbb{N}$, $f(2^n) = (-1)^n$. Ceci implique entre autres que la série $\sum a_k z^k$ ait un rayon de convergence infini. On peut voir le calcul des a_k comme la résolution d'un système de taille infinie d'équations linéaires :

$$\begin{cases} a_0 + a_1 + a_2 + \cdots + a_n + \cdots & = & 1 \\ a_0 + 2a_1 + 2^2 a_2 + \cdots + 2^n a_n + \cdots & = & -1 \\ & \vdots & \\ a_0 + 2^n a_1 + 2^{2n} a_2 + \cdots + 2^{n^2} a_n + \cdots & = & (-1)^n \\ & \vdots & \end{cases}$$

D'où l'idée de se ramener à un système fini, en cherchant d'abord, pour $n \in \mathbb{N}^*$, à résoudre le système de $n+1$ équations à $n+1$ inconnues $a_{0,n}, a_{1,n}, \ldots, a_{n,n}$:

$$\begin{cases} a_{0_n} + a_{1,n} + a_{2,n} + \cdots + a_{n,n} & = & 1 \\ a_{0_n} + 2a_{1,n} + 2^2 a_{2,n} + \cdots + 2^n a_{n,n} & = & -1 \\ & \vdots & \\ a_{0_n} + 2^n a_{1,n} + 2^{2n} a_{2,n} + \cdots + 2^{n^2} a_{n,n} & = & (-1)^n \end{cases}$$

On remarque que le déterminant de ce système est le déterminant de Vandermonde $\mathrm{V}(1, 2, \ldots, 2^n)$. Les entiers $1, 2, \ldots, 2^n$ étant distincts, on en déduit que ce déterminant n'est pas nul. Le système possède donc une solution unique. Si on remplace une des colonnes de ce déterminant par le deuxième membre de l'équation, on obtient encore un déterminant de Vandermonde. En appliquant les formules de Cramer, on trouve donc après simplification, pour $0 \leqslant k \leqslant n$,

$$a_{k,n} = \frac{V(1,\dots,2^{k-1},-1,2^{k+1},\dots,2^n)}{V(1,\dots,2^{k-1},2^k,2^{k+1},\dots,2^n)} = \frac{\prod_{j\neq k}(2^j+1)}{\prod_{j\neq k}(2^j-2^k)}$$

$$= \frac{\prod_{j=0}^{k-1}(1+2^j)\prod_{j=k+1}^{n}(1+2^j)}{\prod_{j=0}^{k-1}(2^j-2^k)\prod_{j=k+1}^{n}(2^j-2^k)}.$$

En mettant 2^j en facteur dans chaque terme au numérateur et au dénominateur, on obtient

$$a_{k,n} = \frac{\prod_{j=0}^{k-1}(1+2^{-j})\prod_{j=k+1}^{n}(1+2^{-j})}{\prod_{j=0}^{k-1}(1-2^{k-j})\prod_{j=k+1}^{n}(1-2^{k-j})} = \frac{\prod_{j=0}^{k-1}(1+2^{-j})\prod_{j=k+1}^{n}(1+2^{-j})}{\prod_{j=1}^{k}(1-2^j)\prod_{j=1}^{n-k}(1-2^{-j})}$$

$$= \frac{\prod_{j=0}^{n}(1+2^{-j})}{(1+2^{-k})\prod_{j=1}^{k}(1-2^j)\prod_{j=1}^{n-k}(1-2^{-j})}.$$

On fait maintenant tendre n vers l'infini. Les suites de terme général $u_n = \prod_{j=0}^{n}(1+2^{-j})$ et $v_n = \prod_{j=1}^{n}(1-2^{-j})$ convergent vers des limites non nulles. En effet, on a, pour $n \geqslant 1$, $\ln u_n = \sum_{j=0}^{n}\ln(1+2^{-j})$ et $\ln v_n = \sum_{j=1}^{n}\ln(1-2^{-j})$ et les séries $\sum\ln(1+2^{-j})$ et $\sum\ln(1-2^{-j})$ convergent.

On note $\prod_{j=0}^{+\infty}(1+2^{-j})$ et $\prod_{j=1}^{+\infty}(1-2^{-j})$ les limites de ces suites et on pose $K = \dfrac{\prod_{j=0}^{+\infty}(1+2^{-j})}{\prod_{j=1}^{+\infty}(1-2^{-j})}$. On trouve alors que, pour tout $k\in\mathbb{N}$, la suite $(a_{k,n})_{n\in\mathbb{N}}$ converge vers $a_k = \dfrac{K}{(1+2^{-k})\prod_{j=1}^{k}(1-2^j)}$.

Il est tentant de penser que la suite (a_k) répond au problème posé. Remarquons, pour commencer, que pour tout $k\in\mathbb{N}$, on a

$$\left|\frac{a_{k+1}}{a_k}\right| = \frac{(1+2^{-k})}{(1+2^{-k-1})(2^{k+1}-1)} \underset{k\to\infty}{\sim} \frac{1}{2^{k+1}} \xrightarrow[k\to+\infty]{} 0.$$

On en déduit que la série entière $\sum a_k z^k$ a un rayon de convergence infini. Il faut maintenant démontrer que, pour tout $p \in \mathbb{N}$, on a $f(2^p) = (-1)^p$. On a, par construction, pour $n \geqslant p$, $\sum_{k=0}^{n} a_{k,n} 2^{pk} = (-1)^p$. Il s'agit d'un problème d'interversion d'une limite et d'une somme. On remarque que, k étant fixé, la suite $(|a_{k,n}|)_{n \in \mathbb{N}^*}$ croît car, pour passer d'un terme au suivant, on multiplie par $\dfrac{1+2^{-(n+1)}}{1-2^{-(n-k+1)}} > 1$. On a donc, pour $n \geqslant k$, $|a_{k,n}| \leqslant |a_k|$, c'est-à-dire une majoration indépendante de n, qui va conduire au résultat.

Fixons donc p. On a alors, pour $n \geqslant p$,

$$\begin{aligned} |f(2^p) - (-1)^p| &= \left| \sum_{k=0}^{+\infty} a_k 2^{pk} - \sum_{k=0}^{n} a_{k,n} 2^{pk} \right| \\ &\leqslant \sum_{k=0}^{n} |a_k - a_{k,n}| 2^{kp} + \sum_{k=n+1}^{+\infty} |a_k| 2^{pk}. \end{aligned}$$

La série $\sum |a_k| 2^{pk}$ est convergente. Pour tout $\varepsilon > 0$, il existe $n_0 \in \mathbb{N}^*$ tel que $\sum_{k=n_0+1}^{+\infty} |a_k| 2^{pk} \leqslant \varepsilon$. On obtient alors, pour $n \geqslant \max(n_0, p)$,

$$\begin{aligned} |f(2^p) - (-1)^p| &\leqslant \sum_{k=0}^{n_0} |a_k - a_{k,n}| 2^{kp} + 2 \sum_{k=n_0+1}^{n} |a_k| 2^{pk} + \sum_{n+1}^{+\infty} |a_k| 2^{pk} \\ &\leqslant \sum_{k=0}^{n_0} |a_k - a_{k,n}| 2^{kp} + 2 \sum_{k=n_0+1}^{+\infty} |a_k| 2^{pk} \\ &\leqslant \sum_{k=0}^{n_0} |a_k - a_{k,n}| 2^{kp} + 2\varepsilon. \end{aligned}$$

En faisant tendre n vers $+\infty$, on obtient $|f(2^p) - (-1)^p| \leqslant 2\varepsilon$. Ceci étant vrai pour tout $\varepsilon > 0$, on a donc $f(2^p) = (-1)^p$.

2. Si (b_n) est une suite telle $g(z) = \sum_{n=0}^{+\infty} b_n z^n$ soit définie sur $\mathbb{C}$, (b_n) est solution du problème posé si f et g prennent les mêmes valeurs en 2^p, pour tout $p \in \mathbb{N}$, c'est-à-dire si $f - g$ s'annule pour toutes ces valeurs. On est donc ramené à chercher s'il existe des fonctions développables en séries entières, non nulles, s'annulant en 2^p pour tout p. On pense assez naturellement à la fonction définie par $h(z) = \prod_{n=0}^{+\infty} \left(1 - \dfrac{z}{2^n}\right)$, qui visiblement s'annule où il convient. Il faudrait démontrer que cette fonction est définie sur $\mathbb{C}$ et développable en série entière de rayon de convergence infini. On peut remarquer que, pour tout $z \in \mathbb{C}$, on

$h(z) = (1-z)\prod_{n=1}^{+\infty}\left(1-\frac{z}{2^n}\right) = (1-z)h\left(\frac{z}{2}\right)$ et chercher une fonction développable en série entière vérifiant cette propriété. On la cherche sous la forme $\varphi(z) = \sum_{n=0}^{+\infty} c_n z^n$. On a alors, sous réserve de convergence, pour $z \in \mathbb{C}$,

$$\begin{aligned}\varphi(z) - (1-z)\varphi\left(\frac{z}{2}\right) &= \sum_{n=0}^{+\infty} c_n z^n - \sum_{n=0}^{+\infty} c_n \frac{z^n}{2^n} + \sum_{n=0}^{+\infty} c_n \frac{z^{n+1}}{2^n} \\ &= \sum_{n=1}^{+\infty} c_n \left(1-\frac{1}{2^n}\right) z^n + \sum_{n=1}^{+\infty} \frac{c_{n-1}}{2^{n-1}} z^n.\end{aligned}$$

D'après l'unicité du développement en série entière, la condition $\varphi(z) = (1-z)\varphi\left(\frac{z}{2}\right) = 0$ s'écrit : pour tout $n \in \mathbb{N}^*$, $c_n = \frac{2}{1-2^n} c_{n-1}$. Si on se donne $c_0 \neq 0$, cette relation définit c_n pour $n \geqslant 1$. On remarque qu'on a alors $\lim_{n\to+\infty} \frac{c_n}{c_{n-1}} = \lim_{n\to+\infty} \frac{2}{1-2^n} = 0$. Le rayon de convergence de la série entière $\sum c_n z^n$ est donc infini. La fonction φ est définie sur $\mathbb{C}$ et par construction, vérifie $\varphi(z) = (1-z)\varphi\left(\frac{z}{2}\right)$. Une récurrence immédiate montre que, pour $z \in \mathbb{C}$ et $n \in \mathbb{N}$, on a

$$\varphi(z) = \prod_{k=0}^{n}\left(1-\frac{z}{2^k}\right)\cdot\varphi\left(\frac{z}{2^{n+1}}\right).$$

Quand n tend vers $+\infty$, $\varphi\left(\frac{z}{2^{n+1}}\right)$ tend vers $\varphi(0) = c_0 \neq 0$. On en déduit que le produit

$$\prod_{k=0}^{n}\left(1-\frac{z}{2^k}\right)$$

converge vers $\frac{\varphi(z)}{c_0}$. On a donc, pour tout $z \in \mathbb{C}$,

$$\varphi(z) = c_0 \prod_{n=0}^{+\infty}\left(1-\frac{z}{2^n}\right) = c_0 h(z).$$

Pour $c_0 = 1$, on obtient $\varphi = h$. La fonction h est donc développable en série entière et vérifie $h(2^p) = 0$ pour tout $p \in \mathbb{N}$. Elle n'est pas nulle ($h(0) = 1$). Elle possède toutes les propriétés voulues. On peut conclure que la suite (a_n) cherchée n'est pas unique. ◁

On peut noter que toute fonction de la forme $h\psi$, où ψ est développable en série entière, avec un rayon de convergence infini, a les mêmes propriétés que h.

Chapitre 4

Séries de Fourier

C'est autour de l'équation des cordes vibrantes que naît l'idée de la représentation des fonctions périodiques sous forme de séries trigonométriques : si t représente le temps, y l'oscillation de la corde à l'abscisse x, l'équation est de la forme

$$\frac{\partial^2 y}{\partial t^2} = \alpha^2 \frac{\partial^2 y}{\partial x^2},$$

où α est une constante, homogène à une vitesse. C'est d'Alembert (1749) qui introduit cette équation, et qui montre que la solution générale est de la forme

$$f(x + \alpha t) + g(x - \alpha t).$$

Reprenant les travaux de d'Alembert. Euler est conduit pour la première fois à considérer des fonctions définies par des expressions différentes sur des intervalles différents, qu'il appelle « discontinues », notamment pour représenter la forme triangulaire d'une corde pincée.

Une intuition géniale revient à Daniel Bernoulli (1755) qui observe qu'une corde peut donner à entendre non seulement le son fondamental et les harmoniques de celui-ci, de fréquence double, triple, etc., mais aussi la superposition de ces sons. Il en tire l'idée que toutes les solutions pour lesquelles on a $y = 0$ aux points d'attache de la corde, $x = 0$ et $x = \mathrm{L}$, et qui correspondent à une corde au repos à $t = 0$, doivent s'écrire comme somme d'une série trigonométrique i.e. une combinaison linéaire infinie des fonctions

$$\sin n\omega x \cos n\omega \alpha t,$$

avec $\omega = \dfrac{\pi}{\mathrm{L}}$.

Cela implique qu'une fonction arbitraire pour $0 < x < \mathrm{L}$ et nulle pour $x = 0$ et $x = \mathrm{L}$ doit pouvoir s'écrire comme combinaison linéaire infinie des fonctions $\sin n\omega x$. Euler, en particulier, n'est pas convaincu. Et comme Bernoulli ne fait aucun effort pour prouver ce qu'il avance, ni pour calculer les coefficients, les objections finissent par l'emporter et par décourager toute étude sérieuse de la question. Ces objections, à propos de la périodicité, de la parité, ont en gros le même fond : on n'imagine pas à cette époque qu'une expression (fût-elle infinie) puisse être égale à une autre expression sur un intervalle sans lui être égale pour toutes les valeurs de la variable.

À partir de 1805, Fourier s'implique dans l'étude de l'équation de la chaleur selon une méthode déjà éprouvée pour l'équation des cordes vibrantes (Lagrange) ou l'équation de la « chaînette » (Huygens) : on écrit les équations pour un système composé d'un nombre fini de corps et on obtient l'équation pour le système continu en faisant tendre la dimension de ces corps vers 0 et leur nombre vers l'infini. Fourier résout le problème discret, mais n'effectue pas le passage à la limite. En faisant $t = 0$ dans la solution obtenue pour l'équation de la chaleur, il aurait obtenu l'écriture d'une fonction arbitraire (de période 2π) sous la forme d'une série trigonométrique sous la forme

$$\begin{aligned} f(x) = {} & \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx \\ & + \left(\frac{1}{\pi}\int_0^{2\pi} f(x)\cos x\,dx\right)\cos x + \left(\frac{1}{\pi}\int_0^{2\pi} f(x)\cos 2x\,dx\right)\cos 2x + \cdots \\ & + \left(\frac{1}{\pi}\int_0^{2\pi} f(x)\sin x\,dx\right)\sin x + \left(\frac{1}{\pi}\int_0^{2\pi} f(x)\sin 2x\,dx\right)\sin 2x + \cdots \end{aligned}$$

Il imagine pour prouver l'existence du développement trigonométrique d'une fonction arbitraire une toute autre méthode. Voici, par exemple, comment il procède dans le cas particulier de la fonction $f(x)$ qui vaut 1 sur $]-\pi/2, \pi/2[$, -1 sur $]\pi/2, 3\pi/2[$, où on cherche un développement de la forme

$$f(x) = a_1 \cos x + a_3 \cos 3x + a_5 \cos 5x + \cdots$$

(les fonctions $\cos x$, $\cos 3x$, $\cos 5x$, ... ont les mêmes symétries que la fonction donnée). En écrivant que ce développement, ainsi que les développements obtenus par dérivations successives, sont valables pour $x = 0$, il obtient le système d'équations linéaires infini

$$\begin{aligned} 1 &= a_1 + a_3 + a_5 + a_7 + \cdots \\ 0 &= a_1 + 3^2 a_3 + 5^2 a_5 + 7^2 a_7 + \cdots \\ 0 &= a_1 + 3^4 a_3 + 5^4 a_5 + 7^4 a_7 + \cdots \\ 0 &= a_1 + 3^6 a_3 + 5^6 a_5 + 7^6 a_7 + \cdots \\ &\cdots \end{aligned}$$

Fourier résout d'abord ce système par élimination en ne conservant que les m premières équations, et dans chaque équation que les m premières inconnues, puis passe à la limite $m \to \infty$ en utilisant la formule de Wallis

$$\frac{\pi}{2} = \frac{2.2}{1.3}\cdot\frac{4.4}{3.5}\cdot\frac{6.6}{5.7}\cdots = 2\,\frac{3^2-1}{3^2}\cdot\frac{5^2-1}{5^2}\cdot\frac{7^2-1}{7^2}\cdots$$

Il obtient ainsi, après contrôle de la convergence,

$$f(x) = \frac{4}{\pi}\left(\cos x - \frac{1}{3}\cos 3x + \frac{1}{5}\cos 5x - \cdots\right)$$

Ce n'est que plus tard qu'il redécouvre la méthode « expéditive » pour calculer les coefficients de l'écriture d'une fonction périodique comme série trigonométrique, à supposer que cette écriture soit possible : multiplier par $\cos nx$ *ou* $\sin nx$ *et intégrer la série terme à terme. Mais dans sa* Théorie analytique de la chaleur, *achevée en 1816 et publiée en 1822, où il conserve la méthode des systèmes linéaires infinis, en même temps qu'il met au point la méthode de discrétisation, il précise : savoir calculer les coefficients* a_n *et* b_n *d'un développement hypothétique* $f(x) = \sum a_n \cos nx + \sum b_n \sin nx$ *ne prouve en aucune façon l'existence de ce développement, ou, si on préfère, la convergence de la série du second membre vers la fonction* f.

Entre temps, Deflers (1819) et Poisson (1820) auront donné des preuves de la convergence de la « série de Fourier » vers la fonction, valables sous des hypothèses non explicitées (mais $\mathcal{C}^2$ *suffit). Dirichlet, dans un travail dont on dira plus tard qu'il établit un nouveau standard de rigueur en analyse, donnera en 1829 des conditions de convergence adaptées aux fonctions continues par morceaux qui étaient devenues depuis Euler un objet de préoccupation pour les mathématiciens.*

Voici quelques rappels du cours qui permettent de fixer les notations. Pour une fonction f, 2π*-périodique, continue par morceaux, les coefficients de Fourier trigonométriques de* f *sont définis par*

$$a_n(f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt \quad et \quad b_n(f) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt,$$

n *étant dans* $\mathbb{N}$ *et la série de Fourier associée est*

$$\frac{a_0}{2} + \sum_{n=1}^{+\infty}(a_n\cos nt + b_n \sin nt).$$

Les coefficients de Fourier exponentiels sont définis par

$$c_n(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt \text{ pour } n \in \mathbb{Z}$$

et la série de Fourier correspondante est $\sum_{n=-\infty}^{+\infty} c_n(f)e^{int}$.

Nous commençons par une série d'exercices sur les polynômes trigonométriques. Un polynôme trigonométrique est une combinaison linéaire à coefficients complexes de fonctions de la forme $t \longmapsto e^{int}$, *avec* $n \in \mathbb{Z}$. *Si* f *est définie par* $f(t) = \sum_{k=-n}^{n} c_k e^{ikt}$, *avec* $(c_{-n}, c_n) \neq (0,0)$, *on dit que* f *est de degré* n. *On peut encore écrire* f *sous la forme*

$f(t) = \sum_{k=0}^{n} a_k \cos kt + b_k \sin kt$. *La fonction f est à valeur réelles si et seulement si les coefficients a_k et b_k sont réels.*

Les deux premiers exercices utilisent simplement les relations d'orthogonalité entre les fonctions $e_n : t \longmapsto e^{int}$:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{int}e^{-imt}\,dt = \begin{cases} 1 \text{ si } n = m \\ 0 \text{ si } n \neq m. \end{cases}$$

Elles sont qualifiées ainsi car, si E désigne l'espace vectoriel complexe préhilbertien des fonctions 2π-périodiques continues de $\mathbb{R}$ dans $\mathbb{C}$ muni du produit hermitien défini par

$$(f|g) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \overline{f(t)}g(t)\,dt \quad \text{pour } f \text{ et } g \text{ dans } \mathrm{E},$$

la famille $(e_n)_{n\in\mathbb{Z}}$ est une famille orthonormale. On gardera également à l'esprit que $c_n(f) = (e_n|f)$.

4.1. Calcul d'intégrales

Soit $(m, n) \in \mathbb{N}^2$. Justifier l'existence de l'intégrale

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\sin(2m+1)\frac{t}{2}}{\sin\frac{t}{2}} \frac{\sin(2n+1)\frac{t}{2}}{\sin\frac{t}{2}}\,\mathrm{d}t$$

et la calculer.

(École polytechnique)

▷ **Solution.**

Pour tout entier naturel n la fonction $f_n : t \longmapsto \dfrac{\sin(2n+1)\frac{t}{2}}{\sin\frac{t}{2}}$ prolongée par $f_n(0) = (2n+1)$ est continue sur $[-\pi, \pi]$, ce qui justifie l'existence de $\mathrm{I}_{m,n} = \dfrac{1}{2\pi}\displaystyle\int_{-\pi}^{\pi} f_m(t)f_n(t)\,\mathrm{d}t$.

Transformons l'expression de f_n à l'aide des formules d'Euler. On obtient, pour $t \neq 0$ dans $[-\pi, \pi]$,

$$f_n(t) = \frac{e^{i(n+\frac{1}{2})t} - e^{-i(n+\frac{1}{2})t}}{e^{i\frac{t}{2}} - e^{-i\frac{t}{2}}} = \frac{e^{-i(n+\frac{1}{2})t}}{e^{-i\frac{t}{2}}} \cdot \frac{e^{i(2n+1)t} - 1}{e^{it} - 1}$$

$$= e^{-int}\sum_{k=0}^{2n} e^{ikt} = \sum_{k=-n}^{n} e^{ikt},$$

résultat qui reste encore vrai si $t = 0$. En transformant de la même façon f_m, on obtient

$$\mathrm{I}_{m,n} = \sum_{|k|\leqslant m, |\ell|\leqslant n} \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ikt} e^{i\ell t}\, \mathrm{d}t.$$

Compte tenu des relations d'orthogonalité entre les fonctions $t \longmapsto e^{int}$, les seuls termes non nuls sont obtenus pour $k + \ell = 0$ et valent 1. Le choix de k fixe la valeur de ℓ et on peut prendre pour k tous les entiers entre $-\min(n, m)$ et $\min(n, m)$. D'où il résulte que

$$\boxed{\mathrm{I}_{m,n} = 2\min(m, n) + 1}. \quad \lhd$$

L'espace $\mathcal{T}_n$ des polynômes trigonométriques de degré inférieur ou égal à n est stable par dérivation. Comme il s'agit d'un espace de dimension finie, l'opérateur linéaire de dérivation $\mathrm{D} : \mathrm{P} \longmapsto \mathrm{P}'$ est continu et on peut s'intéresser au calcul de sa norme triple lorsque qu'on munit $\mathcal{T}_n$ de la norme de la convergence uniforme. Cela consiste donc à déterminer la plus petite constante $c > 0$ telle que $\|\mathrm{P}'\|_\infty \leqslant c\|\mathrm{P}\|_\infty$ pour tout $\mathrm{P} \in \mathcal{T}_n$.

Le premier énoncé ci-après montre que la constante $c = 2n$ convient en utilisant le noyau de Fejér. L'énoncé suivant démontrera l'inégalité de Bernstein qui montre que $c = n$ convient, cette constante étant optimale.

4.2. Majoration de la norme de la dérivation

Soit $\mathrm{P}(x) = \sum_{k=-n}^{n} a_k e^{ikx}$ un polynôme trigonométrique de degré inférieur ou égal à n, où les a_k sont des nombres complexes. Trouver une constante c ne dépendant que de n telle que $\|\mathrm{P}'\|_\infty \leqslant c\|\mathrm{P}\|_\infty$.

On pourra poser $\mathrm{F}_n(x) = \sum_{k=-n}^{n} \left(1 - \frac{|k|}{n}\right) e^{ikx}$ et considérer l'intégrale $\mathrm{I}_n(x) = \int_0^{2\pi} \mathrm{P}(x - y)\mathrm{F}_n(y) \sin(ny)\mathrm{d}y$.

(École normale supérieure)

▷ **Solution.**

Suivons les indications de l'énoncé et calculons $\mathrm{I}_n(x)$. En écrivant

$$\mathrm{P}(x-y)=\sum_{\ell=-n}^{n} a_\ell e^{i\ell x}e^{-i\ell y} \quad \text{et} \quad \sin ny=\frac{1}{2i}(e^{iny}-e^{-iny}),$$

l'intégrale $\mathrm{I}_n(x)$ vaut, par linéarité,

$$\frac{1}{2i}\sum_{\substack{-n\leqslant k\leqslant n\\ -n\leqslant \ell\leqslant n}} a_\ell e^{i\ell x}\left(1-\frac{|k|}{n}\right)\left(\int_0^{2\pi} e^{i(-\ell+k+n)y}\mathrm{d}y-\int_0^{2\pi} e^{i(-\ell+k-n)y}\mathrm{d}y\right).$$

Pour tout entier k, l'intégrale $\int_0^{2\pi} e^{ikx}\mathrm{d}x$ est nulle si $k\neq 0$ et vaut 2π sinon. Dans la somme définissant $\mathrm{I}_n(x)$, les seuls termes non nuls correspondent donc soit à $\ell=k+n$ (avec $-n\leqslant k\leqslant 0$), soit à $\ell=k-n$ (avec $0\leqslant k\leqslant n$). On en déduit que

$$\begin{aligned}
\mathrm{I}_n(x)&=\frac{2\pi}{2i}\left(\sum_{k=-n}^{0} a_{k+n}\left(1-\frac{|k|}{n}\right)e^{i(k+n)x}-\sum_{k=0}^{n} a_{k-n}\left(1-\frac{|k|}{n}\right)e^{i(k-n)x}\right)\\
&=\frac{\pi}{i}\left(\sum_{k=-n}^{0} a_{k+n}\frac{n+k}{n}e^{i(k+n)x}+\sum_{k=0}^{n} a_{k-n}\frac{k-n}{n}e^{i(k-n)x}\right)\\
&=\frac{\pi}{i}\left(\sum_{k=0}^{n} a_k\frac{k}{n}e^{ikx}+\sum_{k=-n}^{0} a_k\frac{k}{n}e^{ikx}\right)=\frac{\pi}{in}\sum_{k=-n}^{n} ka_ke^{ikx}.
\end{aligned}$$

En remarquant alors que $\mathrm{P}'(x)=\sum_{k=-n}^{n} ika_ke^{ikx}$, on obtient finalement, pour tout réel x,

$$\mathrm{I}_n(x)=\frac{\pi}{ni}\frac{1}{i}\mathrm{P}'(x)=-\frac{\pi}{n}\mathrm{P}'(x).$$

À l'aide de cette égalité, on va majorer $\|\mathrm{P}'\|_\infty$. On a, pour tout $x\in\mathbb{R}$,

$$\begin{aligned}
|\mathrm{P}'(x)|&=\frac{n}{\pi}|\mathrm{I}_n(x)|\leqslant\frac{n}{\pi}\int_0^{2\pi}|\mathrm{P}(x-y)\mathrm{F}_n(y)\sin ny|\mathrm{d}y\\
&\leqslant\frac{n}{\pi}\|\mathrm{P}\|_\infty\int_0^{2\pi}|\mathrm{F}_n(y)|\mathrm{d}y \quad \text{et donc,}\\
\|\mathrm{P}'\|_\infty&\leqslant\frac{n}{\pi}\|\mathrm{P}\|_\infty\int_0^{2\pi}|\mathrm{F}_n(y)|\mathrm{d}y.
\end{aligned}$$

Pour avoir un majorant explicite, on va calculer cette dernière intégrale, et pour cela calculer F_n. Pour $x\in\mathbb{R}$ et $n\in\mathbb{N}$, on a en regroupant l'indice $-k$ avec l'indice k,

$$\mathrm{F}_n(x)=-1+\sum_{k=0}^{n-1}\left(1-\frac{k}{n}\right)(e^{ikx}+e^{-ikx})=-1+\frac{2}{n}\,\mathrm{Re}\left(\sum_{k=0}^{n-1}(n-k)e^{ikx}\right).$$

Si $e^{ix} = 1$, alors

$$\mathrm{F}_n(x) = -1 + \frac{2}{n}\left(\sum_{k=0}^{n-1}(n-k)\right) = -1 + \frac{2n(n+1)}{2n} = n.$$

Si $e^{ix} \neq 1$, alors

$$\begin{aligned}\sum_{k=0}^{n-1}(n-k)e^{ikx} &= \sum_{k=0}^{n-1}\left(\sum_{\ell=0}^{k} e^{i\ell x}\right) = \sum_{k=0}^{n-1}\left(\frac{1-e^{i(k+1)x}}{1-e^{ix}}\right)\\ &= \frac{n}{1-e^{ix}} - \frac{e^{ix}(1-e^{inx})}{(1-e^{ix})^2} = \frac{nc^{-i\frac{x}{2}}}{-2i\sin\dfrac{x}{2}} - \frac{1-e^{inx}}{-4\sin^2\dfrac{x}{2}}.\end{aligned}$$

On en déduit que

$$\mathrm{F}_n(x) = -1 + \frac{2}{n}\left(\frac{n}{2} + \frac{1-\cos nx}{4\sin^2\dfrac{x}{2}}\right) = \frac{\sin^2\dfrac{nx}{2}}{n\sin^2\dfrac{x}{2}}.$$

On a donc $\mathrm{F}_n(y) \geqslant 0$, pour tout y. En revenant à notre intégrale, cela permet d'enlever la valeur absolue pour obtenir

$$\int_0^{2\pi} |\mathrm{F}_n(y)|\mathrm{d}y = \int_0^{2\pi} \mathrm{F}_n(y)\mathrm{d}y = \sum_{k=-n}^{n}\left(1 - \frac{|k|}{n}\right)\int_0^{2\pi} e^{iky}\mathrm{d}y = 2\pi$$

et donc finalement $\boxed{\|\mathrm{P}'\|_\infty \leqslant 2n\|\mathrm{P}\|_\infty}$. ◁

Le polynôme trigonométrique F_n est appelé noyau de Fejér d'ordre n. Son caractère positif (qui intervient déjà fortement dans la solution ci-dessus) est la clé du théorème de convergence uniforme de Fejér que le lecteur trouvera dans l'exercice 4.31.

Notons qu'on peut montrer que $\int_0^{2\pi} \mathrm{F}_n(y)|\sin ny|dy = 4$, ce qui permet d'améliorer un peu la majoration précédente puisqu'on obtient alors $c = \dfrac{4n}{\pi}$. En fait, la meilleure constante, qui est donc la norme de l'opérateur de dérivation, est $c = n$. Cela va découler de l'exercice suivant qui établit l'inégalité de Bernstein (dans le cas réel). Notons que cette dernière est démontrée d'une autre manière dans le tome 1 d'algèbre (exercice 5.37). La méthode mise en œuvre ici n'utilise que les théorèmes usuels de l'analyse réelle (théorème des valeurs intermédiaires, théorème de Rolle,...).

4.3. Inégalité de Bernstein (1912)

Soit $n \in \mathbb{N}^*$ et f un polynôme trigonométrique réel de degré inférieur à n, c'est-à-dire une fonction pour laquelle il existe des réels $a_0, a_1, \ldots, a_n, b_1, \ldots, b_n$ tels que, pour tout $x \in \mathbb{R}$,

$$f(x) = a_0 + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx).$$

1. Montrer que si f admet au moins $2n+1$ racines distinctes sur $[0, 2\pi[$, alors f est nul.

2. On suppose que $f'(0) = \|f'\|_\infty > n\|f\|_\infty$. On pose pour tout $x \in \mathbb{R}$, $g(x) = \frac{1}{n}\|f'\|_\infty \sin(nx) - f(x)$.
Montrer que g admet au moins $2n$ racines distinctes sur $[0, 2\pi[$.
Montrer que g' admet au moins $2n+1$ racines distinctes sur $[0, 2\pi]$.
Montrer que g'' admet au moins $2n+1$ racines distinctes sur $[0, 2\pi[$.
Que peut-on en déduire ?

3. Montrer que $\|f'\|_\infty \leqslant n\|f\|_\infty$. *C'est l'inégalité de Bernstein.*

(École polytechnique)

▷ **Solution.**

1. À l'aide des deux formules d'Euler $\cos kx = \frac{1}{2}(e^{ikx} + e^{-ikx})$ et $\sin kx = \frac{1}{2i}(e^{ikx} - e^{-ikx})$, on voit que

$$f(x) = a_0 + \frac{1}{2}\sum_{k=1}^{n}\Big[(a_k - ib_k)e^{ikx} + (a_k + ib_k)e^{-ikx}\Big] = e^{-inx}\mathrm{P}(e^{ix})$$

où $\mathrm{P(X)} = \frac{1}{2}\sum_{k=1}^{n}(a_k + ib_k)\mathrm{X}^{n-k} + a_0\mathrm{X}^n + \frac{1}{2}\sum_{k=1}^{n}(a_k - ib_k)\mathrm{X}^{n+k}$ est un polynôme de degré inférieur à $2n$. Par hypothèse P possède $2n+1$ racines distinctes à savoir les images par $x \longmapsto e^{ix}$ des racines distinctes dans $[0, 2\pi[$ de f. Il s'agit donc du polynôme nul. On en déduit que les coefficients a_k et b_k sont tous nuls et par suite que f est nul. On peut évidemment remplacer l'intervalle $[0, 2\pi[$ par n'importe quel intervalle semi-ouvert de longueur 2π.

2. L'idée est de regarder g aux points où le sinus est maximal ou minimal. Posons, $t_k = \frac{\pi}{2n} + \frac{k\pi}{n}$ pour $k \in [\![0, 2n]\!]$. On a $nt_k = \frac{\pi}{2} + k\pi$ et

$$g(t_k) = \frac{(-1)^k\|f'\|_\infty - nf(t_k)}{n}.$$

Comme $\|f'\|_\infty > n\|f\|_\infty$, $g(t_k)$ est strictement positif pour k pair et strictement négatif pour k impair. Le théorème des valeurs intermédiaires

garantit donc l'existence d'un zéro x_k de g dans chaque intervalle ouvert $]t_k, t_{k+1}[$, $k \in [\![0, 2n-1]\!]$. Ces racines sont deux à deux distinctes. Elles sont toutes dans $[0, 2\pi[$ sauf éventuellement x_{2n-1} qui appartient à $[t_{2n-1}, t_{2n}[= \left]2\pi - \frac{\pi}{2n}, 2\pi + \frac{\pi}{2n}\right[$. Mais dans le cas où $x_{2n-1} \geqslant 2\pi$, il suffit de la remplacer par $x_{2n-1} - 2\pi$ qui est encore racine par 2π-périodicité de g et qui appartient à l'intervalle $[0, t_0[$. Donc g admet au moins $2n$ racines distinctes dans $[0, 2\pi[$.

On a déjà $g'(0) = g'(2\pi) = 0$ et le théorème de Rolle appliqué entre deux racines consécutives de g donne au moins $2n-1$ racines deux à deux distinctes de g' dans $]0, 2\pi[$. Ainsi, g' admet au moins $2n+1$ racines distinctes dans $[0, 2\pi]$. Toujours par le théorème de Rolle, g'' admet au moins $2n$ racines dans $]0, 2\pi[$. Mais comme f' présente un maximum en 0, $f''(0) = 0$ et par suite $g''(0) = 0$. Donc g'' admet au moins $2n+1$ racines dans $[0, 2\pi[$. La question précédente permet d'affirmer que $g'' = 0$. Donc g est affine, donc constante (car bornée) et finalement nulle puisqu'elle s'annule.

On a alors $f(x) = \frac{1}{n}\|f'\|_\infty \sin(nx)$ et cela contredit l'hypothèse initiale car $\|f\|_\infty = n\|f'\|_\infty$.

3. Il est clair qu'il existe $x_0 \in \mathbb{R}$ tel que $|f'(x_0)| = \|f'\|_\infty$. Quitte à prendre $-f$ on peut supposer que $f'(x_0) > 0$. On considère alors $\tilde{f}(x) = f(x + x_0)$. C'est encore un polynôme trigonométrique de degré inférieur à n auquel il est loisible d'appliquer la question 2. Comme $\|f\|_\infty = \|\tilde{f}\|_\infty$ et de même pour les dérivées, on a $\|f'\|_\infty \leqslant n\|f\|_\infty$. ◁

L'inégalité de Bernstein s'étend aux polynômes trigonométriques de degré $\leqslant n$ à coefficients complexes. En effet, soit P *un tel polynôme. Il existe un réel x_0 tel que $|\mathrm{P}'(x_0)| = \|\mathrm{P}'\|_\infty$ car* P′ *est continu et 2π-périodique donc atteint ses bornes. Il existe alors un réel α tel que $e^{i\alpha}\mathrm{P}'(x_0) = \|\mathrm{P}'\|_\infty$. Considérons le polynôme trigonométrique réel* Q *défini par $\mathrm{Q}(x) = \mathrm{Re}(e^{i\alpha}\mathrm{P}(x))$. Il est de degré $\leqslant n$ donc, d'après le résultat de l'exercice, $\|\mathrm{Q}'\|_\infty \leqslant n\|\mathrm{Q}\|_\infty$. On a $|\mathrm{Q}(x)| \leqslant |\mathrm{P}(x)|$ pour tout réel x, donc $\|\mathrm{Q}\|_\infty \leqslant \|\mathrm{P}\|_\infty$. Par ailleurs, $\mathrm{Q}'(x_0) = e^{i\alpha}\mathrm{P}'(x_0) = \|\mathrm{P}'\|_\infty$ de sorte que $\|\mathrm{Q}'\|_\infty \geqslant \|\mathrm{P}'\|_\infty$. Il en résulte donc que $\|\mathrm{P}'\|_\infty \leqslant n\|\mathrm{P}\|_\infty$: c'est l'inégalité de Bernstein pour* P. *Comme il y a égalité pour le polynôme trigonométrique* $\sin nx$, *on a finalement démontré que la norme triple de l'opérateur de dérivation sur l'espace des polynômes trigonométriques de degré $\leqslant n$ est égale à n.*

Dans l'exercice suivant, on se ramène encore à des raisonnements sur les polynômes.

4.4. Fonction 2π-périodique orthogonale aux polynômes trigonométriques de degré $< n$

Soit $f : \mathbb{R} \to \mathbb{R}$, continue 2π-périodique et $n \in \mathbb{N}^*$. On suppose que $a_k(f) = b_k(f) = 0$ pour tout $k \in [\![0, n-1]\!]$. Montrer que f admet au moins $2n$ zéros dans $[0, 2\pi[$.

(École normale supérieure)

▷ **Solution.**

On a, par hypothèse, pour tout $k \in [\![0, n-1]\!]$,

$$\int_0^{2\pi} f(x)\cos kx\,\mathrm{d}x = \int_0^{2\pi} f(x)\sin kx\,\mathrm{d}x = 0$$

ou encore, pour tout $k \in \mathbb{Z}$, tel que $|k| < n$, $\frac{1}{\pi}\int_0^{2\pi} f(x)e^{ikx}\mathrm{d}x = 0$. Il faut en déduire que f s'annule au moins $2n$ fois sur $[0, 2\pi[$.

Cela nous rappelle l'exercice classique 1.17. Soit $f : [0,1] \to \mathbb{R}$, continue, telle que, pour $0 \leqslant k \leqslant n-1$, $\int_0^1 f(t)t^k\mathrm{d}t = 0$. Alors f s'annule au moins n fois sur $[0,1]$. La méthode consiste à remarquer qu'on a $\int_0^1 f(t)\mathrm{P}(t)\mathrm{d}t = 0$, pour tout polynôme P de degré inférieur ou égal à $n-1$, et que, si f change moins de n fois de signe sur $]0,1[$, on peut trouver $\mathrm{P} \in \mathbb{R}_{n-1}[\mathrm{X}]$ tel que $\mathrm{P}f$ garde un signe constant.

On va faire la même chose en remplaçant les polynômes par les polynômes trigonométriques. Par linéarité, on a $\int_0^{2\pi} f(x)g(x)\mathrm{d}x = 0$ pour toute fonction g de la forme $g(x) = \sum_{|k|\leqslant n-1} d_k e^{ikx}$, c'est-à-dire tout polynôme trigonométrique de degré $\leqslant n-1$.

On raisonne par l'absurde et on suppose que f s'annule moins de $2n$ fois sur $[0, 2\pi[$. On considère les r valeurs $\alpha_1, \alpha_2, \ldots, \alpha_r$ de $[0, 2\pi[$ en lesquelles f s'annule et change de signe. Nécessairement r est pair puisque, par périodicité, le signe de f est le même sur $]\alpha_r - 2\pi, \alpha_1[$ et sur $]\alpha_r, \alpha_1 + 2\pi[$ et entre les deux, il y a r changements de signe. On pose $r = 2s$. On a donc, par hypothèse, $r \leqslant 2n-2$ et $s \leqslant n-1$. Considérons le polynôme $\mathrm{P}(\mathrm{X}) = \mathrm{C}\prod_{k=1}^{r}(\mathrm{X} - e^{i\alpha_k})$, où C est une constante complexe à choisir, et g la fonction définie par $g(x) = e^{-isx}\mathrm{P}(e^{ix})$. La fonction g est un polynôme trigonométrique de degré s et, d'après les remarques précédentes, on a $\int_0^{2\pi} f(x)g(x)\mathrm{d}x = 0$.

Montrons que l'on peut choisir C pour que g soit à valeurs réelles. On a, pour tout $x \in \mathbb{R}$,

$$\begin{aligned}\overline{g(x)} &= \overline{\mathrm{C}}e^{isx}\prod_{k=1}^{r}(e^{-ix}-e^{-i\alpha_k})\\ &= \overline{\mathrm{C}}e^{isx}\prod_{k=1}^{r}(-e^{-ix}e^{-i\alpha_k})\prod_{k=1}^{r}(e^{ix}-e^{i\alpha_k})\\ &= \overline{\mathrm{C}}e^{-isx}e^{-i\sum_{k=1}^{r}\alpha_k}\prod_{k=1}^{r}(e^{ix}-e^{i\alpha_k}).\end{aligned}$$

On a $g(x)=\overline{g(x)}$ pour tout réel x si $\mathrm{C}=\overline{\mathrm{C}}e^{-i\sum_{k=1}^{r}\alpha_k}$. On peut prendre par exemple $\mathrm{C}=e^{-\frac{i}{2}\sum_{k=1}^{r}\alpha_k}$.

La fonction g est alors à valeurs réelles. En $\alpha_1,\dots,\alpha_r$, elle s'annule, mais pas sa dérivée car $g'(\alpha_k)=ie^{i(-s+1)\alpha_k}\mathrm{P}'(e^{i\alpha_k})\neq 0$, les racines de P étant simples. Elle change donc de signe en ces points. La fonction fg garde donc un signe constant sur $[0,2\pi]$. Comme elle est continue et non identiquement nulle, on a $\int_0^{2\pi}f(x)g(x)\mathrm{d}x\neq 0$, ce qui nous donne la contradiction cherchée. ◁

4.5. Une inégalité

Pour tout $n\in\mathbb{N}$, on note T_n le $\mathbb{R}$-espace vectoriel engendré par les applications $x\longmapsto\cos(2k\pi x)$ et $x\longmapsto\sin(2k\pi x)$ de $[0,1]$ dans $\mathbb{R}$, pour $0\leqslant k\leqslant n$, et V_n l'ensemble des éléments P de T_n qui vérifient, pour tout $x\in[0,1]$, $\mathrm{P}(x)\geqslant x-\dfrac{1}{2}$.

1. Montrer que, pour tout $\mathrm{P}\in\mathrm{T}_n$, on a

$$\int_0^1\mathrm{P}(x)\mathrm{d}x=\frac{1}{n+1}\sum_{k=0}^{n}\mathrm{P}\left(\frac{k}{n+1}\right).$$

2. Montrer que, pour tout $\mathrm{P}\in\mathrm{V}_n$, on a $\int_0^1\mathrm{P}(x)\mathrm{d}x\geqslant\dfrac{1}{2(n+1)}$.

(École normale supérieure)

Solution.

1. En utilisant les formules d'Euler, on peut écrire P sous la forme $\mathrm{P}(x)=\sum_{k=-n}^{n}c_ke^{i2k\pi x}$, où les c_k sont des nombres complexes. Sachant que $\int_0^1 e^{i2k\pi x}$ est égal à 0 si k n'est pas nul et à 1 sinon, on en déduit, par linéarité, que $\int_0^1\mathrm{P}(x)\mathrm{d}x=c_0$.

On calcule ensuite $\sum_{k=0}^{n} \mathrm{P}\left(\frac{k}{n+1}\right) = \sum_{\ell=-n}^{n} c_\ell \left(\sum_{k=0}^{n} e^{i2\ell\pi\frac{k}{n+1}}\right)$. On obtient, pour $\ell \in [\![-n, n]\!]$,

$$\sum_{k=0}^{n} e^{i2\ell\pi\frac{k}{n+1}} = \sum_{k=0}^{n} \left(e^{i\frac{2\ell\pi}{n+1}}\right)^k = \begin{cases} n+1 & \text{si } \ell = 0 \\ \dfrac{e^{i2\ell\pi}-1}{e^{i\frac{2\ell\pi}{n+1}}-1} = 0 & \text{si } \ell \neq 0. \end{cases}$$

On a donc $\sum_{k=0}^{n} \mathrm{P}\left(\frac{k}{n+1}\right) = (n+1)c_0 = (n+1)\int_0^1 \mathrm{P}(x)\mathrm{d}x$.

2. On applique la relation démontrée dans la première question. Il résulte de la définition de V_n que l'on a $\mathrm{P}\left(\frac{k}{n+1}\right) \geqslant \frac{k}{n+1} - \frac{1}{2}$, pour $0 \leqslant k \leqslant n$. Pour $k = 0$, on peut améliorer cette minoration. En effet, par 1-périodicité de P, on obtient $\mathrm{P}(0) = \mathrm{P}(1) \geqslant 1 - \frac{1}{2} \geqslant \frac{1}{2}$. On a finalement

$$\begin{aligned} \int_0^1 \mathrm{P}(x)\mathrm{d}x &\geqslant \frac{1}{n+1}\left(\frac{1}{2} + \sum_{k=1}^{n}\left(\frac{k}{n+1} - \frac{1}{2}\right)\right) \\ &\geqslant \frac{1}{n+1}\left(\frac{1}{2} + \frac{n(n+1)}{2(n+1)} - n\frac{1}{2}\right) = \frac{1}{2(n+1)}. \quad \lhd \end{aligned}$$

Les deux exercices suivants sont consacrés à l'inégalité de Hilbert. Le premier, dans le cas réel, utilise la formule de Parseval. Il faut garder à l'esprit que si un polynôme trigonométrique P *s'écrit* $\sum_{k=-n}^{n} c_k e^{ikx}$, *les coefficients de Fourier de* P *sont les* c_k *et qu'on peut lui appliquer tous les théorèmes de l'analyse de Fourier.*

4.6. Inégalité de Hilbert (1)

Soit $\mathrm{P} = \sum_{k=0}^{n} a_k \mathrm{X}^k \in \mathbb{R}[\mathrm{X}]$.

1. Montrer que $\int_{-1}^{1} \mathrm{P}^2(x)\mathrm{d}x = -i\int_0^{\pi} \mathrm{P}^2(e^{i\theta})e^{i\theta}\mathrm{d}\theta$.

2. En déduire l'inégalité de Hilbert,

$$\sum_{0\leqslant k,\ell\leqslant n} \frac{a_k a_\ell}{k+\ell+1} \leqslant \pi \sum_{k=0}^{n} a_k^2.$$

(École polytechnique)

▷ **Solution.**

1. Démontrons, plus généralement que l'on a, pour tout $Q \in \mathbb{R}[X]$, $\int_{-1}^{1} Q(x)dx = -i\int_0^{\pi} Q(e^{i\theta})e^{i\theta}d\theta$. Par linéarité de l'intégrale, il suffit de le démontrer pour $Q = X^n$. On trouve

$$-i\int_0^{\pi} Q(e^{i\theta})e^{i\theta}d\theta = -i\int_0^{\pi} e^{i(n+1)\theta}d\theta = \frac{-i}{i(n+1)}((-1)^{n+1}-1)$$
$$= \frac{1}{n+1}(1-(-1)^{n+1}) = \int_{-1}^{1} Q(x)dx.$$

2. Le premier membre de l'inégalité à démontrer est égal à

$$\sum_{0\leqslant k,\ell\leqslant n} a_k a_\ell \int_0^1 x^k x^\ell dx = \int_0^1 \left(\sum_{0\leqslant k,\ell\leqslant n} a_k a_\ell x^k x^\ell\right) dx = \int_0^1 P^2(x)dx.$$

De la question 1, on tire la majoration

$$\int_0^1 P^2(x)dx \leqslant \int_{-1}^1 P^2(x)dx = \left|\int_0^{\pi} P^2(e^{i\theta})e^{i\theta}d\theta\right| \leqslant \int_0^{\pi} |P(e^{i\theta})|^2 d\theta.$$

Le polynôme P étant à coefficients réels, on a $|P(e^{i\theta})|^2 = P(e^{i\theta})P(e^{-i\theta})$. La fonction qu'on intègre étant paire, on a donc

$$\int_0^{\pi} |P(e^{i\theta})|^2 d\theta = \frac{1}{2}\int_{-\pi}^{\pi} |P(e^{i\theta})|^2 d\theta = \frac{1}{2}\cdot 2\pi\sum_{k=0}^{n} a_k^2 = \pi\sum_{k=0}^{n} a_k^2,$$

d'après la formule de Parseval, ce qui donne l'inégalité voulue. ◁

On peut noter que l'inégalité est stricte si les coefficients a_k ne sont pas tous nuls puisque si P *n'est pas nul* $\int_0^1 P^2(x)dx < \int_{-1}^1 P^2(x)dx$. *Cette preuve de l'inégalité de Hilbert est due à Fejér et Riesz.*

L'énoncé suivant donne un résultat plus général.

4.7. Inégalité de Hilbert (2)

On note, pour n dans $\mathbb{Z}$, $e_n : t \in \mathbb{R} \longmapsto e^{int} \in \mathbb{C}$.

1. Calculer les coefficients de Fourier de la fonction 2π-périodique h telle que $h(t) = \pi - t$ pour t dans $[0, 2\pi[$.

2. Soient $P = \sum_{k=0}^{n} a_k e_{-k}$ et $Q = \sum_{\ell=0}^{n} b_\ell e_{-\ell}$ où les a_k et les b_ℓ sont des nombres complexes. Calculer $\frac{1}{2\pi}\int_0^{2\pi} P(t)Q(t)h(t)e_{-1}(t)dt$.

3. *Inégalité de Hilbert.* Trouver une constante C (indépendante des a_k, b_ℓ et de n) telle que

$$\left|\sum_{0\leqslant k,\ell\leqslant n}\frac{a_k b_\ell}{k+\ell+1}\right|^2 \leqslant \mathrm{C}\sum_{k=0}^{n}|a_k|^2\sum_{\ell=0}^{n}|b_\ell|^2.$$

4. On pose, pour $0\leqslant k\leqslant n$, $c_k=\sum_{j=0}^{n}\frac{a_j}{j+k+1}$. Montrer que

$$\sum_{k=0}^{n}|c_k|^2\leqslant\pi^2\sum_{j=0}^{n}|a_j|^2.$$

(École normale supérieure)

▷ **Solution.**

1. Les coefficients de Fourier de h sont $c_n(h)=\frac{1}{2\pi}\int_0^{2\pi}(\pi-t)e^{-int}\mathrm{d}t$, pour n dans $\mathbb{Z}$. Si $n\neq 0$, on trouve, en intégrant par parties

$$c_n(h)=\frac{1}{2\pi}\left(\left[\frac{(\pi-t)e^{-int}}{-in}\right]_0^{2\pi}-\int_0^{2\pi}\frac{e^{-int}}{in}\mathrm{d}t\right)=\frac{1}{in}$$

et

$$c_0(h)=\frac{1}{2\pi}\int_0^{2\pi}(\pi-t)\mathrm{d}t=\frac{1}{2\pi}\left[\pi t-\frac{t^2}{2}\right]_0^{2\pi}=0.$$

2. On notera que pour tous entiers n et m, on a $e_m e_n=e_{m+n}$. Cela donne, en développant

$$\mathrm{PQ}e_{-1}=\left(\sum_{k=0}^{n}a_k e_{-k}\right)\left(\sum_{\ell=0}^{n}b_\ell e_{-\ell}\right)e_{-1}=\sum_{0\leqslant k,\ell\leqslant n}a_k b_\ell e_{-(k+\ell+1)}.$$

On en déduit que

$$\begin{aligned}\frac{1}{2\pi}\int_0^{2\pi}\mathrm{PQ}he_{-1}&=\sum_{0\leqslant k,\ell\leqslant n}a_k b_\ell\times\frac{1}{2\pi}\int_0^{2\pi}e_{-(k+\ell+1)}h\\&=\sum_{0\leqslant k,\ell\leqslant n}a_k b_\ell c_{k+\ell+1}(h)=\frac{1}{i}\sum_{0\leqslant k,\ell\leqslant n}\frac{a_k b_\ell}{k+\ell+1}.\end{aligned}$$

3. D'après la question précédente, on a l'égalité

$$\left|\sum_{0\leqslant k,\ell\leqslant n}\frac{a_k b_\ell}{k+\ell+1}\right|^2=\left(\frac{1}{2\pi}\left|\int_0^{2\pi}\mathrm{PQ}he_{-1}\right|\right)^2$$

ce qu'on peut majorer ainsi :

$$\left|\int_0^{2\pi} \mathrm{P}\mathrm{Q}h e_{-1}\right| \leqslant \int_0^{2\pi} |\mathrm{P}||\mathrm{Q}||h||e_{-1}| \leqslant \|h\|_\infty \int_0^{2\pi} |\mathrm{P}||\mathrm{Q}| \leqslant \pi \int_0^{2\pi} |\mathrm{P}||\mathrm{Q}|.$$

D'après l'inégalité de Cauchy-Schwarz, on en déduit que :

$$\left|\sum_{0\leqslant k,\ell\leqslant n} \frac{a_k b_\ell}{k+\ell+1}\right|^2 \leqslant \frac{1}{4}\left(\int_0^{2\pi} |\mathrm{P}||\mathrm{Q}|\right)^2 \leqslant \frac{1}{4}\left(\int_0^{2\pi} |\mathrm{P}|^2\right)\left(\int_0^{2\pi} |\mathrm{Q}|^2\right).$$

En vertu de l'égalité de Parseval, $\displaystyle\int_0^{2\pi} |\mathrm{P}|^2 = 2\pi \sum_{k=0}^{n} |a_k|^2$ et $\displaystyle\int_0^{2\pi} |\mathrm{Q}|^2 = 2\pi \sum_{\ell=0}^{n} |b_\ell|^2$. On obtient finalement

$$\boxed{\left|\sum_{0\leqslant k,\ell\leqslant n} \frac{a_k b_\ell}{k+\ell+1}\right|^2 \leqslant \pi^2 \sum_{k=0}^{n} |a_k|^2 \sum_{\ell=0}^{n} |b_\ell|^2},$$

ce qui est l'inégalité demandée avec $\mathrm{C} = \pi^2$.

4. En posant $c_k = \displaystyle\sum_{j=0}^{n} \frac{a_j}{j+k+1}$, l'inégalité précédente s'écrit

$$\left|\sum_{\ell=0}^{n} b_\ell c_\ell\right|^2 \leqslant \pi^2 \sum_{k=0}^{n} |a_k|^2 \sum_{\ell=0}^{n} |b_\ell|^2,$$

inégalité valable pour tout $(b_0, b_1, \ldots, b_n)$ de $\mathbb{C}^{n+1}$. En particulier, si on prend $b_\ell = \overline{c_\ell}$ pour $0 \leqslant \ell \leqslant n$, on obtient

$$\left(\sum_{\ell=0}^{n} |c_\ell|^2\right)^2 \leqslant \pi^2 \sum_{k=0}^{n} |a_k|^2 \sum_{\ell=0}^{n} |c_\ell|^2.$$

Si $\displaystyle\sum_{\ell=0}^{n} |c_\ell|^2 \neq 0$, il reste en divisant :

$$\sum_{\ell=0}^{n} |c_\ell|^2 \leqslant \pi^2 \sum_{k=0}^{n} |a_k|^2,$$

inégalité triviale lorsque $\displaystyle\sum_{\ell=0}^{n} |c_\ell|^2 = 0$. ◁

La dernière question de l'exercice signifie que si A *est la matrice de Hilbert* $\left(\dfrac{1}{i+j+1}\right)_{0\leqslant i,j\leqslant n}$ *et* $\|\ \|_2$ *la norme triple de* $\mathcal{M}_n(\mathbb{C})$ *associée à la norme hermitienne canonique de* $\mathbb{C}^n$, *on a* $\|\mathrm{A}\|_2 \leqslant \pi$.

Commence ici une série d'exercices où il s'agit d'écrire le développement en série de Fourier de diverses fonctions – il s'agit en général de fonctions de classe $\mathcal{C}^1$ par morceaux pour lesquelles le théorème de Dirichlet s'applique – et d'utiliser ces développements pour le calcul de sommes de séries numériques, d'intégrales, ou de séries de fonctions.

Nous commençons par des exemples très classiques. On notera que dans la première question de l'exercice suivant les fonctions considérées sont de période 2. *On rappelle que pour une fonction de période* T, *les coefficients de Fourier sont définis par*

$$c_n(f) = \frac{1}{\mathrm{T}} \int_0^{\mathrm{T}} f(t) e^{-2\pi i n t/\mathrm{T}}\, dt \quad (n \in \mathbb{Z}),$$

$$a_n(f) = \frac{2}{\mathrm{T}} \int_0^{\mathrm{T}} f(t) \cos\left(\frac{2\pi}{\mathrm{T}} nt\right) dt\ ,\ b_n(f) = \frac{2}{\mathrm{T}} \int_0^{\mathrm{T}} f(t) \sin\left(\frac{2\pi}{\mathrm{T}} nt\right) dt$$

$(n \in \mathbb{N})$, et la série de Fourier de f est

$$\sum_{n=-\infty}^{+\infty} c_n(f) e^{2\pi i n t/\mathrm{T}}$$

ou

$$\frac{a_0}{2} + \sum_{n=1}^{+\infty} \left(a_n \cos \frac{2\pi}{\mathrm{T}} nt + b_n \sin \frac{2\pi}{\mathrm{T}} nt\right).$$

4.8. Calcul de sommes de séries

1. Soit f_1 (resp. f_2, f_3) la fonction 2-périodique définie par $f_1(t) = -1$ (resp. $-t$, $-t^2$) sur $]-1, 0[$ et $f_1(t) = 1$ (resp. t, t^2) sur $[0, 1]$. Développer ces fonctions en série de Fourier. Calculer

$$\sum_{p=0}^{+\infty} \frac{(-1)^p}{(2p+1)^3}.$$

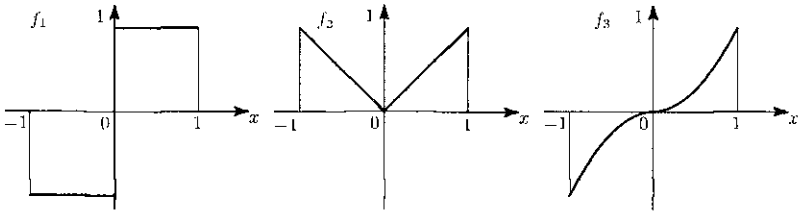
2. Calculer les coefficients de Fourier de la fonction 2π-périodique qui vaut x^2 pour $x \in [-\pi, \pi]$.

En déduire la valeur de $\displaystyle\sum_{n=1}^{+\infty} \frac{1}{n^2}$.

(École polytechnique)

Solution.

1. Les fonctions f_1, f_2, f_3 sont $\mathcal{C}^1$ par morceaux sur $\mathbb{R}$. La restriction de f_1 (resp. f_2, resp. f_3) à $]-1,0[\cup]0,1[$ est impaire, (resp. paire, resp. impaire).



On en déduit que $a_n(f_1) = a_n(f_3) = b_n(f_2) = 0$ pour tout $n \in \mathbb{N}$. Calculons les autres coefficients. Pour $n \in \mathbb{N}^*$,

$$b_n(f_1) = \int_{-1}^{1} f_1(t)\sin(\pi nt)\mathrm{d}t = 2\int_0^1 \sin(\pi nt)\mathrm{d}t = \frac{2}{\pi n}(1-(-1)^n).$$

Les coefficients d'indice pair sont nuls et $b_{2n+1}(f_1) = \dfrac{4}{\pi(2n+1)}$. On calcule $a_n(f_2) = \int_{-1}^{1} f_2(t)\cos(\pi nt)\mathrm{d}t = 2\int_0^1 t\cos(\pi nt)\mathrm{d}t$ en intégrant par parties. On obtient $a_0(f_2) = 1$ et pour $n \in \mathbb{N}^*$,

$$a_n(f_2) = \left[2t\frac{\sin(\pi nt)}{n\pi}\right]_0^1 - \frac{2}{\pi n}\int_0^1 \sin(\pi nt)\mathrm{d}t = 0 + \frac{2}{\pi^2 n^2}((-1)^n - 1).$$

Les coefficients d'indice pair (non nul) sont tous nuls et pour tout $n \in \mathbb{N}$, $a_{2n+1}(f_2) = -\dfrac{4}{\pi^2(2n+1)^2}$.

Pour $n \in \mathbb{N}^*$, $b_n(f_3) = \int_{-1}^{1} f_2(t)\sin(\pi nt)\mathrm{d}t = 2\int_0^1 t^2\sin(\pi nt)\mathrm{d}t$ s'obtient en intégrant deux fois par parties. On trouve

$$\begin{aligned} b_n(f_3) &= \left[2t^2\frac{-\cos(\pi nt)}{\pi n}\right]_0^1 + \frac{4}{\pi n}\int_0^1 t\cos(\pi nt)\mathrm{d}t \\ &= \frac{2}{\pi n}(-1)^{n+1} + \frac{4}{\pi^2 n^2}\left[t\sin(\pi nt)\right]_0^1 - \frac{4}{\pi^2 n^2}\int_0^1 \sin(\pi nt)\mathrm{d}t \\ &= \frac{2}{\pi n}(-1)^{n+1} - \frac{4}{\pi^3 n^3}(1-(-1)^n). \end{aligned}$$

En distinguant selon la parité de n, on écrit

$$b_{2n}(f_3) = -\frac{1}{\pi n} \quad \text{et} \quad b_{2n+1}(f_3) = \frac{2}{\pi(2n+1)} - \frac{8}{\pi^3(2n+1)^3}.$$

Les fonctions f_1, f_2, f_3 étant $\mathcal{C}^1$ par morceaux, leur série de Fourier converge en tout point x vers $\frac{1}{2}(f(x+0)+f(x-0))$ et donc vers $f(x)$ si $x \notin \mathbb{Z}$. On a donc

$$\sum_{n=0}^{+\infty} \frac{4}{\pi(2n+1)} \sin(\pi(2n+1)x) = \begin{cases} -1 & \text{si } x \in]-1,0[, \\ 0 & \text{si } x \in \mathbb{Z} \\ 1 & \text{si } x \in]0,1[\end{cases}$$

$$\frac{1}{2} - \sum_{n=0}^{+\infty} \frac{4}{\pi^2(2n+1)^2} \cos(\pi(2n+1)x) = \begin{cases} -x & \text{si } x \in]-1,0[, \\ 0 & \text{si } x \in \mathbb{Z} \\ x & \text{si } x \in]0,1[\end{cases}$$

$$\sum_{n=1}^{+\infty} \left(\frac{-1}{\pi n}\right) \sin(2\pi nx) + \sum_{n=0}^{+\infty} \left(\frac{2}{\pi(2n+1)} - \frac{8}{\pi^3(2n+1)^3}\right) \sin(\pi(2n+1)x)$$
$$= \begin{cases} -x^2 & \text{si } x \in]-1,0[, \\ 0 & \text{si } x \in \mathbb{Z} \\ x^2 & \text{si } x \in]0,1[. \end{cases}$$

En particulier, si on prend $x = \frac{1}{2}$ dans la dernière formule, on obtient

$$\sum_{n=0}^{+\infty} \frac{2}{\pi(2n+1)}(-1)^n - \sum_{n=0}^{+\infty} \frac{8}{\pi^3(2n+1)^3}(-1)^n = \frac{1}{4}.$$

Mais, la première égalité donne, pour $x = \frac{1}{2}$, $\sum_{n=0}^{+\infty} \frac{4(-1)^n}{\pi(2n+1)} = 1$. On a donc $\sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+1)} = \frac{\pi}{4}$ et $\sum_{n=0}^{+\infty} \frac{(-1)^n}{(2n+1)^3} = \frac{\pi^3}{8}\left(\frac{2}{\pi}\cdot\frac{\pi}{4} - \frac{1}{4}\right) = \frac{\pi^3}{32}$.

2. La fonction 2π-périodique f qui vaut x^2 pour $x \in [-\pi,\pi]$ est paire, continue et $\mathcal{C}^1$ par morceaux. Elle est partout égale à sa série de Fourier. On a $a_0 = \frac{2}{\pi}\int_0^\pi t^2 \mathrm{d}t = \frac{2\pi^2}{3}$ et, pour tout $n \in \mathbb{N}$, en intégrant par parties,

$$a_n(f) = \frac{2}{\pi}\int_0^\pi t^2\cos(nt)\mathrm{d}t = \frac{2}{\pi n}\Big[t^2\sin(nt)\Big]_0^\pi - \frac{4}{\pi n}\int_0^\pi t\sin(nt)\mathrm{d}t$$
$$= 0 + \frac{4}{\pi n^2}\left[t\cos(nt)\right]_0^\pi - \frac{4}{\pi n^2}\int_0^\pi \cos(nt)\mathrm{d}t = \frac{4}{n^2}(-1)^n.$$

On a donc, pour tout $x \in \mathbb{R}$,

$$f(x) = \frac{\pi^2}{3} + \sum_{n=1}^{+\infty} \frac{4(-1)^n}{n^2}\cos nx.$$

En particulier en prenant $x = \pi$ on obtient $\pi^2 = \dfrac{\pi^2}{3} + \sum_{n=1}^{+\infty} \dfrac{4}{n^2}$ ce qui donne la somme classique $\sum_{n=1}^{+\infty} \dfrac{1}{n^2} = \dfrac{\pi^2}{6}$. $\lhd$

L'énoncé suivant redémontre le développement eulérien de la fonction cotangente, déjà obtenu par une méthode plus élémentaire dans l'exercice 2.26.

4.9. Développement eulérien de la fonction cotangente

Soit $0 < u < 1$ et $f : \mathbb{R} \longrightarrow \mathbb{R}$, 2π-périodique, telle que pour tout $x \in [-\pi, \pi]$, $f(x) = \cos(ux)$.

1. Calculer les coefficients de Fourier de f.
2. Calculer $g(u) = \sum_{n=1}^{+\infty} \dfrac{2u}{u^2 - n^2}$.
3. Pour $t \in]0, 1[$, calculer $\mathrm{I}(t) = \int_0^t g(u)\mathrm{d}u$.

(École polytechnique)

$\rhd$ **Solution.**

La fonction f est continue sur $\mathbb{R}$, car $\lim_{t \to \pi^-} f(t) = \cos(u\pi) = \lim_{t \to -\pi^+} f$, et de classe $\mathcal{C}^1$ par morceaux.

1. La parité de f nous invite à considérer plutôt les coefficients trigonométriques. Les b_n sont tous nuls. Calculons donc pour $n \geqslant 0$

$$\begin{aligned} a_n &= \frac{1}{\pi} \int_{-\pi}^{\pi} \cos(nx)\cos(ux)\mathrm{d}u = \frac{2}{\pi} \int_0^{\pi} \cos(nx)\cos(ux)\mathrm{d}u \\ &= \frac{1}{\pi} \int_0^{\pi} (\cos(n+u)x + \cos(n-u)x)\,\mathrm{d}u \\ &= \frac{1}{\pi} \left(\frac{\sin(n\pi + u\pi)}{n+u} - \frac{\sin(u\pi - n\pi)}{n-u} \right) \\ &= \frac{(-1)^{n+1}}{\pi} \frac{2u}{n^2 - u^2} \sin \pi u. \end{aligned}$$

Puisque f est continue, $\mathcal{C}^1$ par morceaux, sa série de Fourier converge normalement sur $\mathbb{R}$ et a pour somme f. On a donc, si $x \in [-\pi, \pi]$,

$$\cos ux = \frac{\sin \pi u}{\pi u} + \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{\pi} \frac{2u}{n^2 - u^2} \sin \pi u \cos nx.$$

2. On obtient, en substituant π à x dans la dernière formule,

$$\cos \pi u = \frac{\sin \pi u}{\pi u} - \frac{1}{\pi} \sum_{n=1}^{+\infty} \frac{2u}{n^2 - u^2} \sin \pi u = \frac{\sin \pi u}{\pi u} + \frac{g(u) \sin \pi u}{\pi}.$$

En divisant par $\sin \pi u$, il reste $\operatorname{cotan} \pi u = \dfrac{1}{\pi u} + \dfrac{g(u)}{\pi}$, et finalement

$$\boxed{g(u) = \pi \operatorname{cotan} \pi u - \frac{1}{u}}.$$

3. Il est clair sur l'expression ci-dessus que g est continue sur $]0,1[$ et qu'elle se prolonge par continuité en 0 par $g(0) = 0$. Si $\alpha \in]0,t[$, on peut écrire

$$\begin{aligned} \int_\alpha^t g(u)\mathrm{d}u &= \int_\alpha^t \left(\pi \operatorname{cotan} \pi u - \frac{1}{u}\right) \mathrm{d}u = \left[\ln \sin \pi u - \ln u\right]_\alpha^t \\ &= \ln\left(\frac{\sin \pi t}{t}\right) - \ln\left(\frac{\sin \pi \alpha}{\alpha}\right) \\ &\xrightarrow[\alpha \to 0]{} \ln\left(\frac{\sin \pi t}{t}\right) - \ln \pi = \ln\left(\frac{\sin \pi t}{\pi t}\right) \end{aligned}$$

On conclut donc $\boxed{\displaystyle\int_0^t g(u)\mathrm{d}u = \ln\left(\frac{\sin \pi t}{\pi t}\right)}$. ◁

Il manque sans doute une dernière question à l'exercice : en intégrant terme à terme la série (ce qui est possible par convergence normale sur le segment $[0,t]$) on obtient

$$\int_0^t g(u)\,du = \sum_{n=1}^{+\infty} \ln\left(1 - \frac{t^2}{n^2}\right).$$

En passant à l'exponentielle et en tenant compte du résultat de la question 3 on obtient donc $\dfrac{\sin \pi t}{\pi t} = \displaystyle\prod_{n=1}^{+\infty} \left(1 - \frac{t^2}{n^2}\right)$ *pour tout* $t \in]0,1[$. *Il s'agit du développement eulérien de la fonction sinus, qui reste valable sur* $\mathbb{C}$ *comme le montre l'exercice 2.29.*

Dans les trois exercices suivants, on ne calcule pas directement les coefficients de Fourier de la fonction f considérée, mais on obtient, à l'aide d'un développement en série entière, l'écriture de f comme somme d'une série trigonométrique. On démontre ensuite que celle-ci est la série de Fourier de f. L'argument utilisé doit être retenu : si $\sum c_n e^{inx}$ (ou $\sum(a_n \cos nx + b_n \sin nx)$) converge uniformément (et en particulier normalement) sur $\mathbb{R}$ et si on note $f(x)$ la somme de la série[1], *alors les*

1. Naturellement, $f(x) = \dfrac{a_0}{2} + \displaystyle\sum_{n \geqslant 1} (a_n \cos nx + b_n \sin nx)$ dans le cas des coefficients trigonométriques.

c_n (ou les a_n et b_n) sont les coefficients de Fourier de f. Démontrons ce résultat pour les coefficients exponentiels, la démonstration étant la même pour les coefficients trigonométriques. On a, pour tout réel x, $f(x) = \sum_{n=-\infty}^{+\infty} c_n e^{inx}$. La fonction f est continue sur $\mathbb{R}$, puisque la convergence est uniforme et, pour tout $n \in \mathbb{Z}$,

$$c_n(f) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(t)e^{-int}\,dt = \frac{1}{2\pi}\int_{-\pi}^{\pi} \sum_{k=-\infty}^{+\infty} c_k e^{i(k-n)t}\,dt.$$

La convergence de la série trigonométrique étant uniforme, on peut intervertir sommation et intégration et on obtient

$$c_n(f) = \sum_{k=-\infty}^{+\infty} c_k \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i(k-n)t}\,dt = c_n,$$

compte tenu des relations d'orthogonalité entre les fonctions $t \longmapsto e^{int}$.

4.10. Développement en série de Fourier (1)

Développer la fonction $x \longmapsto \dfrac{1+\cos x}{4-2\cos x}$ en série de Fourier.

(École polytechnique)

▷ **Solution.**

On exprime $f(x) = \dfrac{1+\cos x}{4-2\cos x}$ en fonction de e^{ix} et on utilise un développement en série entière de la fraction rationnelle obtenue. On a pour tout $x \in \mathbb{R}$,

$$\begin{aligned}\frac{1+\cos x}{4-2\cos x} &= -\frac{1}{2} + \frac{3}{4-2\cos x} = -\frac{1}{2} + \frac{3}{4-e^{ix}-e^{-ix}} \\ &= -\frac{1}{2} + \frac{3e^{ix}}{-(e^{ix})^2 + 4e^{ix} - 1}.\end{aligned}$$

On décompose en éléments simples la fraction rationnelle $\dfrac{3X}{-X^2+4X-1}$. On obtient

$$\begin{aligned}\frac{3X}{-X^2+4X-1} &= \frac{3X}{-(X-(2-\sqrt{3}))(X-(2+\sqrt{3}))} \\ &= \frac{\sqrt{3}}{2}\left(\frac{2-\sqrt{3}}{X-(2-\sqrt{3})} - \frac{2+\sqrt{3}}{X-(2+\sqrt{3})}\right)\end{aligned}$$

et donc, pour tout réel x,

$$f(x) = -\frac{1}{2} + \frac{\sqrt{3}}{2}\left(\frac{2-\sqrt{3}}{e^{ix}-(2-\sqrt{3})} - \frac{2+\sqrt{3}}{e^{ix}-(2+\sqrt{3})}\right).$$

On peut écrire alors, pour tout réel x,

$$\begin{aligned}\frac{1}{e^{ix}-(2-\sqrt{3})} &= \frac{e^{-ix}}{1-(2-\sqrt{3})e^{-ix}} = e^{-ix}\sum_{n=0}^{+\infty}\left((2-\sqrt{3})e^{-ix}\right)^n \\ &= \sum_{n=0}^{+\infty}(2-\sqrt{3})^n e^{-i(n+1)x},\end{aligned}$$

car $|(2-\sqrt{3})e^{-ix}| \leqslant 1$. On a, de même, sachant que $\dfrac{1}{2+\sqrt{3}} = 2-\sqrt{3}$,

$$\begin{aligned}\frac{1}{e^{ix}-(2+\sqrt{3})} &= -\frac{1}{(2+\sqrt{3})\left(1-(2-\sqrt{3})e^{ix}\right)} \\ &= -(2-\sqrt{3})\sum_{n=0}^{+\infty}\left((2-\sqrt{3})e^{ix}\right)^n\end{aligned}$$

On en déduit que, pour tout réel x,

$$\begin{aligned}f(x) &= -\frac{1}{2} + \frac{\sqrt{3}}{2}\left(\sum_{n=0}^{+\infty}(2-\sqrt{3})^{n+1}e^{-i(n+1)x} + \sum_{n=0}^{+\infty}(2-\sqrt{3})^n e^{inx}\right) \\ &= \frac{-1+\sqrt{3}}{2} + \sqrt{3}\sum_{n=1}^{+\infty}(2-\sqrt{3})^n\cos nx.\end{aligned}$$

Cette série converge normalement sur $\mathbb{R}$ car, pour $n \geqslant 1$,

$$|(2-\sqrt{3})^n\cos nx| \leqslant (2-\sqrt{3})^n.$$

On en déduit que c'est le développement en série de Fourier de f.

Conclusion. Le développement de f en série de Fourier est donc

$$\boxed{f(x) = \frac{-1+\sqrt{3}}{2} + \sqrt{3}\sum_{n=1}^{+\infty}(2-\sqrt{3})^n\cos nx}. \quad \triangleleft$$

On peut en déduire la valeur des intégrales $\int_{-\pi}^{\pi} f(t)\cos nt\,dt$ pour tout entier n. Cette idée est à la base de l'énoncé qui suit : on calcule des intégrales en les interprétant comme les coefficients de Fourier d'une fonction dont on détermine le développement en série de Fourier via l'utilisation de séries entières.

4.11. Calcul d'intégrales par développement en série de Fourier

Calculer pour $a \in \mathbb{R}_+^*$ et $n \in \mathbb{N}$,

$$\mathrm{I}_n = \frac{1}{\pi}\int_0^{2\pi} \frac{\cos nx}{\operatorname{ch} a - \sin x}\mathrm{d}x \quad \text{et} \quad \mathrm{J}_n = \frac{1}{\pi}\int_0^{2\pi} \frac{\sin nx}{\operatorname{ch} a - \sin x}\mathrm{d}x.$$

(École polytechnique)

▷ **Solution.**

Soit f la fonction $x \longmapsto \dfrac{1}{\operatorname{ch} a - \sin x}$. Pour $a > 0$, f est définie sur $\mathbb{R}$ et 2π-périodique. Il s'agit de calculer ses coefficients de Fourier. La fonction f est de classe $\mathcal{C}^1$ sur $\mathbb{R}$. Elle est donc égale à sa série de Fourier. Pour déterminer celle-ci, on exprime $f(x)$ en fonction de e^{ix} et on utilise des développements en série entière.

On a, pour tout $x \in \mathbb{R}$,

$$f(x) = \frac{1}{\operatorname{ch} a - \dfrac{1}{2i}(e^{ix} - e^{-ix})} = \frac{2ie^{ix}}{-e^{2ix} + 2ie^{ix}\operatorname{ch} a + 1}.$$

On considère la fraction rationnelle $\mathrm{F} = \dfrac{2i\mathrm{X}}{-\mathrm{X}^2 + 2i\mathrm{X}\operatorname{ch} a + 1}$ que l'on décompose en éléments simples. On obtient

$$\mathrm{F} = \frac{-2i\mathrm{X}}{(\mathrm{X} - ie^a)(\mathrm{X} - ie^{-a})} = -\frac{i}{\operatorname{sh} a}\left(\frac{e^a}{\mathrm{X} - ie^a} - \frac{e^{-a}}{\mathrm{X} - ie^{-a}}\right).$$

On en déduit que, pour tout réel x,

$$f(x) = -\frac{i}{\operatorname{sh} a}\left(\frac{e^a}{e^{ix} - ie^a} - \frac{e^{-a}}{e^{ix} - ie^{-a}}\right).$$

On développe chaque terme en série :

$$\frac{e^a}{e^{ix} - ie^a} = \frac{i}{1 + ie^{-a}e^{ix}} = i\sum_{n=0}^{+\infty}(-ie^{-a}e^{ix})^n = \sum_{n=0}^{+\infty}(-1)^n i^{n+1}e^{-na}e^{inx},$$

car $|ie^{-a}e^{ix}| = e^{-a} < 1$ et, de même,

$$\begin{aligned}\frac{e^{-a}}{e^{ix} - ie^{-a}} &= \frac{e^{-a}e^{-ix}}{1 - ie^{-a}e^{-ix}} = e^{-a}e^{-ix}\sum_{n=0}^{+\infty}(ie^{-a}e^{-ix})^n \\ &= \sum_{n=0}^{+\infty} i^n e^{-(n+1)a}e^{-(n+1)ix}.\end{aligned}$$

Finalement, on peut écrire, pour tout réel x,

$$\begin{aligned} f(x) &= \frac{-i}{\operatorname{sh} a}\left(i+\sum_{n=1}^{+\infty} e^{-na}\left(i^{n+1}(-1)^n e^{inx}-i^{n-1}e^{-inx}\right)\right) \\ &= \frac{1}{\operatorname{sh} a}\left(1+\sum_{n=1}^{+\infty} e^{-na}\, i^n\left((-1)^n e^{inx}+e^{-inx}\right)\right) \\ &= \frac{1}{\operatorname{sh} a}\left(1+\sum_{p=1}^{+\infty} 2e^{-2pa}(-1)^p\cos 2px \right. \\ &\qquad \left. +\sum_{p=0}^{+\infty} 2e^{-(2p+1)a}(-1)^p\sin(2p+1)x\right). \end{aligned}$$

Les séries $\sum e^{-2pa}(-1)^p\cos 2px$ et $\sum e^{-(2p+1)a}(-1)^p\sin(2p+1)x$ sont normalement convergentes, car pour tout réel x et tout entier p, $|e^{-2pa}(-1)^p\cos 2px|$ et $|e^{-(2p+1)a}(-1)^p\sin(2p+1)x|$ sont inférieurs à $(e^{-a})^{2p}$, avec $e^{-a}<1$.

On en déduit que l'écriture précédente est le développement de f en série de Fourier. D'où l'on tire, pour tout $p\in\mathbb{N}$,

$$\mathrm{I}_0=\frac{2}{\operatorname{sh} a}\ ;\ \mathrm{I}_{2p}=\frac{2(-1)^p e^{-2pa}}{\operatorname{sh} a}\ \text{si } p\geqslant 1\ ;\ \mathrm{I}_{2p+1}=0\ ;$$

$$\mathrm{J}_{2p}=0\ \text{ et }\ \mathrm{J}_{2p+1}=\frac{2(-1)^p e^{-(2p+1)a}}{\operatorname{sh} a}\cdot\ \lhd$$

4.12. Développement en série de Fourier (2)

On pose $f(x,\theta)=\arctan\left(\dfrac{1-x}{1+x}\tan\theta\right)$.

1. Quel est le domaine de définition de f ?

2. Soit $\theta\in\left]0,\dfrac{\pi}{2}\right[$. Développer en série entière $x\longmapsto f(x,\theta)$.

3. Soit $x\in\left]-1,1\right[$. Développer en série de Fourier $\theta\longmapsto f(x,\theta)$.

(**École polytechnique**)

▷ **Solution.**

1. Si $(x,\theta)\in\mathbb{R}^2$, $f(x,\theta)$ est défini pour $x\neq -1$ et $\theta\not\equiv\dfrac{\pi}{2}\pmod{\pi}$.

2. La fonction $x\longmapsto f(x,\theta)$ est $\mathcal{C}^\infty$ sur $\mathbb{R}\backslash\{-1\}$. Pour $x\neq -1$,

$$\frac{\partial f}{\partial x}(x,\theta)=\frac{\tan\theta\left(\dfrac{-2}{(1+x)^2}\right)}{1+\left(\dfrac{1-x}{1+x}\right)^2\tan^2\theta}\cdot$$

Après simplifications, $\dfrac{\partial f}{\partial x}(x,\theta) = \dfrac{-\sin 2\theta}{x^2+2x\cos 2\theta+1}$. Puisque 0 n'est pas un pôle de cette fraction rationnelle, $x \longmapsto \dfrac{\partial f}{\partial x}(x,\theta)$ est développable en série entière en 0 (voir l'exercice 3.12). Décomposons-la en éléments simples. Comme

$$\begin{aligned} x^2+2x\cos 2\theta+1 &= (x+\cos 2\theta)^2+\sin^2(2\theta) \\ &= (x+e^{2i\theta})(x+e^{-2i\theta}) \\ &= (1+xe^{-2i\theta})(1+xe^{2i\theta}), \end{aligned}$$

il existe $(\lambda,\mu)\in\mathbb{C}^2$ tel que

$$\frac{-\sin 2\theta}{\mathrm{X}^2+2\mathrm{X}\cos 2\theta+1} = \frac{\lambda}{1+\mathrm{X}e^{2i\theta}} + \frac{\mu}{1+\mathrm{X}e^{-2i\theta}}.$$

Comme la première fraction est réelle, on a $\lambda=\overline{\mu}$. On obtient λ en multipliant la décomposition ci-dessus par $1+\mathrm{X}e^{2i\theta}$ et en évaluant l'égalité en $-e^{-2i\theta}$. Il vient

$$\lambda = \frac{-\sin 2\theta}{1-e^{-4i\theta}} = \frac{-e^{2i\theta}\sin 2\theta}{2i\sin 2\theta} = \frac{-e^{2i\theta}}{2i}.$$

On obtient finalement

$$\frac{\partial f}{\partial x}(x,\theta) = -\frac{1}{2i}\left(\frac{e^{2i\theta}}{1+xe^{2i\theta}} - \frac{e^{-2i\theta}}{1+xe^{-2i\theta}}\right).$$

La fonction $x\longmapsto \dfrac{\partial f}{\partial x}(x,\theta)$ est développable en série entière sur $]-1,1[$ et, pour $|x|<1$,

$$\begin{aligned} \frac{\partial f}{\partial x}(x,\theta) &= -\frac{1}{2i}\left(e^{2i\theta}\sum_{n=0}^{+\infty}(-1)^n(e^{2i\theta})^n x^n - e^{-2i\theta}\sum_{n=0}^{+\infty}(-1)^n(e^{-2i\theta})^n x^n\right) \\ &= -\frac{1}{2i}\left(\sum_{n=0}^{+\infty}(-1)^n(e^{2i(n+1)\theta}-e^{-2i(n+1)\theta})x^n\right) \\ &= \sum_{n=0}^{+\infty}(-1)^{n+1}\sin(2(n+1)\theta)x^n. \end{aligned}$$

Ainsi, la fonction $x\longmapsto f(x,\theta)$ est développable en série entière sur $]-1,1[$. Puisque $\theta\in\left]0,\dfrac{\pi}{2}\right[$, on a $f(0,\theta)=\arctan(\tan\theta)=\theta$ et par intégration terme à terme, on obtient

$$f(x,\theta) = \theta + \sum_{n=0}^{+\infty}(-1)^{n+1}\frac{\sin 2(n+1)\theta}{n+1}x^{n+1}$$

c'est-à-dire

$$\boxed{f(x,\theta) = \theta + \sum_{n=1}^{+\infty} (-1)^n \frac{\sin 2n\theta}{n} x^n}.$$

3. La fonction $g : \theta \longmapsto f(x,\theta)$ est définie pour $\theta \in \mathbb{R}\backslash\left(\frac{\pi}{2} + \pi\mathbb{Z}\right)$; elle est π-périodique et impaire. On a $\lim\limits_{\theta \to \frac{\pi}{2}-} f(x,\theta) = \frac{\pi}{2}$ et $\lim\limits_{\theta \to -\frac{\pi}{2}+} f(x,\theta) = -\frac{\pi}{2}$. Donc g est continue par morceaux. D'après la question précédente, pour $\theta \in \left]0, \frac{\pi}{2}\right[$ on a

$$g(\theta) = \theta + \sum_{n=1}^{+\infty} (-1)^n \frac{\sin 2n\theta}{n} x^n.$$

Comme g est impaire, ainsi que les fonctions $\theta \longmapsto \sin 2n\theta$, cette relation reste vraie pour $\theta \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$.

Le développement en série de Fourier de la série ne pose pas de problème : pour $\theta \in \mathbb{R}$, posons $k(\theta) = \sum\limits_{n=1}^{+\infty} (-1)^n \frac{\sin 2n\theta}{n} x^n$. Cette série converge normalement car, pour $n \geqslant 1$, $\left|(-1)^n \frac{\sin 2n\theta}{n} x^n\right| \leqslant |x|^n$. En particulier, k est continue et $\sum\limits_{n=1}^{+\infty} (-1)^n \frac{\sin 2n\theta}{n} x^n$ est précisément sa série de Fourier.

Il ne reste plus qu'à s'intéresser au premier terme θ. Considérons la fonction π-périodique h définie par $h(\theta) = \theta$ pour tout $\theta \in \left]-\frac{\pi}{2}, \frac{\pi}{2}\right]$ et $h\left(\frac{\pi}{2}\right) = 0$. Elle est $\mathcal{C}^1$ par morceaux et d'après le théorème de convergence simple de Dirichlet, sa série de Fourier converge simplement vers h, puisque $h\left(\frac{\pi}{2}\right) = 0 = \frac{1}{2}\left(h\left(\frac{\pi}{2}+\right) + h\left(\frac{\pi}{2}-\right)\right)$. La fonction h étant impaire et π-périodique, on a pour $n \geqslant 1$,

$$b_n(h) = \frac{2}{\pi} \int_0^{\pi} \theta \sin n\theta \mathrm{d}\theta.$$

En intégrant par parties, on obtient

$$b_n(h) = \frac{2}{\pi}\left(\left[-\frac{\theta \cos n\theta}{n}\right]_0^{\pi} + \int_0^{\pi} \frac{\cos n\theta}{n} \mathrm{d}\theta\right) = -\frac{2\cos n\pi}{n}.$$

Ainsi $b_n(h) = 0$ si n est impair et $b_{2n} = \frac{(-1)^{n-1}}{n}$ si $n \geqslant 1$. On obtient donc, pour tout $\theta \in \mathbb{R}$.

$$h(\theta) = \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n} \sin 2n\theta.$$

Les fonctions g, h et k sont π-périodiques et g coïncide sur $\left]-\frac{\pi}{2}, \frac{\pi}{2}\right[$ avec $h+k$. Elles coïncident donc sur $\mathbb{R}\setminus\left(\frac{\pi}{2}+\pi\mathbb{Z}\right)$. Disposant des séries de Fourier de h et k, on obtient celle de g en faisant la somme :

$$f(x,\theta) = g(\theta) = \sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n} \sin(2n\theta) + \sum_{n=1}^{+\infty} \frac{(-1)^n}{n} \sin(2n\theta)x^n,$$

ce qui s'écrit encore

$$\boxed{f(x,\theta) = \sum_{n=1}^{+\infty} \frac{(-1)^n}{n} \sin(2n\theta)(x^n-1)}. \quad \triangleleft$$

4.13. Calcul de la somme d'une série trigonométrique

1. Soit φ impaire, 2π-périodique, avec $\varphi(x) = \frac{x}{2}$ si $x \in]-\pi, \pi[$. Déterminer la série de Fourier de φ et prouver que celle-ci converge uniformément sur tout segment inclus dans $]-\pi, \pi[$.

2. Soit $a \in \mathbb{R}_+ \setminus \mathbb{N}$. On pose $f(x) = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1} \sin nx}{n(n^2-a^2)}$. Prouver que $f \in \mathcal{C}^1(\mathbb{R},\mathbb{R})$. Quelle est sa série de Fourier ?

3. Prouver que f est de classe $\mathcal{C}^2$ sur l'intervalle $]-\pi, \pi[$ et calculer $f''(x) + a^2 f(x)$ pour tout $x \in]-\pi, \pi[$.

4. Calculer $f(x)$ pour tout $x \in [-\pi, \pi]$.

(École polytechnique)

▷ **Solution.**

1. La fonction φ n'étant pas continue sur $\mathbb{R}$, nous ne sommes pas dans les conditions d'application du théorème de convergence uniforme de la série de Fourier du programme. On va démontrer directement le résultat demandé. Puisque φ est impaire, on a $a_n(\varphi) = 0$ pour tout $n \in \mathbb{N}$ et pour $n \geqslant 1$,

$$b_n(\varphi) = \frac{2}{\pi}\int_0^\pi \frac{x}{2} \sin nx \, \mathrm{d}x = \frac{1}{\pi}\left(\left[-\frac{x\cos nx}{n}\right]_0^\pi + \underbrace{\int_0^\pi \frac{\cos nx}{n}\mathrm{d}x}_{=0}\right)$$

$$= \frac{(-1)^{n+1}}{n}.$$

Posons, pour $n \in \mathbb{N}^*$, $s_n(x) = \sum_{k=1}^{n} \frac{(-1)^{k+1}}{k} \sin kx$. Le théorème de Dirichlet assure que $s_n(x)$ converge vers $\varphi(x)$ pour tout x de $]-\pi, \pi[$. Il

nous faut montrer que la convergence est uniforme sur tout segment inclus dans $]-\pi,\pi[$. Un tel segment étant inclus dans un segment de la forme $[-\alpha,\alpha]$, avec $\alpha\in[0,\pi[$, on se placera dans ce dernier cas. On effectue une transformation d'Abel, en posant $\mathrm{U}_n(x)=\sum_{k=0}^{n}(-1)^{k+1}\sin kx$. Il s'agit de l'opposé de la partie imaginaire de

$$\sum_{k=0}^{n}(-e^{ix})^k=\frac{1-(-e^{ix})^{n+1}}{1+e^{ix}}=\frac{1-(-e^{ix})^{n+1}}{e^{i\frac{x}{2}}2\cos\dfrac{x}{2}},$$

ce calcul étant valide pour $x\in[-\alpha,\alpha]$ car dans ce cas $-e^{ix}$ est distinct de 1. Il en résulte que $|\mathrm{U}_n(x)|\leqslant\dfrac{1}{\cos\dfrac{x}{2}}\cdot$ Sachant que, pour tout x de $[-\alpha,\alpha]$, on a $\cos\dfrac{x}{2}\geqslant\cos\dfrac{\alpha}{2}$, on en déduit que pour tout entier $n\geqslant 1$ et tout $x\in[-\alpha,\alpha]$, $|\mathrm{U}_n(x)|\leqslant\dfrac{1}{\cos\dfrac{\alpha}{2}}\cdot$ Autrement dit, la suite (U_n) est uniformément bornée sur $[-\alpha,\alpha]$. On a alors, pour $n\geqslant 1$ et $p\geqslant 0$,

$$\sum_{k=n}^{n+p}\frac{(-1)^{k+1}}{k}\sin kx=\sum_{k=n}^{n+p}\frac{\mathrm{U}_k(x)-\mathrm{U}_{k-1}(x)}{k}=\sum_{k=n}^{n+p}\frac{\mathrm{U}_k(x)}{k}-\sum_{k=n-1}^{n+p-1}\frac{\mathrm{U}_k(x)}{k+1}$$

ce qui conduit en regroupant les deux sommes à

$$\sum_{k=n}^{n+p}\frac{(-1)^{k+1}}{k}\sin kx=\frac{\mathrm{U}_{n+p}(x)}{n+p}-\frac{\mathrm{U}_{n-1}}{n}+\sum_{k=n}^{n+p-1}\mathrm{U}_k(x)\left(\frac{1}{k}-\frac{1}{k+1}\right).$$

Compte tenu de ce qui précède, on peut alors majorer cette tranche de Cauchy uniformément : pour tout x dans $[-\alpha,\alpha]$,

$$\begin{aligned}\left|\sum_{k=n}^{n+p}\frac{(-1)^{k+1}}{k}\sin kx\right|&\leqslant\frac{1}{\cos\dfrac{\alpha}{2}}\left(\frac{1}{n+p}+\frac{1}{n}+\sum_{k=n}^{n+p-1}\left(\frac{1}{k}-\frac{1}{k+1}\right)\right)\\&\leqslant\frac{2}{n\cos\dfrac{\alpha}{2}}.\end{aligned}$$

Le critère de Cauchy uniforme permet donc de conclure que $(s_n(x))_{n\geqslant 1}$ converge uniformément vers φ sur l'intervalle $[-\alpha,\alpha]$.

2. Posons, pour tout $n\in\mathbb{N}$ et $x\in\mathbb{R}$, $u_n(x)=\dfrac{(-1)^{n-1}\sin nx}{n(n^2-a^2)}\cdot$ La fonction u_n est définie sur $\mathbb{R}$ car $a\notin\mathbb{N}$.

On a, pour tout $x\in\mathbb{R}$, $|u_n(x)|\leqslant\dfrac{1}{n|n^2-a^2|}\cdot$ La série $\sum u_n$ qui définit f est donc normalement convergente. De plus, les fonctions u_n sont de

classe $\mathcal{C}^1$ sur $\mathbb{R}$ et pour tout $x \in \mathbb{R}$, on a $|u_n'(x)| = \left|\dfrac{\cos nx}{n^2 - a^2}\right| \leqslant \dfrac{1}{|n^2 - a^2|}$. La série $\sum u_n'$ est donc normalement convergente. On en déduit, par le théorème de dérivation des séries de fonctions, que f est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et que, pour tout réel x,

$$f'(x) = \sum_{n=1}^{+\infty} \frac{(-1)^{n-1} \cos nx}{n^2 - a^2}.$$

La série trigonométrique qui définit f étant normalement convergente, on en déduit que c'est la série de Fourier de f.

3. Pour tout $n \in \mathbb{N}^*$, u_n est de classe $\mathcal{C}^2$ et, pour tout $x \in \mathbb{R}$, on a

$$u_n''(x) = \frac{(-1)^n n \sin nx}{n^2 - a^2}.$$

Sachant que $\dfrac{n}{n^2 - a^2} - \dfrac{1}{n} = \dfrac{a^2}{n(n^2 - a^2)}$, on en déduit que, pour tout $x \in \mathbb{R}$:

$$u_n''(x) = -\frac{(-1)^{n+1}}{n} \sin nx - a^2 u_n(x).$$

D'après la première question, la série $\sum \dfrac{(-1)^{n+1}}{n} \sin nx$ est uniformément convergente sur $[-\alpha, \alpha]$, pour $\alpha \in]-\pi, \pi[$ vers φ ; la série $\sum u_n$ est normalement convergente, donc uniformément convergente sur $\mathbb{R}$ vers f. On en déduit que $\sum u_n''$ est uniformément convergente sur $[-\alpha, \alpha]$. Le théorème de dérivation des séries s'applique de nouveau : f est de classe $\mathcal{C}^2$ sur $[-\alpha, \alpha]$ et $f'' = \sum_{n=1}^{+\infty} u_n''$. Ceci étant vrai pour tout α de $[0, \pi[$, on en déduit que f est de classe $\mathcal{C}^2$ sur $]-\pi, \pi[$. De plus, on a, pour tout $x \in]-\pi, \pi[$,

$$\begin{aligned} f''(x) &= \sum_{n=1}^{+\infty} u_n''(x) = -\sum_{n=1}^{+\infty} \frac{(-1)^{n+1}}{n} \sin nx - a^2 \sum_{n=1}^{+\infty} u_n(x) \\ &= -\varphi(x) - a^2 f(x). \end{aligned}$$

La fonction f vérifie sur $]-\pi, \pi[$ l'équation différentielle $y'' + a^2 y = -\varphi$.

4. La fonction $x \longmapsto -\dfrac{x}{2a^2}$ étant solution de l'équation différentielle, on en déduit qu'il existe $(\mathrm{A}, \mathrm{B}) \in \mathbb{R}^2$ tel que, pour tout $x \in]-\pi, \pi[$,

$$f(x) = -\frac{x}{2a^2} + \mathrm{A} \cos ax + \mathrm{B} \sin ax.$$

f étant de plus continue, cette relation est en fait vérifiée sur $[-\pi, \pi]$. Or il est clair que $f(0) = f(\pi) = 0$. On obtient $\mathrm{A} = 0$, puis $\mathrm{B} = \dfrac{\pi}{2a^2 \sin a\pi}$.

Conclusion. Pour tout $x \in [-\pi, \pi]$, $f(x) = -\dfrac{x}{2a^2} + \dfrac{\pi}{2a^2}\dfrac{\sin ax}{\sin a\pi}$. $\lhd$

Dans l'exercice suivant on passe par un développement en série de Fourier pour majorer la norme infinie d'une fonction de classe $\mathcal{C}^2$ par la norme de la convergence en moyenne quadratique de sa dérivée seconde.

4.14. Inégalité entre normes

Soit f une application de classe $\mathcal{C}^2$ de $[0,1]$ dans $\mathbb{R}$ telle que $f(0) = f(1) = 0$. Montrer que $\|f\|_\infty \leqslant \dfrac{\|f''\|_2}{3\sqrt{5}}$.

(École polytechnique)

$\rhd$ **Solution.**

Soit g la fonction impaire et 2-périodique qui prolonge f à $\mathbb{R}$. Comme $f(0) = f(1) = 0$, la fonction g est continue en 0 et 1 donc sur $\mathbb{R}$. La fonction f est de classe $\mathcal{C}^1$ donc g est $\mathcal{C}^1$ par morceaux. Elle est donc égale à la somme de sa série de Fourier. En posant pour tout $n \in \mathbb{N}^*$,

$$a_n(g) = \frac{1}{2}\int_0^2 g(t)\sin(\pi nt)\,\mathrm{d}t = \int_0^1 f(t)\sin(\pi nt)\,\mathrm{d}t,$$

on obtient, pour tout $x \in [0,1]$,

$$f(x) = g(x) = \sum_{n=1}^{+\infty} a_n(g)\sin(\pi nx).$$

La fonction g'' est définie sur $\mathbb{R}\backslash\mathbb{Z}$ et se prolonge en une fonction continue par morceaux impaire sur $\mathbb{R}$ en posant $g''(k) = 0$ pour tout entier k. On sait que ses coefficients de Fourier vérifient $a_n(g'') = -(\pi n)^2 a_n(g)$. Par la formule de Parseval, on obtient

$$\begin{aligned}\|f''\|_2^2 &= \int_0^1 (f''(x))^2\,\mathrm{d}x = \frac{1}{2}\int_{-1}^1 (g''(x))^2\,\mathrm{d}x \\ &= \frac{1}{2}\sum_{n=1}^{+\infty}(\pi n)^4 a_n(g)^2 = \frac{\pi^4}{2}\sum_{n=1}^{+\infty} n^4 a_n(g)^2.\end{aligned}$$

On a pour tout $x \in [0,1]$,

$$|f(x)| \leqslant \sum_{n=1}^{+\infty}|a_n(g)| \leqslant \sum_{n=1}^{+\infty}|n^2 a_n(g)|\frac{1}{n^2} \leqslant \sqrt{\sum_{n=1}^{+\infty} n^4 a_n(g)^2}\sqrt{\sum_{n=1}^{+\infty}\frac{1}{n^4}},$$

d'après l'inégalité de Cauchy-Schwarz.

On sait que $\sum_{n=1}^{+\infty} \frac{1}{n^4} = \frac{\pi^4}{90}$ et $\sum_{n=1}^{+\infty} n^4 a_n(g)^2 = \frac{2}{\pi^4}\|f''\|_2^2$. On en déduit, pour tout $x \in [0,1]$, $|f(x)| \leqslant \frac{\|f''\|_2}{\sqrt{45}}$ et donc $\boxed{\|f\|_\infty \leqslant \frac{\|f''\|_2}{3\sqrt{5}}}$. $\lhd$

Dans l'exercice suivant, on détermine une fonction f dont la suite des coefficients de Fourier est une suite (a_n) à étudier, ce qui permet de déterminer un équivalent de a_n en utilisant les résultats de l'analyse de Fourier.

4.15. Recherche d'équivalent

Pour $n \in \mathbb{Z}$, on pose $a_n = \iint_{[0,1]^2} e^{-|x-y|} e^{2i\pi n(x-y)} \mathrm{d}x\mathrm{d}y$. Trouver un équivalent en $+\infty$ de

$$S_N = \sum_{(m,n)\in I_N} a_m a_n, \text{ où } I_N = \{(m,n) \in \mathbb{Z}^2, |m| \geqslant N \text{ ou } |n| \geqslant N\}.$$

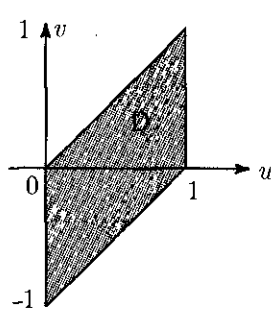
(École normale supérieure)

▷ **Solution.**

La première chose à faire est bien d'entendu d'expliciter un peu la suite a_n. La présence du terme $x - y$ dans l'intégrale double incite à faire le changement de variables suivant : on pose $v = x - y$ et $u = x$. Soit $\varphi : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ qui à (x, y) associe $\varphi(x, y) = (u, v) = (x, x - y)$. L'application φ réalise un difféomorphisme de $\mathbb{R}^2$ sur $\mathbb{R}^2$ dont le jacobien en tout point est égal à -1. Le théorème de changement de variables permet donc de dire que

$$a_n = \iint_D e^{-|v|} e^{2i\pi nv} \mathrm{d}u\mathrm{d}v,$$

où $D = \varphi([0,1]^2) = \{(u, v),\ u \in [0,1] \text{ et } u - v \in [0,1]\}$.



On a aussi $D = \{(u,v),\ u \in [0,1] \text{ et } u-1 \leqslant v \leqslant u\}$; il s'agit d'un parallélogramme (voir figure ci-dessus).

En intégrant d'abord selon u, le théorème de Fubini permet d'écrire que

$$\begin{aligned} a_n &= \int_{-1}^{0} e^{-|v|}e^{2i\pi nv}\left(\int_0^{1+v} \mathrm{d}u\right)\mathrm{d}v + \int_0^1 e^{-|v|}e^{2i\pi nv}\left(\int_v^1 \mathrm{d}u\right)\mathrm{d}v \\ &= \int_{-1}^{0} e^{v}e^{2i\pi nv}(1+v)\mathrm{d}v + \int_0^1 e^{-v}e^{2i\pi nv}(1-v)\mathrm{d}v \\ &= \int_0^1 e^{-v}e^{-2i\pi nv}(1-v)\mathrm{d}v + \int_0^1 e^{-v}e^{2i\pi nv}(1-v)\mathrm{d}v \end{aligned}$$

Et finalement, $\boxed{a_n = 2\displaystyle\int_0^1 e^{-v}\cos(2\pi nv)(1-v)\mathrm{d}v}$. $(*)$

On peut alors terminer le calcul de a_n en intégrant par parties. On obtient, pour $n \in \mathbb{N}$,

$$\begin{aligned} \int_0^1 (1-v)e^{v(-1+2i\pi n)}\mathrm{d}v &= \left[(1-v)\frac{e^{v(-1+2i\pi n)}}{-1+2i\pi n}\right]_0^1 + \int_0^1 \frac{e^{v(-1+2i\pi n)}}{-1+2i\pi n}\mathrm{d}v \\ &= \frac{1}{1-2i\pi n} + \frac{e^{-1}-1}{(1-2i\pi n)^2} \\ &= \frac{1+2i\pi n}{1+4\pi^2n^2} + \frac{(e^{-1}-1)(1-4\pi^2n^2+4i\pi n)}{(1+4\pi^2n^2)^2}. \end{aligned}$$

On en déduit que

$$\begin{aligned} a_n &= 2\,\mathrm{Re}\left(\int_0^1 (1-v)e^{v(-1+2i\pi n)}\mathrm{d}v\right) \\ &= \frac{2}{1+4\pi^2n^2} + \frac{(e^{-1}-1)(2-8\pi^2n^2)}{(1+4\pi^2n^2)^2}. \end{aligned}$$

Si on évalue l'intégrale double en intégrant d'abord selon la variable v on tombe directement sur ce résultat, mais nous allons voir que la formule $(*)$ ci-dessus est importante.

Ce calcul permet déjà de justifier l'existence de la suite S_N. En effet, de $\lim\limits_{n\to+\infty} n^2a_n = \dfrac{2}{4\pi^2} + \dfrac{(e^{-1}-1)(-8\pi^2)}{16\pi^4} = \dfrac{2-e^{-1}}{2\pi^2}$, on déduit que

$$a_n \underset{n\to\infty}{\sim} \frac{2-e^{-1}}{2\pi^2n^2}.$$

La série $\sum a_n$ est absolument convergente donc la série double $\sum a_m a_n$ est sommable. En particulier S_N existe pour tout entier N. On peut

écrire, pour $N \in \mathbb{N}^*$,

$$\begin{aligned} S_N &= \sum_{(n,m)\in\mathbb{Z}^2} a_n a_m - \sum_{\substack{|n|<N\\|m|<N}} a_n a_m = \left(\sum_{n\in\mathbb{Z}} a_n\right)^2 - \left(\sum_{|n|<N} a_n\right)^2 \\ &= \left(\sum_{n\in\mathbb{Z}} a_n + \sum_{|n|<N} a_n\right)\left(\sum_{n\in\mathbb{Z}} a_n - \sum_{|n|<N} a_n\right) \\ &= \left(\sum_{n\in\mathbb{Z}} a_n + \sum_{|n|<N} a_n\right)\left(2\sum_{n\geqslant N} a_n\right), \end{aligned}$$

car la relation $(*)$ montre que $a_{-n} = a_n$ pour tout n. Comme on a $\lim_{N\to+\infty} \sum_{n\in\mathbb{Z}} a_n + \sum_{|n|<N} a_n = 2\sum_{n\in\mathbb{Z}} a_n$, on en tire que

$$S_N \underset{N\to\infty}{\sim} 4\sum_{n\in\mathbb{Z}} a_n \sum_{n\geqslant N} a_n,$$

si du moins $\sum_{n\in\mathbb{Z}} a_n$ n'est pas nul. Il nous faut donc encore faire deux choses : trouver la valeur de $\sum_{n\in\mathbb{Z}} a_n$ puis déterminer un équivalent du reste $\sum_{n\geqslant N} a_n$.

• On va commencer par calculer la valeur de la somme $\sum_{n\in\mathbb{Z}} a_n$. Pour cela on utilise la relation $(*)$. Considérons la fonction 1-périodique f définie par $f(x) = e^{-x}(1-x)$ si $x \in [0,1[$. D'après $(*)$, a_n est un coefficient de Fourier trigonométrique de f. La fonction f est $\mathcal{C}^1$ par morceaux. On en déduit que sa série de Fourier converge en tout point vers $\frac{1}{2}(f(x+0)+f(x-0))$. On a, pour tout $x \in \mathbb{R}$,

$$\frac{1}{2}(f(x+0)+f(x-0)) = \frac{a_0}{2} + \sum_{n=1}^{+\infty}(a_n\cos 2\pi nx + b_n \sin 2\pi nx).$$

En prenant $x = 0$, on obtient $\frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n = \frac{1}{2}(1+0) = \frac{1}{2}$. Puisque $a_n = a_{-n}$ pour tout entier n, on en déduit que

$$\sum_{n\in\mathbb{Z}} a_n = 2\left(\frac{a_0}{2} + \sum_{n=1}^{+\infty} a_n\right) = 1.$$

• On termine enfin la recherche de l'équivalent demandé. On a vu précédemment que

$$a_n \underset{n\to\infty}{\sim} \frac{2-e^{-1}}{2\pi^2 n^2} \underset{n\to\infty}{\sim} \frac{2-e^{-1}}{2\pi^2}\left(\frac{1}{n-1} - \frac{1}{n}\right).$$

Le théorème de sommation des relations d'équivalence dans le cas de séries convergentes à termes positifs donne, pour $N \in \mathbb{N}^*$,

$$\sum_{n\geqslant N} a_n \underset{N\to\infty}{\sim} \frac{2-e^{-1}}{2\pi^2} \sum_{n\geqslant N}\left(\frac{1}{n-1}-\frac{1}{n}\right) = \frac{(2-e^{-1})}{2\pi^2(N-1)}$$

et on peut conclure finalement que

$$\boxed{S_N \underset{N\to\infty}{\sim} \frac{2(2-e^{-1})}{\pi^2 N}}. \quad \triangleleft$$

4.16. Formule sommatoire de Poisson

Soit f une application de classe $\mathcal{C}^\infty$ de $\mathbb{R}$ dans $\mathbb{C}$ telle que, pour tout $(n,m) \in \mathbb{N}^2$, $\lim\limits_{x\to\pm\infty} x^n f^{(m)}(x) = 0$. On pose

$$\varphi(x) = \sum_{k\in\mathbb{Z}} f(x+2k\pi).$$

1. Que dire de φ ? Calculer les coefficients de Fourier de φ.
2. Montrer que

$$\varphi(x) = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} \hat{f}(n)e^{inx},$$

où $\hat{f}(n) = \int_{\mathbb{R}} f(t)e^{-int}dt$.

(École polytechnique)

▷ **Solution.**

1. Montrons que φ est définie et $\mathcal{C}^\infty$ sur $\mathbb{R}$. Pour commencer, montrons que la série définissant φ converge uniformément sur tout intervalle compact $[-\alpha,\alpha]$ $(\alpha > 0)$. De l'hypothèse $\lim\limits_{x\to\pm\infty} x^2 f(x) = 0$, on déduit l'existence de $a > 0$ tel que $|x^2 f(x)| \leqslant 1$ si $|x| \geqslant a$. Pour $|x| \leqslant \alpha$ et $k \in \mathbb{Z}$, on a $|x+2k\pi| \geqslant |2k\pi| - |x| \geqslant |2k\pi| - \alpha$ et donc $|x+2k\pi| \geqslant a$ si $|k| \geqslant \dfrac{\alpha+a}{2\pi}$. On en déduit, pour $|k| \geqslant \dfrac{\alpha+a}{2\pi}$,

$$|f(x+2k\pi)| \leqslant \frac{1}{(x+2k\pi)^2} \leqslant \frac{1}{(2|k|\pi-\alpha)^2}.$$

La série $\sum \dfrac{1}{(2|k|\pi-\alpha)^2}$ étant convergente, on en déduit que la série $\sum f(x+2k\pi)$ converge uniformément sur $[-\alpha,\alpha]$. La fonction φ est donc définie sur $\mathbb{R}$.

Pour tout $n \in \mathbb{N}$, la fonction $f^{(n)}$ vérifie la même hypothèse que f. On en déduit que la série $\sum f^{(n)}(x+2k\pi)$ converge uniformément sur tout intervalle $[-\alpha,\alpha]$. Le théorème de dérivation des séries de fonctions montre que, pour tout $\alpha > 0$, la fonction f est $\mathcal{C}^\infty$ sur $[-\alpha,\alpha]$ et que les dérivées successives de φ sont obtenues en dérivant terme à terme. Finalement f est $\mathcal{C}^\infty$ et, pour $n \in \mathbb{N}$ et $x \in \mathbb{R}$, on a

$$\varphi^{(n)}(x) = \sum_{k\in\mathbb{Z}} f^{(n)}(x+2k\pi).$$

La fonction φ est clairement 2π-périodique. Notons c_n $(n \in \mathbb{Z})$ ses coefficients de Fourier. La série définissant φ convergeant uniformément sur $[0,2\pi]$, on peut intervertir la sommation et l'intégration et écrire, pour $n \in \mathbb{Z}$,

$$c_n = \frac{1}{2\pi}\int_0^{2\pi} \varphi(x)e^{-inx}\mathrm{d}x = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \int_0^{2\pi} f(x+2k\pi)e^{-inx}\mathrm{d}x.$$

On obtient ensuite

$$c_n = \frac{1}{2\pi}\sum_{k\in\mathbb{Z}} \int_{2k\pi}^{2(k+1)\pi} f(x)e^{-inx}\mathrm{d}x = \frac{1}{2\pi}\int_{\mathbb{R}} f(x)e^{-inx}\mathrm{d}x,$$

cette dernière intégrale ayant un sens car $|f(x)e^{-inx}| = |f(x)| = o\left(\dfrac{1}{x^2}\right)$ au voisinage de $\pm\infty$, par hypothèse.

2. La fonction φ étant $\mathcal{C}^\infty$ donc $\mathcal{C}^1$ sur $\mathbb{R}$, elle est égale à la somme de sa série de Fourier en tout point, et celle-ci est normalement convergente. On a donc, pour tout réel x,

$$\varphi(x) = \sum_{n\in\mathbb{Z}} c_n e^{inx} = \sum_{n\in\mathbb{Z}} \left(\frac{1}{2\pi}\int_{\mathbb{R}} f(t)e^{-int}\mathrm{d}t\right) e^{inx}. \quad \triangleleft$$

En particulier pour $x = 0$, on obtient l'égalité

$$\varphi(0) = \sum_{k\in\mathbb{Z}} f(2k\pi) = \frac{1}{2\pi}\sum_{n\in\mathbb{Z}} \hat{f}(n).$$

Cette formule qui relie les valeurs de f aux points $2k\pi$ aux valeurs de sa transformée de Fourier aux points entiers est usuellement appelée formule sommatoire de Poisson. Indiquons qu'une fonction f vérifiant comme dans l'énoncé $\lim\limits_{x\to\pm\infty} x^n f^{(m)}(x) = 0$ pour tout $(n,m) \in \mathbb{N}^2$ est dite à décroissance rapide. La formule de Poisson reste valide avec des hypothèses plus faibles sur f.

L'exercice suivant fait redémontrer cette formule sommatoire de Poisson dans un cas particulier avec une fonction 1-périodique. La formule obtenue dans ce cas est plus jolie car les 2π disparaissent.

4.17. Un cas particulier de la formule sommatoire de Poisson

On pose pour $t > 0$ fixé $f(x) = \sum_{k=-\infty}^{+\infty} \exp\left(-\frac{(k-x)^2}{2t}\right)$.

1. Montrer que f est définie et de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et 1-périodique.
2. Déterminer ses coefficients de Fourier.
3. On pose, pour $t > 0$, $\theta(t) = \sum_{k=-\infty}^{+\infty} \exp\left(-\pi \frac{k^2}{t}\right)$. Déterminer une relation entre $\theta(t)$ et $\theta\left(\frac{1}{t}\right)$.

(École polytechnique)

▷ **Solution.**

1. Montrons que la série définissant $f(x)$ converge normalement sur tout intervalle $[-a, a]$ $(a > 0)$. On a, pour $x \in [-a, a]$ et $k \in \mathbb{Z}$,

$$e^{-\frac{(k-x)^2}{2t}} = e^{\frac{-k^2-x^2+2kx}{2t}} \leqslant \exp\left(\frac{-k^2+2|k|a}{2t}\right) = o\left(\frac{1}{k^2}\right),$$

ce qui démontre le résultat annoncé. La fonction f est donc définie sur $[-a, a]$ et continue sur cette intervalle, car somme d'une série de fonctions continues qui converge uniformément.

De même, la série dérivée $\sum \frac{k-x}{t} \exp\left(-\frac{(k-x)^2}{2t}\right)$ converge normalement sur $[-a, a]$. En effet, on a

$$\frac{|k-x|}{t} e^{-\frac{(k-x)^2}{2t}} \leqslant \frac{|k|+a}{t} e^{\frac{-k^2+2|k|a}{2t}}$$

et la série $\sum \frac{|k|+a}{t} e^{\frac{-k^2+2|k|a}{2t}}$ converge. On en déduit que la fonction f est de classe $\mathcal{C}^1$ sur $[-a, a]$ et que sa dérivée s'obtient en dérivant terme à terme la série donnant $f(x)$. Ceci étant vrai pour tout $a > 0$, la fonction f est $\mathcal{C}^1$ sur $\mathbb{R}$.

On a, pour tout réel x,

$$\begin{aligned} f(x+1) &= \lim_{n\to+\infty} \sum_{k=-n}^{n} \exp\left(-\frac{(k-x-1)^2}{2t}\right) \\ &= \lim_{n\to+\infty} \sum_{k=-n-1}^{n-1} \exp\left(-\frac{(k-x)^2}{2t}\right) = f(x). \end{aligned}$$

La fonction f est 1-périodique.

2. Calculons les coefficients de Fourier de f. On a pour $n \in \mathbb{Z}$,

$$c_n(f) = \int_0^1 f(x)e^{-2i\pi nx}\mathrm{d}x = \int_0^1 \sum_{k\in\mathbb{Z}} \exp\left(-\frac{(k-x)^2}{2t}\right) e^{-2i\pi nx}\mathrm{d}x$$
$$= \sum_{k\in\mathbb{Z}} \int_0^1 \exp\left(-\frac{(k-x)^2}{2t}\right) e^{-2i\pi nx}\mathrm{d}x,$$

l'interversion de la sommation et de l'intégration étant licite puisque la série converge normalement sur $[0,1]$. On remarque alors, en effectuant le changement de variable $y = x - k$, que

$$\int_0^1 \exp\left(-\frac{(k-x)^2}{2t}\right) e^{-2i\pi nx}\mathrm{d}x = \int_{-k}^{-k+1} e^{-\frac{y^2}{2t}} e^{-2i\pi ny}\mathrm{d}y,$$

d'où l'on déduit que $c_n(f) = \int_{\mathbb{R}} e^{-\frac{x^2}{2t}} e^{-2i\pi nx}\mathrm{d}x$.

Pour déterminer ces coefficients de Fourier, nous allons calculer. plus généralement, $\varphi(u) = \int_{\mathbb{R}} e^{-\frac{x^2}{2t}} e^{iux}\mathrm{d}x$ pour tout $u \in \mathbb{R}$.

La fonction $g : (u,x) \longmapsto e^{-\frac{x^2}{2t}} e^{iux}$ est continue sur $\mathbb{R}^2$ et on a, pour tout $(u,x) \in \mathbb{R}^2$, $|g(u,x)| \leqslant e^{-\frac{x^2}{2t}}$. La fonction $x \longmapsto g(u,x)$ est donc intégrable sur $\mathbb{R}$ pour tout réel u et vérifie une hypothèse de domination. Le théorème de continuité sous le signe $\int$ permet de conclure que φ est continue sur $\mathbb{R}$.

De plus, g admet une dérivée partielle par rapport à u et pour $(u,x) \in \mathbb{R}^2$, on a $\left|\frac{\partial g(u,x)}{\partial u}\right| = \left|ixe^{-\frac{x^2}{2t}} e^{iux}\right| = |x|e^{-\frac{x^2}{2t}}$. La fonction $x \longmapsto \frac{\partial g(u,x)}{\partial u}$ est donc intégrable, pour tout $u \in \mathbb{R}$ et vérifie une hypothèse de domination. On en déduit, d'après le théorème de dérivation sous le signe $\int$ que φ est de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et que

$$\varphi'(u) = \int_{\mathbb{R}} ixe^{-\frac{x^2}{2t}} e^{iux}\mathrm{d}x.$$

On peut intégrer par parties, en intégrant $x \longmapsto ixe^{-\frac{x^2}{2t}}$ et en dérivant $x \longmapsto e^{iux}$. La fonction $x \longmapsto -ite^{-\frac{x^2}{2t}} e^{iux}$ ayant une limite nulle en $-\infty$ et $+\infty$, on obtient

$$\varphi'(u) = -\int_{\mathbb{R}} \left(-ite^{-\frac{x^2}{2t}}\right)\left(iue^{iux}\right)\mathrm{d}x = -ut\varphi(u).$$

La fonction φ vérifie l'équation différentielle $\varphi'(u) = -ut\varphi(u)$. Il existe donc une constante C telle que, pour tout $u \in \mathbb{R}$, $\varphi(u) = \mathrm{C}e^{-\frac{u^2t}{2}}$. On calcule C en remarquant que

$$\mathrm{C} = \varphi(0) = \int_{\mathbb{R}} e^{-\frac{x^2}{2t}}\mathrm{d}x = \sqrt{2t}\int_{\mathbb{R}} e^{-x^2}\mathrm{d}x = \sqrt{2\pi t}.$$

d'après un résultat classique. On a donc $\varphi(u) = \sqrt{2\pi t}\, e^{-\frac{u^2 t}{2}}$ et en particulier, pour tout $n \in \mathbb{Z}$,

$$c_n(f) = \varphi(-2\pi n) = \sqrt{2\pi t}\, e^{-2\pi^2 n^2 t}.$$

3. La fonction f étant de classe $\mathcal{C}^1$, sa série de Fourier converge sur $\mathbb{R}$ vers f. En particulier, on a, pour tout $t > 0$,

$$\sum_{k\in\mathbb{Z}} e^{-\frac{k^2}{2t}} = f(0) = \sqrt{2\pi t} \sum_{n\in\mathbb{Z}} e^{-2\pi^2 n^2 t}.$$

En remplaçant t par $\dfrac{t}{2\pi}$, on obtient, pour tout $t > 0$,

$$\sum_{k\in\mathbb{Z}} e^{-\pi\frac{k^2}{t}} = \sqrt{t} \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t},$$

c'est-à-dire

$$\boxed{\theta(t) = \sqrt{t}\,\theta\left(\frac{1}{t}\right)}. \quad \lhd$$

La fonction θ étudiée dans l'exercice est reliée à la fonction Θ de Jacobi, définie par $\Theta(x) = \sum_{n\in\mathbb{Z}} x^{n^2}$: on a en effet $\theta(t) = \Theta\left(e^{-\frac{\pi}{t}}\right)$. Introduite par Euler pour l'étude de problèmes de partitions de nombre entiers, la fonction Θ fut étudiée par Jacobi. En 1828, il découvrit la formule suivante : $\left(\sum_{n\in\mathbb{Z}} x^{n^2}\right)^4 = 1 + 8\sum_{m=1}^{+\infty}\left(\sum_{\substack{d\mid m\\ 4\nmid d}} d\right) x^m$. En observant le coefficient de x^n dans les deux membres de cette égalité, il en déduisit une expression du nombre de représentations d'un entier naturel n comme somme de quatre carrés.

La relation démontrée dans la question 3 de l'exercice relie le comportement de Θ en 1 à son comportement en 0.

Dans l'exercice suivant, on cherche un développement en série entière à partir d'un développement en série de Fourier.

4.18. Série génératrice des nombres de Bernoulli

Soit φ la fonction 2π-périodique définie par $\varphi(x) = \exp\left(\dfrac{zx}{2\pi}\right)$ pour $x \in]-\pi, \pi]$, où $z \in \mathbb{C} \setminus 2i\pi\mathbb{Z}$.

Du développement en série de Fourier de φ, déduire le développement en série entière en 0 de la fonction $f : z \longmapsto \dfrac{z}{e^z - 1}$.

(École polytechnique)

▷ **Solution.**

La fonction φ est continue par morceaux sur $\mathbb{R}$. On peut calculer ses coefficients de Fourier. On obtient, pour $n \in \mathbb{Z}^*$,

$$c_n(\varphi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \varphi(x)e^{-inx}\mathrm{d}x = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{\frac{zx}{2\pi}}e^{-inx}\mathrm{d}x$$
$$= \frac{1}{2\pi}\cdot\frac{1}{\frac{z}{2\pi} - ni}\left[e^{\frac{zx}{2\pi}}e^{-inx}\right]_{-\pi}^{\pi} = \frac{(-1)^n(e^{\frac{z}{2}} - e^{-\frac{z}{2}})}{z - 2i\pi n}.$$

La fonction φ est $\mathcal{C}^1$ par morceaux sur $\mathbb{R}$. Le théorème de Dirichlet permet d'affirmer que sa série de Fourier converge en tout point x de $\mathbb{R}$ vers $\frac{1}{2}(\varphi(x+0) + \varphi(x-0))$. On a donc, pour tout $x \in \mathbb{R}$,

$$\frac{1}{2}(\varphi(x+0) + \varphi(x-0)) = (e^{\frac{z}{2}} - e^{\frac{-z}{2}})\sum_{n\in\mathbb{Z}} \frac{(-1)^n e^{inx}}{z - 2i\pi n}.$$

Afin d'en déduire le développement en série entière de la fonction $f : z \longmapsto \dfrac{z}{e^z - 1}$, on va choisir une valeur de x à laquelle appliquer cette formule. On songe naturellement à 0 ou à π. Pour 0, on ne voit pas apparaître $f(z)$. Pour $x = \pi$, on a $\frac{1}{2}(\varphi(x+0)+\varphi(x-0)) = \frac{1}{2}(e^{\frac{z}{2}} + e^{\frac{-z}{2}})$ et on obtient,

$$\frac{1}{2}(e^{\frac{z}{2}} + e^{\frac{-z}{2}}) = (e^{\frac{z}{2}} - e^{\frac{-z}{2}})\sum_{n\in\mathbb{Z}} \frac{1}{z - 2i\pi n} = (e^{\frac{z}{2}} - e^{\frac{-z}{2}})\sum_{n\in\mathbb{Z}} \frac{z + 2i\pi n}{z^2 + 4\pi^2 n^2}.$$

En divisant par $e^{\frac{z}{2}} - e^{-\frac{z}{2}}$ et en regroupant les termes correspondant à deux valeurs opposées de n, l'égalité devient

$$\frac{e^{\frac{z}{2}} + e^{-\frac{z}{2}}}{2(e^{\frac{z}{2}} - e^{-\frac{z}{2}})} = \frac{1}{z} + 2\sum_{n=1}^{+\infty} \frac{z}{z^2 + 4\pi^2 n^2}.$$

On transforme le premier membre. Pour cela, on écrit

$$\frac{e^{\frac{z}{2}} + e^{-\frac{z}{2}}}{2(e^{\frac{z}{2}} - e^{-\frac{z}{2}})} = \frac{e^z + 1}{2(e^z - 1)} = \frac{1}{2} + \frac{1}{(e^z - 1)}.$$

En multipliant notre égalité par z et en soustrayant $\frac{z}{2}$, on obtient finalement, pour tout $z \in \mathbb{C} \setminus 2i\pi\mathbb{Z}$,

$$f(z) = 1 - \frac{z}{2} + 2\sum_{n=1}^{+\infty} \frac{z^2}{z^2 + 4\pi^2 n^2}.$$

Il reste à développer cette somme en série entière. Pour cela, on développe en série entière $\dfrac{z^2}{z^2 + 4\pi^2 n^2}$. On a, pour $|z| < 2\pi$ et $n \in \mathbb{N}^*$, $\left|\dfrac{z}{2\pi n}\right| < 1$ et donc

$$\frac{z^2}{z^2 + 4\pi^2 n^2} = \frac{z^2}{4\pi^2 n^2} \frac{1}{1 + \dfrac{z^2}{4\pi^2 n^2}} = \frac{z^2}{4\pi^2 n^2} \sum_{k=0}^{+\infty} (-1)^k \frac{z^{2k}}{(4\pi^2 n^2)^k}$$
$$= \sum_{k=0}^{+\infty} (-1)^k \frac{z^{2(k+1)}}{(4\pi^2 n^2)^{k+1}} = \sum_{k=1}^{+\infty} (-1)^{k-1} \frac{z^{2k}}{(2\pi n)^{2k}}.$$

Considérons la série double $\sum u_{n,k}$ avec $u_{n,k} = (-1)^{k-1} \dfrac{z^{2k}}{(2\pi n)^{2k}}$ pour $(n,k) \in (\mathbb{N}^*)^2$. Pour tout $n \in \mathbb{N}^*$, $\sum\limits_k |u_{n,k}|$ converge et

$$\sum_{k \in \mathbb{N}^*} |u_{n,k}| = \sum_{k \in \mathbb{N}^*} \frac{|z|^{2k}}{(2\pi n)^{2k}} = \frac{|z|^2}{4\pi^2 n^2} \frac{1}{1 - \dfrac{|z|^2}{4\pi^2 n^2}} = \frac{|z|^2}{4\pi^2 n^2 - |z|^2}.$$

La série $\sum\limits_n \dfrac{|z|^2}{4\pi^2 n^2 - |z|^2}$ converge. On déduit que la série double est sommable. On sait qu'alors on peut échanger l'ordre des sommations et on obtient, pour $|z| < 2\pi$, $z \neq 0$,

$$f(z) = 1 - \frac{z}{2} + 2 \sum_{n \in \mathbb{N}^*} \sum_{k \in \mathbb{N}^*} u_{n,k} = 1 - \frac{z}{2} + 2 \sum_{k \in \mathbb{N}^*} \sum_{n \in \mathbb{N}^*} u_{n,k}$$
$$= 1 - \frac{z}{2} + 2 \sum_{k \in \mathbb{N}^*} \frac{(-1)^{k-1}}{(2\pi)^{2k}} \left(\sum_{n \in \mathbb{N}^*} \frac{1}{n^{2k}} \right) z^{2k}.$$

Remarquons que f admet un prolongement par continuité en 0 défini par $f(0) = 1$ et que l'égalité est encore vérifiée pour $z = 0$. On obtient donc, pour tout $|z| < 2\pi$,

$$\boxed{f(z) = 1 - \frac{z}{2} + 2 \sum_{k \in \mathbb{N}^*} \frac{(-1)^{k-1}}{(2\pi)^{2k}} \left(\sum_{n \in \mathbb{N}^*} \frac{1}{n^{2k}} \right) z^{2k}.}$$

C'est le développement cherché. Il est valable pour $|z| < 2\pi$. On ne pouvait pas espérer mieux, puisque f n'est pas définie en $2i\pi$. ◁

Les nombres de Bernoulli b_n sont définis par $f(z) = \sum\limits_{n=0}^{+\infty} \dfrac{b_n}{n!} z^n$. On a, d'après l'exercice, $b_0 = 1$, $b_1 = -\dfrac{1}{2}$ et $b_{2n+1} = 0$ pour $n \geqslant 1$.

En considérant le produit de $f(z)$ *par* $e^z - 1$, *qui est égal à* z, *on obtient, par un produit de Cauchy, pour* $n \geqslant 2$, $0 = \sum_{k=0}^{n-1} \frac{b_k}{k!} \cdot \frac{1}{(n-k)!}$, *soit* $\sum_{k=0}^{n-1} C_n^k b_k = 0$. *On a donc, pour* $n \geqslant 2$, $b_{n-1} = -\frac{1}{n} \sum_{k=2}^{n} C_n^k b_k$. *Par une récurrence immédiate, on en déduit que les* b_k *appartiennent à* $\mathbb{Q}$. *Du développement en série obtenu dans l'exercice on tire, pour* $k \geqslant 1$, $\frac{b_{2n}}{(2n)!} = \frac{(-1)^{k-1}}{(2\pi)^{2k}} \sum_{n \in \mathbb{N}^*} \frac{1}{n^{2k}}$ *et donc* $\sum_{n \in \mathbb{N}^*} \frac{1}{n^{2k}} = (-1)^{k-1} \frac{(2\pi)^{2k}}{2(2n)!} b_{2k}$ *(*formule d'Euler*). La relation de récurrence précédente fournit* $b_2 = \frac{1}{6}$, $b_4 = -\frac{1}{30}$, $b_6 = \frac{1}{42}$,... *et donc*

$$\sum_{n \in \mathbb{N}^*} \frac{1}{n^2} = \frac{\pi^2}{6}, \quad \sum_{n \in \mathbb{N}^*} \frac{1}{n^4} = \frac{\pi^4}{90}, \quad \sum_{n \in \mathbb{N}^*} \frac{1}{n^6} = \frac{\pi^6}{945}, \ldots$$

Notons que l'on ne sait par contre quasiment rien de $\sum_{n=1}^{+\infty} \frac{1}{n^k}$ *si* k *est impair. L'un des seuls résultats significatifs connus est que* $\sum_{n=1}^{+\infty} \frac{1}{n^3}$ *est irrationnel (résultat démontré par Apéry en 1978).*

Si la fonction f *est* 2π*-périodique et de classe* $\mathcal{C}^1$, *une simple intégration par parties montre que* $c_n(f') = in\, c_n(f)$, *pour tout* $n \in \mathbb{Z}$. *On voit, dans la solution de l'exercice suivant, que cela reste vrai lorsque* f *est continue et seulement* $\mathcal{C}^1$ *par morceaux. Plus généralement, si* f *est* $\mathcal{C}^{k-1}$ *et* $\mathcal{C}^k$ *par morceaux, alors on a* $c_n(f^{(k)}) = (in)^k c_n(f)$. *On en déduit notamment que* $c_n(f) = o\left(\frac{1}{n^k}\right)$, *puisque* $c_n(f^{(k)})$ *tend vers* 0 *quand* n *tend vers* $+\infty$.

4.19. Série de Fourier d'une primitive

Soit f une fonction continue par morceaux, 2π-périodique de $\mathbb{R}$ dans $\mathbb{C}$. Pour tous $n \in \mathbb{N}^*$ et $x \in \mathbb{R}$, on pose $S_n(x) = \sum_{1 \leqslant |k| \leqslant n} \frac{c_k}{k} e^{ikx}$, où les c_k sont les coefficients de Fourier de f. Étudier la convergence de la suite de fonctions (S_n).

(École polytechnique, école normale supérieure)

▷ **Solution.**

En dérivant S_n, on obtient $S_n'(x) = \sum_{1 \leqslant |k| \leqslant n} ic_k e^{ikx}$. Au facteur i près, c'est une somme partielle de la série de Fourier de $f - c_0$. Il est donc assez naturel de considérer une primitive de $f - c_0$. Soit F la fonction définie par $F(x) = \int_{-\pi}^{x} (f(t) - c_0)\,dt$. On a, pour tout x,

$$\begin{aligned} F(x + 2\pi) &= F(x) + \int_{x}^{x+2\pi} (f(t) - c_0)\,dt = F(x) + \int_{-\pi}^{\pi} (f(t) - c_0)\,dt \\ &= F(x) + \int_{-\pi}^{\pi} f(t)\,dt - 2\pi c_0 = F(x). \end{aligned}$$

La fonction F est donc 2π-périodique. Elle est lipschitzienne car la fonction $f - c_0$ est bornée. Donc en particulier F est continue sur $\mathbb{R}$. Elle est aussi dérivable en tout point x où f est continue et en un tel point $F'(x) = f(x) - c_0$. Comme f est continue par morceaux, la fonction F est donc continue et $\mathcal{C}^1$ par morceaux. D'après le théorème de Dirichlet, sa série de Fourier converge uniformément vers F.

La fonction F n'étant pas dérivable sur $\mathbb{R}$, on ne peut pas directement utiliser les relations entre $c_n(F)$ et $c_n(F')$. Si $]a.b[$ est un intervalle sur lequel f est continue et $n \in \mathbb{N}^*$, on a

$$\int_a^b F(t)e^{-int}\mathrm{d}x = \left[\frac{F(t)}{-in}\right]_a^b + \frac{1}{in}\int_a^b (f(t) - c_0)e^{-int}\mathrm{d}t.$$

Si $a_1, a_2, \ldots, a_p$ sont les points de discontinuité de f sur $[-\pi, \pi]$, on obtient, en posant $a_0 = -\pi$ et $a_{p+1} = \pi$,

$$\begin{aligned} c_n(F) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} F(t)e^{-int}\mathrm{d}x = \frac{1}{2\pi}\sum_{j=0}^{p}\int_{a_j}^{a_{j+1}} F(t)e^{-int}\mathrm{d}x \\ &= \frac{1}{2\pi}\sum_{j=0}^{p}\left(\left[\frac{F(t)}{-in}e^{-int}\right]_{a_j}^{a_{j+1}} + \frac{1}{in}\int_{a_j}^{a_{j+1}} (f(t) - c_0)e^{-int}\mathrm{d}t\right) \\ &= \frac{1}{-2i\pi n}(F(\pi) - F(-\pi)) + \frac{1}{2i\pi n}\int_{-\pi}^{\pi} (f(t) - c_0)e^{-int}\mathrm{d}t \\ &= \frac{1}{in}c_n(f), \end{aligned}$$

car $\int_0^{2\pi} c_0 e^{-int}\,\mathrm{d}t = 0$ (puisque $n \neq 0$).

On obtient de même

$$\begin{aligned} c_0(F) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} F(t)\,\mathrm{d}t = \frac{1}{2\pi}\left[tF(t)\right]_{-\pi}^{\pi} - \frac{1}{2\pi}\int_{-\pi}^{\pi} t(f(t) - c_0)\,\mathrm{d}t \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} tf(t)\,\mathrm{d}t, \end{aligned}$$

car $F(\pi) = F(-\pi) = 0$.

On a, pour $n \in \mathbb{N}^*$, $\sum_{k=-n}^{n} c_n(F)e^{ikx} = c_0(F) + \frac{1}{i}S_n(x)$. La série de Fourier de F converge vers F. On en déduit que, pour tout $x \in \mathbb{R}$, $(S_n(x))$ converge vers

$$i(F(x) - c_0(F)) = i\left(\int_{-\pi}^{x} (f(t) - c_0)\,dt + \frac{1}{2\pi}\int_{-\pi}^{\pi} tf(t)\,dt\right).$$

La convergence de la série de Fourier de F étant uniforme sur $\mathbb{R}$, la convergence de la suite de fonctions (S_n) est également uniforme sur $\mathbb{R}$. ◁

Il résulte en particulier de l'exercice que si les c_n sont les coefficients de Fourier d'une fonction 2π-périodique, continue par morceaux, la suite de terme général $\sum_{1\leqslant|k|\leqslant n} \frac{c_k}{k} = \sum_{k=1}^{n} \frac{c_k - c_{-k}}{k} = -i\sum_{k=1}^{n} \frac{b_k}{k}$ converge. Ainsi, la série $\sum \frac{b_n}{n}$ converge. On en déduit, par exemple, que la série trigonométrique $\sum \frac{\sin nx}{\ln n}$, qui converge sur $\mathbb{R}$ d'après l'exercice 2.7, n'est pas la série de Fourier d'une fonction continue par morceaux, puisque $\sum \frac{1}{n\ln n}$ diverge. Bien entendu cela découle aussi de la formule de Parseval puisque la série $\sum \frac{1}{(\ln n)^2}$ diverge.

L'énoncé suivant donne une nouvelle preuve de l'inégalité de Wirtinger déjà rencontrée dans l'exercice 1.15.

4.20. Inégalité de Wirtinger

1. Soit $f : \mathbb{R} \longrightarrow \mathbb{R}$, 2π-périodique continue et de classe $\mathcal{C}^1$ par morceaux. Démontrer *l'inégalité de Wirtinger* :

$$\int_0^{2\pi} f^2 - \frac{1}{2\pi}\left(\int_0^{2\pi} f\right)^2 \leqslant \int_0^{2\pi} f'^2.$$

2. Soit $g \in \mathcal{C}^2([0,\pi],\mathbb{R})$ non nulle, telle que $g(0) = g(\pi) = 0$. Montrer que

$$\inf\left\{\frac{g''(x)}{g(x)},\ x \in]0,\pi[,\ g(x) \neq 0\right\} \leqslant -1.$$

(École normale supérieure)

▷ **Solution.**

1. L'idée est d'utiliser la formule de Parseval pour les fonctions f et f'. On sait que

$$\int_0^{2\pi} f^2 = 2\pi \sum_{n=-\infty}^{+\infty} |c_n(f)|^2.$$

Par ailleurs, on sait que $c_n(f') = in\, c_n(f)$ pour tout $n \in \mathbb{Z}$. La formule de Parseval pour f' permet donc d'écrire que

$$\begin{aligned}\int_0^{2\pi} f'^2 &= 2\pi \sum_{n=-\infty}^{+\infty} |inc_n(f)|^2 = 2\pi \sum_{n\in\mathbb{Z}^*} n^2 |c_n(f)|^2 \\ &\geqslant 2\pi \sum_{n\in\mathbb{Z}^*} |c_n(f)|^2 = \int_0^{2\pi} f^2 - 2\pi |c_0(f)|^2.\end{aligned}$$

C'est le résultat demandé puisque $c_0(f) = \dfrac{1}{2\pi}\displaystyle\int_0^{2\pi} f$. L'étude du cas d'égalité est facile. Il y a égalité si et seulement si $c_n(f) = 0$ pour $|n| > 1$. Et $f(x)$ est alors de la forme $\mathrm{C} + \mathrm{A}\cos x + \mathrm{B}\sin x$ avec $(\mathrm{A}, \mathrm{B}, \mathrm{C}) \in \mathbb{R}^3$.

2. Raisonnons par l'absurde et supposons que

$$\inf\left\{\frac{g''(x)}{g(x)},\ x \in]0,\pi[,\ g(x) \neq 0\right\} > -1.$$

On a donc, pour tout $x \in [0,\pi]$ tel que $g(x) \neq 0$, $\dfrac{g''(x)}{g(x)} > -1$ et, en multipliant par $g^2(x)$, $g''(x)g(x) > -g(x)^2$. On en déduit que, pour tout $x \in [0,\pi]$, $g''(x)g(x) \geqslant -g(x)^2$, avec égalité pour les x où g s'annule. Puisque les fonctions g et g'' sont continues et que g n'est pas nulle, on en déduit que $\displaystyle\int_0^{\pi} g''(x)g(x)\mathrm{d}x > -\int_0^{\pi} g^2(x)\mathrm{d}x$. En intégrant par parties, on obtient

$$\int_0^{\pi} g''(x)g(x)\mathrm{d}x = \underbrace{\big[g'(x)g(x)\big]_0^{\pi}}_{=0} - \int_0^{\pi} g'^2(x)\mathrm{d}x.$$

On a donc l'inégalité $\displaystyle\int_0^{\pi} g'^2(x)\mathrm{d}x < \int_0^{\pi} g^2(x)\mathrm{d}x$. Considérons la fonction h impaire et 2π-périodique dont la restriction à $[0,\pi]$ est g. La fonction h est continue sur $\mathbb{R}$, car $g(0) = g(\pi) = 0$ et $\mathcal{C}^1$ par morceaux. Par ailleurs $\displaystyle\int_0^{2\pi} h(x)\mathrm{d}x = 0$, puisque h est impaire. De l'inégalité de Wirtinger, on déduit $\displaystyle\int_0^{2\pi} h^2(x)\mathrm{d}x \leqslant \int_0^{2\pi} h'^2(x)\mathrm{d}x$. Or, ceci est en contradiction avec ce qui précède puisque $\displaystyle\int_0^{2\pi} h'^2 = 2\int_0^{\pi} g'^2$ et $\displaystyle\int_0^{2\pi} h^2 = 2\int_0^{\pi} g^2$. ◁

L'exercice suivant a pour objet une généralisation de l'inégalité de Wirtinger qu'on utilise pour établir une minoration de la période d'une solution périodique non constante d'une équation différentielle autonome $x' = \mathrm{F}(x)$ lorsque F *est une application lipschitzienne. La seconde question reprend le même problème mais dans le cas discret.*

4.21. Solutions périodiques de $x' = \mathrm{F}(x)$

1. Soit f une application continue par morceaux T-périodique de $\mathbb{R}$ dans $\mathbb{C}$. Montrer que

$$\iint_{[0,\mathrm{T}]^2} |f(u) - f(v)|^2 \mathrm{d}u\mathrm{d}v = 2 \sum_{k \in \widetilde{\mathbb{Z}}^*} \left| \int_0^{\mathrm{T}} f(t) e^{-i\frac{2k\pi t}{\mathrm{T}}} \mathrm{d}t \right|^2 .$$

Soit f de classe $\mathcal{C}^1$ de période T de $\mathbb{R}$ dans $\mathbb{C}$. Montrer que

$$\iint_{[0,\mathrm{T}]^2} |f'(u) - f'(v)|^2 \mathrm{d}u\mathrm{d}v \geqslant \left(\frac{2\pi}{\mathrm{T}}\right)^2 \iint_{[0,\mathrm{T}]^2} |f(u) - f(v)|^2 \mathrm{d}u\mathrm{d}v.$$

Soit F une application λ-lipschitzienne de $\mathbb{R}$ dans $\mathbb{R}$ et (E) l'équation différentielle $x'(t) = \mathrm{F}(x(t))$. Soit φ une solution non constante et T-périodique de (E). Montrer que $\dfrac{2\pi}{\mathrm{T}} \leqslant \lambda$.

2. Soit $(x_n)_{n\in\mathbb{N}}$ une suite de réels K-périodique. Montrer que

$$\sum_{0 \leqslant n,m \leqslant \mathrm{K}-1} |x_n - x_m|^2 = 2 \sum_{k=1}^{\mathrm{K}-1} \left| \sum_{n=0}^{\mathrm{K}-1} x_n e^{-i\frac{2k\pi n}{\mathrm{K}}} \right|^2 .$$

Soit $\mathrm{D} : x = (x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \longmapsto \mathrm{D}(x) = (x_{n+1} - x_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$. Trouver une constante $\mathrm{C} > 0$ telle que pour toute suite $(x_n)_{n\in\mathbb{N}}$ K-périodique et tout $k \in [\![1, \mathrm{K}-1]\!]$, on ait

$$\left| \sum_{n=0}^{\mathrm{K}-1} \mathrm{D}(x)_n e^{-i\frac{2k\pi n}{\mathrm{K}}} \right| \geqslant \mathrm{C} \left| \sum_{n=0}^{\mathrm{K}-1} x_n e^{-i\frac{2k\pi n}{\mathrm{K}}} \right| .$$

Soit F une application λ-lipschitzienne de $\mathbb{R}$ dans $\mathbb{R}$. Soit $(x_n)_{n\in\mathbb{N}}$ une suite non constante K-périodique vérifiant $x_{n+1} = x_n + \mathrm{F}(x_n)$. Trouver un minorant de K.

(École polytechnique)

▷ **Solution.**

1. Notons c_k $(k \in \mathbb{Z})$ les coefficients de Fourier de f. On a donc pour tout $k \in \mathbb{Z}$, $c_k = \dfrac{1}{\mathrm{T}}\displaystyle\int_0^{\mathrm{T}} f(t)\mathrm{e}^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t$. À partir de l'inégalité de Parseval $\displaystyle\int_0^{\mathrm{T}} |f(t)|^2\mathrm{d}t = \mathrm{T}\sum_{k\in\mathbb{Z}} |c_k|^2$, on peut calculer le premier membre de l'égalité. On obtient

$$\begin{aligned}
\iint_{[0,\mathrm{T}]^2} |f(u)-f(v)|^2\mathrm{d}u\mathrm{d}v &= \iint_{[0,\mathrm{T}]^2} \Big(|f(u)|^2 + |f(v)|^2 - f(u)\overline{f(v)} - f(v)\overline{f(u)}\Big)\mathrm{d}u\mathrm{d}v \\
&= 2\mathrm{T}\int_0^{\mathrm{T}} |f(u)|^2 - 2\int_0^{\mathrm{T}} f(u)\mathrm{d}u \int_0^{\mathrm{T}} \overline{f(v)}\mathrm{d}v \\
&= 2\mathrm{T}^2\sum_{k\in\mathbb{Z}} |c_k|^2 - 2\mathrm{T}^2|c_0|^2 \\
&= 2\mathrm{T}^2\sum_{k\in\mathbb{Z}^*} |c_k|^2 \\
&= 2\sum_{k\in\mathbb{Z}^*} \left|\int_0^{\mathrm{T}} f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t\right|^2.
\end{aligned}$$

Si f est $\mathcal{C}^1$ et T-périodique, la relation précédente s'applique à f', ce qui donne

$$\iint_{[0,\mathrm{T}]^2} |f'(u)-f'(v)|^2\mathrm{d}u\mathrm{d}v = 2\sum_{k\in\mathbb{Z}^*} \left|\int_0^{\mathrm{T}} f'(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t\right|^2.$$

En intégrant par parties, on obtient pour tout $k \in \mathbb{Z}^*$,

$$\begin{aligned}
\int_0^{\mathrm{T}} f'(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t &= \Big[f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\Big]_0^{\mathrm{T}} + \frac{2i\pi k}{\mathrm{T}}\int_0^{\mathrm{T}} f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t \\
&= \frac{2i\pi k}{\mathrm{T}}\int_0^{\mathrm{T}} f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t.
\end{aligned}$$

On en déduit que, pour $k \in \mathbb{Z}^*$,

$$\begin{aligned}
\left|\int_0^{\mathrm{T}} f'(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t\right|^2 &= \left(\frac{2k\pi}{\mathrm{T}}\right)^2\left|\int_0^{\mathrm{T}} f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t\right|^2 \\
&\geqslant \left(\frac{2\pi}{\mathrm{T}}\right)^2\left|\int_0^{\mathrm{T}} f(t)e^{-i\frac{2k\pi t}{\mathrm{T}}}\mathrm{d}t\right|^2.
\end{aligned}$$

De ce qui précède, on déduit que

$$\iint_{[0,\mathrm{T}]^2} |f'(u)-f'(v)|^2\mathrm{d}u\mathrm{d}v \geqslant \left(\frac{2\pi}{\mathrm{T}}\right)^2 \iint_{[0,\mathrm{T}]^2} |f(u)-f(v)|^2\mathrm{d}u\mathrm{d}v.$$

On a, pour tout réel x, $\varphi'(t) = \mathrm{F}(\varphi(t))$. Sachant que F est λ-lipschitzienne, on en déduit pour $(u, v) \in \mathbb{R}^2$, l'inégalité

$$|\varphi'(u) - \varphi'(v)| = |\mathrm{F}(\varphi(u)) - \mathrm{F}(\varphi(v))| \leqslant \lambda |\varphi(u) - \varphi(v)|.$$

On obtient, en intégrant,

$$\iint_{[0,1]^2} |\varphi'(u) - \varphi'(v)|^2 \mathrm{d}u\mathrm{d}v \leqslant \lambda^2 \iint_{[0,1]^2} |\varphi(u) - \varphi(v)|^2 \mathrm{d}u\mathrm{d}v,$$

puis en appliquant l'inégalité précédente à la fonction φ qui est $\mathcal{C}^1$ et T-périodique

$$\left(\frac{2\pi}{\mathrm{T}}\right)^2 \iint_{[0,1]^2} |\varphi(u) - \varphi(v)|^2 \mathrm{d}u\mathrm{d}v \leqslant \lambda^2 \iint_{[0,1]^2} |\varphi(u) - \varphi(v)|^2 \mathrm{d}u\mathrm{d}v.$$

La fonction φ n'étant pas constante, la fonction $(u, v) \longmapsto |\varphi(u) - \varphi(v)|$ est continue et non identiquement nulle sur $[0, \mathrm{T}]^2$. On en déduit que $\iint_{[0,1]^2} |\varphi(u) - \varphi(v)|^2 \mathrm{d}u\mathrm{d}v > 0$. En simplifiant, on obtient l'inégalité

$$\left(\frac{2\pi}{\mathrm{T}}\right)^2 \leqslant \lambda^2, \text{ c'est-à-dire } \frac{2\pi}{\mathrm{T}} \leqslant \lambda.$$

2. En développant le premier membre de l'égalité à démontrer, on obtient

$$\begin{aligned}\sum_{0 \leqslant m,n \leqslant \mathrm{K}-1} |x_m - x_n|^2 &= \sum_{0 \leqslant m,n \leqslant \mathrm{K}-1} \left(|x_n|^2 + |x_m|^2 - x_n\overline{x_m} - x_m\overline{x_n}\right) \\ &= 2\mathrm{K} \sum_{m=0}^{\mathrm{K}-1} |x_m|^2 - 2 \sum_{0 \leqslant m,n \leqslant \mathrm{K}-1} x_m\overline{x_n}.\end{aligned}$$

On transforme ensuite $\mathrm{S} = 2\sum_{k=1}^{\mathrm{K}-1} \left|\sum_{n=0}^{\mathrm{K}-1} x_n e^{-i\frac{2k\pi n}{\mathrm{K}}}\right|^2$. On obtient, en échangeant l'ordre des sommations,

$$\begin{aligned}\mathrm{S} &= 2 \sum_{k=1}^{\mathrm{K}-1} \sum_{0 \leqslant m,n \leqslant \mathrm{K}-1} x_m\overline{x_n} e^{-i\frac{2k\pi}{\mathrm{K}}(m-n)} \\ &= 2 \sum_{0 \leqslant m,n \leqslant \mathrm{K}-1} x_m\overline{x_n} \sum_{k=1}^{\mathrm{K}-1} e^{-i\frac{2k\pi}{\mathrm{K}}(m-n)}.\end{aligned}$$

On calcule ensuite les sommes $\sum_{k=1}^{\mathrm{K}-1} e^{-i\frac{2k\pi}{\mathrm{K}}(m-n)}$.

• Si $m \neq n$, on a $e^{-i\frac{2\pi}{K}(m-n)} \neq 1$ et

$$\sum_{k=1}^{K-1} e^{-i\frac{2k\pi}{K}(m-n)} = -1 + \frac{e^{-i2\pi(m-n)} - 1}{e^{-i\frac{2\pi(m-n)}{K}} - 1} = -1.$$

• Si $m = n$, on a $e^{-i\frac{2k\pi}{K}(m-n)} = 1$ et

$$\sum_{k=1}^{K-1} e^{-i\frac{2k\pi}{K}(m-n)} = K - 1.$$

On en déduit que

$$\begin{aligned} S &= -2 \sum_{m \neq n} x_m \overline{x_n} + 2(K-1) \sum_{m=0}^{K-1} |x_m|^2 \\ &= -2 \sum_{0 \leqslant m,n \leqslant K-1} x_m \overline{x_n} + 2K \sum_{m=0}^{K-1} |x_m|^2, \end{aligned}$$

ce qui donne le résultat voulu.

On transforme l'expression à minorer. On obtient

$$\begin{aligned} \left| \sum_{n=0}^{K-1} (x_{n+1} - x_n) e^{-i\frac{2k\pi n}{K}} \right| &= \left| \sum_{n=0}^{K-1} x_{n+1} e^{-i\frac{2k\pi n}{K}} - \sum_{n=0}^{K-1} x_n e^{-i\frac{2k\pi n}{K}} \right| \\ &= \left| \sum_{n=1}^{K} x_n e^{-i\frac{2k\pi(n-1)}{K}} - \sum_{n=0}^{K-1} x_n e^{-i\frac{2k\pi n}{K}} \right| \\ &= \left| e^{i\frac{2k\pi}{K}} - 1 \right| \left| \sum_{n=0}^{K-1} x_n e^{-i\frac{2k\pi n}{K}} \right|, \end{aligned}$$

car $x_K = x_0$, la suite (x_n) étant K-périodique.

En écrivant

$$\left| e^{i\frac{2k\pi}{K}} - 1 \right| = \left| e^{i\frac{k\pi}{K}} \left(e^{i\frac{k\pi}{K}} - e^{i\frac{-k\pi}{K}} \right) \right| = 2 \sin \frac{k\pi}{K},$$

on obtient, pour $k \in [\![1, K-1]\!]$, la minoration $\left| e^{i\frac{2k\pi}{K}} - 1 \right| \geqslant 2 \sin \frac{\pi}{K}$. Ceci montre l'existence de la constante C, égale à $2 \sin \frac{\pi}{K}$.

En sommant les égalités obtenues en faisant varier k de 1 à $K-1$, on obtient

$$\sum_{k=1}^{K-1} \left| \sum_{n=0}^{K-1} D(x)_n e^{-i\frac{2k\pi n}{K}} \right|^2 \geqslant C^2 \sum_{k=1}^{K-1} \left| \sum_{n=0}^{K-1} x_n e^{-i\frac{2k\pi n}{K}} \right|^2,$$

ce qui équivaut, compte tenu de ce qui précède à

$$\sum_{0\leqslant n,m\leqslant K-1} |D(x)_n - D(x)_m|^2 \geqslant C^2 \sum_{0\leqslant n,m\leqslant K-1} |x_n - x_m|^2.$$

L'application F étant λ-lipschitzienne, on a

$$\sum_{0\leqslant n,m\leqslant K-1} |D(x)_n - D(x)_m|^2 \leqslant \lambda^2 \sum_{0\leqslant n,m\leqslant K-1} |x_n - x_m|^2,$$

ce qui conduit à

$$C^2 \sum_{0\leqslant n,m\leqslant K-1} |x_n - x_m|^2 \leqslant \lambda^2 \sum_{0\leqslant n,m\leqslant K-1} |x_n - x_m|^2.$$

La somme n'étant pas nulle, car la suite (x_n) est K-périodique et non constante, on en déduit que $C \leqslant \lambda$, c'est-à-dire que $2\sin\dfrac{\pi}{K} \leqslant \lambda$. On obtient $\dfrac{\pi}{K} \leqslant \arcsin\dfrac{\lambda}{2}$ et finalement $K \geqslant \dfrac{\pi}{\arcsin\dfrac{\lambda}{2}}$. ◁

Les deux exercices qui suivent ont trait à la résolution d'équations différentielles linéaires dont les coefficients sont des fonctions 2π-périodiques. On recherche des solutions 2π-périodiques. L'équation différentielle se traduit immédiatement par des relations sur les coefficients de Fourier qui permettent de calculer les coefficients de la fonction inconnue et à partir de là de déterminer celle-ci.

Si une fonction est de classe C^k, la relation de Parseval appliquée à $f^{(k)}$ montre que la famille $(\sum n^{2k}|c_n(f)|^2)_{n\in\mathbb{Z}}$ est sommable. Dans ces exercices, on fait l'hypothèse inverse que la famille $(\sum n^{2k}|c_n(f)|^2)_{n\in\mathbb{Z}}$ est sommable et on démontre alors que f est au moins de classe C^{k-1}.

Le premier de ces exercices se place dans le cadre des fonctions 2π-périodiques de $\mathbb{R}$ dans $\mathbb{C}^p$. Cela ne change pas beaucoup du cadre habituel ; il suffit d'appliquer les théorèmes usuels à chacune des composantes.

4.22. Résolution d'une équation différentielle

Soit H l'ensemble des fonctions $f : \mathbb{R} \to \mathbb{C}^p$, continues et 2π-périodiques telles que $\sum n^4(|c_n(f)|^2 + |c_{-n}(f)|^2)$ converge, où $|\ |$ est la norme hermitienne de $\mathbb{C}^p$.

Si $f \in \mathrm{H}$, on pose

$$\mathrm{N}(f) = \sqrt{|c_0(f)|^2 + \sum_{n\in\mathbb{Z}} n^4 |c_n(f)|^2}.$$

1. Montrer que N est une norme. Monter que, si $f \in \mathrm{H}$, f est de classe $\mathcal{C}^1$. Montrer qu'il existe $\alpha > 0$ tel que, pour tout $f \in \mathrm{H}$, $\|f'\|_\infty \leqslant \alpha \mathrm{N}(f)$.

2. Soit $\mathrm{A} \in \mathcal{M}_p(\mathbb{R})$, symétrique, $g \in \mathrm{H}$ et $\mathrm{C} > 0$. Montrer qu'il existe un unique élément $f \in \mathrm{H}$ tel que $f - g + \dfrac{\mathrm{C}}{2}\mathrm{A}(f' + g') = 0$. Montrer que $\mathrm{N}(f) = \mathrm{N}(g)$.

(École normale supérieure)

▷ **Solution.**

1. Soit f et g sont deux éléments de H et $\lambda \in \mathbb{C}$, on a, pour tout $n \in \mathbb{Z}$, $c_n(f+g) = c_n(f) + c_n(g)$, d'où l'on tire

$$\begin{aligned} |c_n(f+g)|^2 &= |c_n(f)|^2 + |c_n(g)|^2 + 2\langle c_n(f), c_n(g)\rangle \\ &\leqslant |c_n(f)|^2 + |c_n(g)|^2 + 2|c_n(f)||c_n(g)| \\ &\leqslant 2(|c_n(f)|^2 + |c_n(g)|^2). \end{aligned}$$

On en déduit que la série $\sum n^4(|c_n(f+g)|^2 + |c_{-n}(f+g)|^2)$ converge, c'est-à-dire que $f + g \in \mathrm{H}$. On note, de plus que, pour tout $n \in \mathbb{Z}$, on a $c_n(\lambda f)|^2 = |\lambda|^2 |c_n(f)|^2$; on en déduit que $\sum n^4(|c_n(\lambda f)|^2 + |c_{-n}(\lambda f)|^2)$ converge, c'est-à-dire que λf est dans H. L'ensemble H est un sous-espace vectoriel de l'espace des applications continues et 2π-périodiques de $\mathbb{R}$ dans $\mathbb{C}^p$.

Posons, pour f et g dans H,

$$\langle f, g\rangle = \overline{\langle c_0(f)}, c_0(g)\rangle + \sum_{n\in\mathbb{Z}} n^4 \overline{\langle c_n(f)}, c_n(g)\rangle.$$

Cette définition a un sens car, pour tout $\mathrm{N} \in \mathbb{N}^*$, on a

$$\begin{aligned} \sum_{|n|\leqslant \mathrm{N}} n^4 |\langle c_n(f), c_n(g)\rangle| &\leqslant \sum_{|n|\leqslant \mathrm{N}} n^2 |c_n(f)| n^2 |c_n(g)| \\ &\leqslant \sqrt{\sum_{|n|\leqslant \mathrm{N}} n^4 |c_n(f)|^2 \cdot \sum_{|n|\leqslant \mathrm{N}} n^4 |c_n(g)|^2} \\ &\leqslant \sqrt{\sum_{n\in\mathbb{Z}} n^4 |c_n(f)|^2 \cdot \sum_{n\in\mathbb{Z}} n^4 |c_n(g)|^2}, \end{aligned}$$

d'après l'inégalité de Cauchy-Schwarz. On définit ainsi clairement une forme sesquilinéaire hermitienne positive sur H.

Montrons qu'elle est définie positive. Supposons que $\langle f, f\rangle = 0$. On a alors, pour tout $n \in \mathbb{Z}$, $c_n(f) = 0$. La formule de Parseval montre que $\int_0^{2\pi} |f(t)|^2 \mathrm{d}t = 0$ et donc que $f = 0$, puisque la fonction f est continue. On notera que si $f_1, f_2, \ldots, f_p$ sont les composantes de f, on a, pour tout $t \in \mathbb{R}$, $|f(t)|^2 = \sum_{k=1}^{p} |f_k(t)|^2$ et, pour tout $n \in \mathbb{Z}$, $c_n(f) = (c_n(f_1), c_n(f_2), \ldots, c_n(f_p))$ et donc $|c_n(f)|^2 = \sum_{k=1}^{p} |c_n(f_k)|^2$. La formule de Parseval pour f s'obtient en sommant les relations obtenues en appliquant la formule de Parseval aux fonctions $f_1, f_2, \ldots, f_p$.

On a défini un produit hermitien sur H et N est clairement la norme issue de ce produit hermitien. C'est donc une norme.

Montrons que si $f \in \mathrm{H}$, alors f est de classe $\mathcal{C}^1$. Pour cela, nous allons appliquer le théorème de dérivation des séries de fonctions. Pour tout $\mathrm{N}^* \in \mathbb{N}$, on a

$$\begin{aligned}\sum_{0<|n|\leqslant \mathrm{N}} |c_n(f) in e^{inx}| &= \sum_{0<|n|\leqslant \mathrm{N}} |c_n(f) n^2| \left|\frac{1}{n}\right| \\ &\leqslant \sqrt{\sum_{0<|n|\leqslant \mathrm{N}} |c_n(f)|^2 n^4} \cdot \sqrt{\sum_{0\leqslant|n|\leqslant \mathrm{N}} \frac{1}{n^2}} \\ &\leqslant \sqrt{\sum_{n\in\mathbb{Z}} |c_n(f)|^2 n^4} \cdot \sqrt{2 \sum_{n\in\mathbb{N}^*} \frac{1}{n^2}}.\end{aligned}$$

On en déduit que la série de fonctions $\sum c_n(f) in e^{inx}$ converge normalement et qu'il en est de même *a fortiori* de la série $\sum c_n(f) e^{inx}$. La fonction h définie par $h(x) = \sum_{n\in\mathbb{Z}} c_n(f) e^{inx}$ est donc de classe $\mathcal{C}^1$ sur $\mathbb{R}$ et sa dérivée h' est donnée par $h'(x) = \sum_{n\in\mathbb{Z}} c_n(f) in e^{inx}$. Il reste à justifier que $h = f$. La convergence normale de la série définissant h implique que les coefficients de Fourier de h sont égaux à ceux de f. La fonction $h - f$ dont les coefficients de Fourier sont nuls est donc nulle d'après la formule de Parseval. La fonction f est donc de classe $\mathcal{C}^1$ et, pour tout $x \in \mathbb{R}$, $f'(x) = \sum_{n\in\mathbb{Z}} c_n(f) in e^{inx}$. On obtient en utilisant l'inégalité démontrée plus haut, pour tout $x \in \mathbb{R}$,

$$|f'(x)| \leqslant \sum_{n\in\mathbb{Z}} |c_n(f) n| \leqslant \sqrt{\sum_{n\in\mathbb{Z}} |c_n(f)|^2 n^4} \cdot \sqrt{2 \sum_{n\in\mathbb{N}^*} \frac{1}{n^2}} \leqslant \frac{\pi}{\sqrt{3}} \mathrm{N}(f)$$

et donc $\|f'\|_\infty \leqslant \frac{\pi}{\sqrt{3}} \mathrm{N}(f)$.

2. Nous allons nous ramener au cas où la matrice A est diagonale. Il existe des matrices D diagonale et P orthogonale telle que $A = {}^tPDP$. En multipliant par P à gauche, l'équation $f - g + \frac{C}{2} A(f' + g') = 0$ équivaut donc à

$$Pf - Pg + \frac{C}{2} DPf' + DPg' = 0.$$

La matrice P étant à coefficients constants, on a, pour $f \in H$ et $n \in \mathbb{Z}$, $c_n(Pf) = Pc_n(f)$ et donc, puisque P est orthogonale, $|c_n(Pf)| = |c_n(f)|$. On en déduit que Pf appartient encore à H. L'application $f \to Pf$ est une bijection de H sur H. On remarque de plus que $(Pf)' = Pf'$. Il suffit donc de démontrer que, pour tout $g \in H$, il existe un unique élément f de H tel que $f - g + \frac{C}{2} D(f' + g') = 0$. Si $D = \mathrm{Diag}(\lambda_1, \lambda_2, \ldots, \lambda_p)$. et si $f_1, f_2, \ldots, f_p$ (resp. $g_1, g_2, \ldots, g_p$) sont les composantes de f (resp. g), l'équation équivaut à

$$\forall k \in [\![1, p]\!], \quad f_k - g_k + \frac{C}{2} \lambda_k (f_k' + g_k') = 0.$$

On écrit que les coefficients de Fourier de $f_k - g_k + \frac{C}{2} \lambda_k (f_k' + g_k')$ sont nuls. On obtient, pour tout $n \in \mathbb{Z}$,

$$c_n(f_k) - c_n(g_k) + \frac{C}{2} \lambda_k (inc_n(f_k) + inc_n(g_k)) = 0$$

et donc

$$c_n(f_k) = \frac{1 - \frac{C}{2} \lambda_k in}{1 + \frac{C}{2} \lambda_k in} c_n(g_k) = \frac{2 - C\lambda_k in}{2 + C\lambda_k in} c_n(g_k).$$

Les coefficients de Fourier des fonctions f_k étant imposés, il y a au plus une solution.

On a, pour $n \in \mathbb{Z}$ et $k \in [\![1, p]\!]$, $\left| \frac{2 - C\lambda_k in}{2 + C\lambda_k in} c_n(g_k) \right| = |c_n(g_k)|$. Puisque $g \in H$, la série $\sum |c_n(g_k)|$ converge, car $|c_n(g_k)| \leqslant |c_n(g)|$. On peut donc poser, pour tout $x \in \mathbb{R}$,

$$f_k(x) = \sum_{n \in \mathbb{Z}} \frac{2 - C\lambda_k in}{2 + C\lambda_k in} c_n(g_k) e^{inx}.$$

La convergence étant normale, on a bien $c_n(f_k) = \frac{2 - C\lambda_k in}{2 + C\lambda_k in} c_n(g_k)$ pour tout n. Posons alors $f = (f_1, f_2, \ldots, f_p)$. On obtient, pour tout $n \in \mathbb{Z}$, $c_n(f) = (c_n(f_1), \ldots, c_n(f_p))$ et

$$|c_n(f)|^2 = \sum_{k=1}^{p} |c_n(f_k)|^2 = \sum_{k=1}^{p} |c_n(g_k)|^2 = |c_n(g)|^2.$$

On en déduit que, comme g, f appartient à H et $\mathrm{N}(f) = \mathrm{N}(g)$. Par construction, f est l'unique solution de l'équation différentielle. $\lhd$

4.23. Étude de convergence uniforme

Pour $k \in \mathbb{N}$, on note $\mathcal{C}^k_{2\pi}$ l'ensemble des fonctions 2π-périodiques de $\mathbb{R}$ dans $\mathbb{C}$ de classe $\mathcal{C}^k$ et

$$\mathrm{H}_k = \{f \in \mathcal{C}^0_{2\pi},\ \sum_{n\in\mathbb{Z}} n^{2k}|c_n(f)|^2 < +\infty\}.$$

1. Montrer que $\mathrm{H}_{k+1} \subset \mathcal{C}^k_{2\pi}$. Que peut-on dire de l'inclusion inverse ?

2. Soit $k \geqslant 1$. Montrer que, pour tout réel $a > 0$ et tout $f \in \mathrm{H}_k$, il existe un unique $g \in \mathrm{H}_{k+2}$ tel que $g'' = \dfrac{g-f}{a}$.

3. Soit $f_0 \in \mathrm{H}_1$, $a > 0$ et (f_p) définie par $f''_{p+1} = \dfrac{f_{p+1}-f_p}{a}$. Étudier la convergence de la suite (f_p).

4. Pour $t \in [pa, (p+1)a[$, on pose $\mathrm{F}_a(t,x) = f_p(x)$. Donner les coefficients de Fourier de $x \longmapsto \mathrm{F}_a(t,x)$. Étudier la convergence uniforme de (F_a) lorsque a tend vers 0.

(École normale supérieure)

$\rhd$ **Solution.**

1. Soit $f \in \mathrm{H}_{k+1}$. Pour montrer que $f \in \mathcal{C}^k_{2\pi}$, nous allons montrer que f est égale à la somme de sa série de Fourier et que celle-ci est dérivable k fois terme à terme. Par l'inégalité de Cauchy-Schwarz, on a pour $\mathrm{N} \in \mathbb{N}$,

$$\begin{aligned}\sum_{1\leqslant|n|\leqslant\mathrm{N}} n^k|c_n(f)| &= \sum_{1\leqslant|n|\leqslant\mathrm{N}} n^{k+1}|c_n(f)| \times \frac{1}{n}\\ &\leqslant \sqrt{\sum_{1\leqslant|n|\leqslant\mathrm{N}} n^{2k+2}|c_n(f)|^2}\sqrt{\sum_{1\leqslant|n|\leqslant\mathrm{N}}\frac{1}{n^2}}\\ &\leqslant \sqrt{\sum_{n\in\mathbb{Z}} n^{2k+2}|c_n(f)|^2}\sqrt{2\sum_{n=1}^{+\infty}\frac{1}{n^2}},\end{aligned}$$

puisque la série $\sum_{n\in\mathbb{Z}} n^{2k+2}|c_n(f)|^2$ converge par hypothèse. Ainsi, la série $\sum_{n\in\mathbb{Z}} n^k|c_n(f)|$ est convergence et il en de même *a fortiori* de $\sum_{n\in\mathbb{Z}} n^\ell|c_n(f)|$ pour $0 \leqslant \ell \leqslant k$.

On pose, pour tout réel x, $g(x) = \sum_{n\in\mathbb{Z}} c_n(f)e^{inx}$. La série qui définit $g(x)$ est normalement convergente. Il en est de même, pour tout $0 \leqslant \ell \leqslant k$ de la série $\sum_{n\in\mathbb{Z}} c_n(f)(in)^\ell e^{inx}$ obtenue en dérivant ℓ fois terme à terme. Une application réitérée du théorème de dérivation des séries de fonctions montre que g est de classe $\mathcal{C}^k$. Enfin, la série définissant g étant normalement convergente, ses coefficients de Fourier peuvent s'obtenir en échangeant intégration et sommation, ce qui donne, pour tout $p \in \mathbb{Z}$,

$$c_p(g) = \frac{1}{2\pi}\int_0^{2\pi} g(x)e^{-ipx}\mathrm{d}x = \sum_{n\in\mathbb{Z}} \frac{1}{2\pi}\int_0^{2\pi} c_n(f)e^{i(n-p)x}\mathrm{d}x = c_p(f).$$

Les fonctions f et g ont donc mêmes coefficients de Fourier. Ceux de la fonction $f-g$ sont donc nuls. De la formule de Parseval on déduit alors que $\int_0^{2\pi} |f(x)-g(x)|^2\mathrm{d}x = 0$ et donc $f = g$, puisque $f-g$ est continue et 2π-périodique. La fonction f qui est égale à g est de classe $\mathcal{C}^k$ et on a l'inclusion $\mathrm{H}_{k+1} \subset \mathcal{C}^k_{2\pi}$.

Supposons réciproquement que $f \in \mathcal{C}^k_{2\pi}$. Si $k \geqslant 1$, on peut calculer les coefficients de Fourier de f'. On sait que, pour $n \in \mathbb{Z}$, $c_n(f') = (in)c_n(f)$. Un récurrence immédiate conduit à $c_n(f^{(k)}) = (in)^k c_n(f)$. La formule de Parseval appliquée à $f^{(k)}$ donne

$$\sum_{n\in\mathbb{Z}} |c_n(f^{(k)})|^2 = \sum_{n\in\mathbb{Z}} n^{2k}|c_n(f)|^2 = \frac{1}{2\pi}\int_0^{2\pi} |f^{(k)}(x)|^2\mathrm{d}x.$$

Ceci montre que $f \in \mathcal{C}^k_{2\pi}$. On a donc l'inclusion $\mathcal{C}^k_{2\pi} \subset \mathrm{H}_k$. Par contre, on n'a pas $\mathcal{C}^k_{2\pi} \subset \mathrm{H}_{k+1}$. Considérons, pour $k \in \mathbb{N}$, la fonction f définie sur $\mathbb{R}$ par $f(x) = \sum_{n\in\mathbb{Z}^*} \frac{e^{inx}}{n^{k+\frac{3}{2}}}$. La série qui définit f est normalement convergente. On en déduit que f est continue et que ses coefficients de Fourier sont définies par $c_n(f) = \frac{1}{n^{k+\frac{3}{2}}}$. Le théorème de dérivation des séries de fonction montre que f est de classe $\mathcal{C}^k$ et que, pour tout réel x, on a $f^{(k)}(x) = \sum_{n\in\mathbb{Z}^*} i^k \frac{e^{inx}}{n^{\frac{3}{2}}}$. Cependant f n'appartient pas à H_{k+1} car la série de terme général $n^{2(k+1)}|c_n(f)|^2 = \frac{n^{2(k+1)}}{n^{2k+3}} = \frac{1}{n}$ diverge.

2. On cherche $g \in \mathrm{H}_{k+2}$ telle que $g - ag'' = f$. On sait d'après la question 1 que si $g \in \mathrm{H}_{k+2}$, alors g est de classe $\mathcal{C}^{k+1}$ et donc de classe $\mathcal{C}^2$, puisque $k \geqslant 1$. Si les fonctions $g - ag''$ et f sont égales, elles ont mêmes coefficients de Fourier. On doit donc avoir, pour tout $n \in \mathbb{Z}$,

$$c_n(f) = c_n(g) - ac_n(g'') = c_n(g) - a(in)^2 c_n(g) = (1+an^2)c_n(g)$$

et $c_n(g) = \frac{c_n(f)}{1+an^2}$. Si g existe, elle est d'après la question 1 égale à la somme de sa série de Fourier et on a, pour tout réel x,

$$g(x) = \sum_{n\in\mathbb{Z}} \frac{c_n(f)}{1+an^2} e^{inx}.$$

La suite $(c_n(f))$ converge vers 0 donc la série définissant g est normalement convergente. Elle définit une fonction g continue et 2π-périodique dont les coefficients de Fourier sont les $\frac{c_n(f)}{1+an^2}$. On a, de plus $n^{2k+4}|c_n(g)|^2 \underset{n\to\infty}{\sim} \frac{1}{a^2} n^{2k}|c_n(f)|^2$ et puisque f appartient à H_k, g appartient à H_{k+2}. La fonction g est au moins de classe $\mathcal{C}^2$ et par construction les fonctions $g - ag''$ et f ont mêmes coefficients de Fourier. Elles sont égales.

3. En utilisant la question 2, on démontre de manière simple, par récurrence sur p, que f_p est définie pour tout $p \in \mathbb{N}$ et est élément de H_{2p+1}. D'après la question 1, $f - P$ est égale à la somme de sa série de Fourier. Les coefficients de Fourier de f_p sont définis, pour tout $n \in \mathbb{Z}$ par $c_n(f_p) = \frac{c_n(f_0)}{(1+an^2)^p}$. On en déduit que, pour tout $n \in \mathbb{Z}^*$, la suite $c_n(f_p)_{p\in\mathbb{N}}$ converge vers 0, alors que la suite $(c_0(f_p))_{p\in\mathbb{N}}$ est constante. Nous allons montrer que (f_p) converge uniformément vers la fonction constante $x \longmapsto c_0(f_0)$. Il s'agit d'inverser une sommation et un passage à la limite. On a, pour tous $p \in \mathbb{N}$ et $n \in \mathbb{Z}$, $|c_n(f_p)| \leqslant |c_n(f_0)|$. Le salut vient de cette majoration uniforme par rapport à p. On a, pour tout $x \in \mathbb{R}$,

$$|f_p(x) - c_0(f_0)| = |f_p(x) - c_0(f_p)| = \left| \sum_{n\in\mathbb{Z}^*} \frac{c_n(f_0)}{(1+an^2)^p} e^{inx} \right|$$

$$\leqslant \sum_{n\in\mathbb{Z}^*} \frac{|c_n(f_0)|}{(1+an^2)^p}.$$

Pour $N \in \mathbb{N}$, on obtient

$$|f_p(x) - c_0(f_0)| \leqslant \sum_{1\leqslant|n|\leqslant N} \frac{|c_n(f_0)|}{(1+an^2)^p} + \sum_{|n|>N} |c_n(f_0)|.$$

Soit $\varepsilon > 0$. On peut choisir N pour que $\sum_{|n|>\mathbb{N}} |c_n(f_0)| \leqslant \varepsilon$, car $\sum |c_n(f_0)|$ converge, puisque f_0 est dans H_1. Ensuite, il existe $p_0 \in \mathbb{N}$ tel que $\sum_{1\leqslant|n|\leqslant N} \frac{|c_n(f_0)|}{(1+an^2)^p} \leqslant \varepsilon$, puisque chaque terme de cette somme (finie) tend vers 0. On a donc, pour tous $p \geqslant p_0$ et $x \in \mathbb{R}$, $|f_p(x) - c_0(f_0)| \leqslant 2\varepsilon$. La suite (f_p) converge uniformément vers $c_0(f_0) = \frac{1}{2\pi}\int_0^{2\pi} f_0(x)\mathrm{d}x$.

4. Pour t fixé, la fonction $x \longmapsto F_a(t,x)$ coïncide avec f_p, où $p = E\left(\frac{t}{a}\right)$. Ses coefficients de Fourier sont donc, d'après la question

précédente, $\dfrac{c_n(f_0)}{(1+an^2)^p} = \dfrac{c_n(f_0)}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}$. On sait que la fonction f_p est, pour tout $p \in \mathbb{N}$, égale à sa série de Fourier. On a donc, pour $a > 0$, $t \geqslant 0$ et $x \in \mathbb{R}$,

$$\mathrm{F}_a(t,x) = \sum_{n\in\mathbb{Z}} \frac{c_n(f_0)}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}} e^{inx}.$$

Posons, pour $x \in \mathbb{R}$ et $t \geqslant 0$, $u_n(a,t,x) = \dfrac{c_n(f_0)}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}} e^{inx}$. Cherchons la limite de $u_n(a,t,x)$ quand a tend vers 0. On peut écrire

$$(1+an^2)^{\mathrm{E}(\frac{t}{a})} = \exp\left(\mathrm{E}\left(\frac{t}{a}\right)\ln(1+an^2)\right).$$

Si $t > 0$, on a

$$\mathrm{E}\left(\frac{t}{a}\right)\ln(1+an^2) \underset{a\to 0}{\sim} \frac{t}{a}\cdot an^2 = tn^2.$$

On en déduit que $\lim\limits_{a\to 0}(1+an^2)^{\mathrm{E}(\frac{t}{a})} = e^{tn^2}$, résultat qui reste vrai pour $t = 0$. On obtient, pour $t \geqslant 0$ et $x \in \mathbb{R}$, $\lim\limits_{a\to 0} u_n(a,t,x) = c_n(f_0)e^{-tn^2}e^{inx}$. On va montrer évidemment que, pour $t \geqslant 0$ et $x \in \mathbb{R}$,

$$\lim_{a\to 0} \mathrm{F}_a(t,x) = \sum_{n\in\mathbb{Z}} c_n(f_0)e^{-tn^2}e^{inx}$$

et que, de plus, cette convergence est uniforme par rapport à t et x. On pose, pour tout $n \in \mathbb{Z}$, $v_n(t,x) = c_n(f_0)e^{-nt^2}e^{inx}$. On remarque que, pour $a > 0$, $t \geqslant 0$ et $x \in \mathbb{R}$, on a $(1+an^2)^{\mathrm{E}(\frac{t}{a})} \geqslant 1$. On en déduit $|u_n(a,t,x)| \leqslant |c_n(f_0)|$ et de plus $|v_n(x,t)| \leqslant |c_n(f_0)|$, majorations indépendantes de t et x. La série $\sum |c_n(f_0)|$ étant convergente, la série $\sum v_n(t,x)$ est normalement convergente. On note $\mathrm{G}(t,x)$ sa somme. On écrit, pour tous t, x, a et $\mathrm{N} \in \mathbb{N}$

$$\begin{aligned}|\mathrm{F}_a(t,x) - \mathrm{G}(t,x)| &\leqslant \sum_{n\in\mathbb{Z}} |u_n(a,t,x) - v_n(t,x)| \\ &\leqslant \sum_{|n|\leqslant \mathrm{N}} |u_n(a,t,x) - v_n(t,x)| + 2\sum_{|n|>\mathrm{N}} |c_n(f_0)|.\end{aligned}$$

Montrons que, pour tout $n \in \mathbb{Z}$, la convergence de $u_n(a,t,x)$ vers $v_n(t,x)$ est uniforme par rapport à t et x. On a pour $t \geqslant 0$ et $x \in \mathbb{R}$,

$$|u_n(a,t,x) - v_n(t,x)| \leqslant |c_n(f_0)|\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}} - e^{-tn^2}\right|$$

et

$$\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}-e^{-tn^2}\right| = e^{-tn^2}\left|\exp(tn^2-\mathrm{E}\left(\frac{t}{a}\right)\ln(1+an^2))-1\right|$$
$$\leqslant e^{-tn^2}\left(\exp(tn^2-\mathrm{E}\left(\frac{t}{a}\right)\ln(1+an^2))-1\right),$$

car $\mathrm{E}\left(\frac{t}{a}\right)\ln(1+an^2)) \leqslant \frac{t}{a}\cdot an^2 \leqslant tn^2$. De la minoration, valable pour tout $x \geqslant 0$, $\ln(1+x) \geqslant x - \frac{x^2}{2}$, on déduit que

$$\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}-e^{-tn^2}\right| \leqslant e^{-tn^2}\left(\exp(tn^2-\mathrm{E}\left(\frac{t}{a}\right)\left(an^2-\frac{a^2n^4}{2}\right)-1\right)$$
$$\leqslant e^{-tn^2}\left(e^{\left(an^2\left(\frac{t}{a}-\mathrm{E}\left(\frac{t}{a}\right)\right)+\mathrm{E}\left(\frac{t}{a}\right)\frac{a^2n^4}{2}\right)}-1\right)$$
$$\leqslant e^{-tn^2}\left(e^{\left(an^2+\frac{tan^4}{2}\right)}-1\right).$$

La majoration $e^u - 1 \leqslant ue^u$, valable pour $u \geqslant 0$ conduit à

$$\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}-e^{-tn^2}\right| \leqslant e^{an^2-tn^2\left(1-\frac{an^2}{2}\right)}\left(an^2+\frac{tan^4}{2}\right).$$

Pour $a \leqslant \frac{1}{n^2}$, on obtient

$$\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}-e^{-tn^2}\right| \leqslant e^{an^2-\frac{1}{2}tn^2}\left(an^2+\frac{an^2}{2}e^{\frac{tn^2}{2}}\right),$$

car $e^u \geqslant u$ pour $u \geqslant 0$. Finalement on obtient

$$\left|\frac{1}{(1+an^2)^{\mathrm{E}(\frac{t}{a})}}-e^{-tn^2}\right| \leqslant e^{an^2}\left(\frac{3}{2}an^2\right)$$

et donc, pour $t \geqslant 0$, $x \in \mathbb{R}$ et $a \leqslant \frac{1}{n^2}$,

$$|u_n(a,t,x)-v_n(t,x)| \leqslant 2|c_n(f_0)|e^{an^2}(an^2).$$

Revenons à l'inégalité

$$|\mathrm{F}_a(t,x)-\mathrm{G}(t,x)| \leqslant \sum_{|n|\leqslant \mathrm{N}}|u_n(a,t,x)-v_n(t,x)| + 2\sum_{|n|>\mathrm{N}}|c_n(f_0)|.$$

Soit $\varepsilon > 0$. On choisit N tel que $2 \sum_{|n|>N} |c_n(f_0)| \leqslant \varepsilon$. On prend ensuite $a \leqslant \frac{1}{N^2}$. On a alors, d'après ce qui précède, pour $t \geqslant 0$, $x \in \mathbb{R}$ et $|n| \leqslant N$, $|u_n(a,t,x) - v_n(t,x)| \leqslant 2|c_n(f_0)|e^{aN^2}(aN^2)$ et donc

$$\sum_{|n|\leqslant N} |u_n(a,t,x) - v_n(t,x)| \leqslant e^{aN^2}(aN^2) \sum_{|n|\leqslant N} |c_n(f_0)|.$$

Pour a assez petit, on aura $\sum_{|n|\leqslant N} |u_n(a,t,x) - v_n(t,x)| \leqslant \varepsilon$ et donc $|F_a(t,x) - G(t,x)| \leqslant 3\varepsilon$, ceci pour tous $t \geqslant 0$ et $x \in \mathbb{R}$.

Conclusion. F_a converge uniformément vers la fonction G.

Une fonction f continue et 2π-périodique est de classe $\mathcal{C}^\infty$, si et seulement pour tout $k \in \mathbb{N}$, on a $c_n = o\left(\frac{1}{|n|^k}\right)$. Nous avons déjà vu que la condition est nécessaire. Réciproquement, si cette condition est réalisée, la fonction est égale à la somme de sa série de Fourier et elle est $\mathcal{C}^\infty$ d'après le théorème de dérivation des séries de fonctions. Dans l'exercice suivant est énoncée une condition suffisante portant sur les coefficients de Fourier pour que la fonction soit analytique, c'est-à-dire localement développable en série entière.

4.24. Condition suffisante d'analycité

Soit $f \in \mathcal{C}^\infty(\mathbb{R},\mathbb{C})$ une fonction 2π-périodique. Montrer que s'il existe $\lambda > 0$ et $C > 0$ tels que $|c_n(f)| \leqslant C\exp(-\lambda|n|)$ pour tout entier $n \in \mathbb{Z}$, alors f est analytique, c'est-à-dire pour tout $x_0 \in \mathbb{R}$ développable en série entière au voisinage de x_0.

(**École normale supérieure**)

▷ **Solution.**

Supposons qu'il existe $\lambda > 0$ et $C > 0$ tels que, pour tout $n \in \mathbb{Z}$, on ait $|c_n(f)| \leqslant C\exp(-\lambda|n|)$. Comme f est de classe $\mathcal{C}^1$, d'après le théorème de convergence normale, f est somme de sa série de Fourier. On a donc, pour tout $x \in \mathbb{R}$.

$$f(x) = \sum_{n\in\mathbb{Z}} c_n(f)e^{inx}.$$

Soit $x_0 \in \mathbb{R}$. Afin de développer f en série entière au voisinage de x_0, commençons par le faire pour les fonctions $x \longmapsto e^{inx}$. On obtient, pour tout $x \in \mathbb{R}$,

$$e^{inx} = e^{inx_0}e^{in(x-x_0)} = e^{inx_0} \sum_{k=0}^{+\infty} \frac{(in)^k(x-x_0)^k}{k!}.$$

On pose, pour $(n,k) \in \mathbb{N}\times\mathbb{Z}$, $a_{n,k} = c_n(f)e^{inx_0}\dfrac{(in)^k(x-x_0)^k}{k!}$. Montrons que la série double $\sum a_{n,k}$ est sommable. Pour tout $n \in \mathbb{Z}$, on a

$$\sum_{k=0}^{+\infty} |a_{n,k}| = |c_n(f)| \sum_{k=0}^{+\infty} \frac{|n(x-x_0)|^k}{k!} = |c_n(f)|e^{|n(x-x_0)|}.$$

On a par hypothèse, pour tout $n \in \mathbb{Z}$,

$$|c_n(f)|e^{|n(x-x_0)|} \leqslant Ce^{-\lambda|n|}e^{|n(x-x_0)|} \leqslant Ce^{(-\lambda+|x-x_0|)|n|}.$$

On en déduit que si $|x-x_0| < \lambda$, la série $\sum |c_n(f)|e^{|n(x-x_0)|}$ converge et donc la série double $\sum a_{n,k}$ est sommable. Sa somme est égale à

$$\sum_{n\in\mathbb{Z}} \left(\sum_{k\in\mathbb{N}} a_{n,k} \right) = \sum_{n\in\mathbb{Z}} c_n(f)e^{inx} = f(x).$$

La série double étant sommable, on peut intervertir l'ordre des sommations et on trouve, pour $|x-x_0| < \lambda$,

$$f(x) = \sum_{k\in\mathbb{N}} \left(\sum_{n\in\mathbb{Z}} a_{n,k} \right) = \sum_{k\in\mathbb{N}} \frac{(x-x_0)^k}{k!} \left(\sum_{n\in\mathbb{Z}} (in)^k c_n(f)e^{inx_0} \right).$$

La fonction f est donc développable en série entière au voisinage de x_0. ◁

Le prochain thème abordé concerne l'étude de la convergence vers une fonction 2π-périodique de sa série de Fourier ou plus généralement d'une suite de polynômes trigonométriques. L'exercice suivant étudie la non-convergence uniforme de la série de Fourier vers la fonction, au voisinage d'un point de discontinuité de celle-ci : c'est le phénomène de Gibbs.

4.25. Le phénomène de Gibbs

1. Calculer les coefficients de la fonction 2π-périodique f qui est affine sur $[0, 2\pi[$ et vérifie $f(0) = 0$, $\lim\limits_{x\to(2\pi)^-} f(x) = 2\pi$.

2. Soit s_n la somme partielle d'indice n de sa série trigonométrique. Déterminer les extrema de s_n. Soit m_n le premier minimum de s_n sur $[0, 2\pi[$. Étudier le comportement de la suite (m_n). Interpréter.

(École polytechnique)

▷ **Solution.**

1. On a, pour tout $t \in [0, 2\pi[$, $f(t) = t$. On obtient

$$c_0(f) = \frac{1}{2\pi}\int_0^{2\pi} t\,\mathrm{d}t = \pi \text{ et, pour } n \in \mathbb{Z}^*,$$

$$c_n(f) = \frac{1}{2\pi}\int_0^{2\pi} te^{-int}\mathrm{d}t = \frac{1}{2\pi}\left(\left[\frac{te^{-int}}{-in}\right]_0^{2\pi} + \int_0^{2\pi}\frac{e^{-int}}{in}\mathrm{d}t\right) = \frac{i}{n}\cdot$$

On en déduit que $a_0(f) = 2\pi$ et, pour $n \in \mathbb{N}^*$,

$$a_n(f) = 0 \text{ et } b_n(f) = -\frac{2}{n}\cdot$$

2. Il résulte de la question 1 que la fonction s_n est définie par

$$s_n(t) = \pi - 2\sum_{k=1}^{n}\frac{\sin kt}{k}\cdot$$

Étudions les extrema de s_n sur $[0, 2\pi[$. On a, pour tout réel t,

$$s_n'(t) = -2\sum_{k=1}^{n}\cos kt = -2\,\mathrm{Re}\left(\sum_{k=1}^{n}e^{ikt}\right).$$

On obtient $s_n'(0) = -2n$ et, pour $t \in]0, 2\pi[$,

$$s_n'(t) = -2\,\mathrm{Re}\left(e^{it}\frac{e^{int}-1}{e^{it}-1}\right) = -2\,\mathrm{Re}\left(e^{i(\frac{n+1}{2})t}\frac{\sin\left(\frac{nt}{2}\right)}{\sin\left(\frac{t}{2}\right)}\right)$$

$$= -2\frac{\cos\left(\frac{n+1}{2}t\right)\sin\left(\frac{nt}{2}\right)}{\sin\left(\frac{t}{2}\right)} = 1 - \frac{\sin\left(n+\frac{1}{2}\right)t}{\sin\frac{t}{2}}\cdot$$

La dérivée de s_n' s'annule et change de signe si $\cos\left(\frac{n+1}{2}t\right) = 0$ ou $\sin\left(\frac{nt}{2}\right) = 0$, c'est-à-dire pour $t = \frac{(2k+1)\pi}{n+1}$ $(0 \leqslant k \leqslant n)$ ou $t = \frac{2k\pi}{n}$ $(1 \leqslant k \leqslant n-1)$. Le premier extremum de s_n sur $[0, 2\pi[$ est obtenu pour $t = \frac{\pi}{n+1}\cdot$ C'est un minimum, car s_n' est négative au voisinage de 0 $(s_n'(0) = -2n)$. On a donc $m_n = s_n\left(\frac{\pi}{n+1}\right)$. Pour exprimer m_n, partons de s_n', qui est continue sur $[0, 2\pi[$ et donnons une forme intégrale à s_n. On a $s_n(0) = \pi$. On en déduit que, pour $t \in [0, 2\pi[$,

$$s_n(t) = \pi + \int_0^t s_n'(x)\mathrm{d}x = \pi + t - \int_0^t \frac{\sin\left(n+\frac{1}{2}\right)x}{\sin\frac{x}{2}}\mathrm{d}x.$$

En particulier, on a $m_n = \pi + \dfrac{\pi}{n+1} - \displaystyle\int_0^{\frac{\pi}{n+1}} \frac{\sin\left(n+\frac{1}{2}\right)x}{\sin\frac{x}{2}}\,\mathrm{d}x$. Pour calculer la limite de l'intégrale $\mathrm{I}_n = \displaystyle\int_0^{\frac{\pi}{n+1}} \frac{\sin\left(n+\frac{1}{2}\right)x}{\sin\frac{x}{2}}\,\mathrm{d}x$, on fait le changement de variable $u = (n+\frac{1}{2})x$. On obtient

$$\mathrm{I}_n = \int_0^{\frac{(n+\frac{1}{2})\pi}{(n+1)}} \frac{\sin u}{(n+\frac{1}{2})\sin\frac{u}{2n+1}}\,\mathrm{d}u.$$

Considérons la fonction f_n définie sur $]0,\pi[$ par

$$f_n(x) = \frac{\sin x}{(n+\frac{1}{2})\sin\frac{x}{2n+1}} \text{ si } x \leqslant \frac{(n+\frac{1}{2})\pi}{n+1} \text{ et } 0 \text{ si } x > \frac{(n+\frac{1}{2})\pi}{n+1}.$$

Les fonctions f_n sont intégrables sur $]0,\pi[$ et pour tout $x \in]0,\pi[$, on a $\lim\limits_{n\to+\infty} f_n(x) = 2\dfrac{\sin x}{x}$; la fonction $x \longmapsto \dfrac{\sin x}{x}$ est intégrable sur $]0,\pi[$. Enfin, on a une relation de domination. En effet, on obtient, pour $n \geqslant 1$ et $0 < x \leqslant \dfrac{(n+\frac{1}{2})\pi}{n+1}$, en utilisant la relation $\sin(x) \geqslant \dfrac{2}{\pi}x$, qui est valable sur $[0, \frac{\pi}{2}]$,

$$0 \leqslant f_n(x) \leqslant \pi\frac{\sin x}{x}.$$

Cette inégalité est vérifiée sur tout l'intervalle $]0,\pi[$. Le théorème de convergence dominée permet de conclure que $\lim\limits_{n\to+\infty} \mathrm{I}_n = 2\displaystyle\int_0^{\pi} \frac{\sin x}{x}\,\mathrm{d}x$. On en déduit enfin que

$$\lim_{n\to+\infty} m_n = \pi - 2\int_0^{\pi} \frac{\sin x}{x}\,\mathrm{d}x.$$

La fonction f étant de classe $\mathcal{C}^1$ par morceaux sur $\mathbb{R}$, sa série de Fourier converge en tout point t vers $\dfrac{1}{2}(f(t^+) + f(t^-))$. On a donc ici, pour $t \notin 2\pi\mathbb{Z}$, $\lim\limits_{n\to+\infty} s_n(t) = f(t)$ et $\lim\limits_{n\to+\infty} s_n(0) = \dfrac{1}{2}(0+2\pi) = \pi$. La convergence n'est uniforme sur aucun intervalle d'intérieur non vide contenant 0. En effet, on a $\lim\limits_{n\to+\infty} \dfrac{\pi}{n+1} = 0$. Si la convergence était uniforme, on aurait $\lim\limits_{n\to+\infty} m_n = \lim\limits_{n\to+\infty} s_n(0) = \pi$. Quand tend vers l'infini, $\dfrac{\pi}{n+1}$ tend vers 0, mais la différence entre $s_n(0)$ et le minimum de s_n sur

$\left[0, \dfrac{\pi}{n+1}\right]$ tend vers une limite finie égale à $2\displaystyle\int_0^\pi \frac{\sin x}{x}\,dx$ (soit environ 3,7). Ce défaut de convergence au voisinage d'un point de discontinuité de f est appelé *phénomène de Gibbs.* ◁

Les deux exercices suivants visent à démontrer que toute fonction continue 2π-périodique de $\mathbb{R}$ dans $\mathbb{C}$ est limite uniforme d'une suite de polynômes trigonométriques (équivalent du théorème d'approximation de Weierstrass pour les fonctions continues sur un segment). La méthode est la même dans les deux cas : on considère la suite de terme général $f_n = f * u_n$, obtenue en convolant f avec une suite de polynômes trigonométriques u_n, positifs, dont l'intégrale sur $[0, 2\pi]$ est égal à 1 et tels que, pour tout $\delta \in]0, \pi[$, le maximum de u_n sur $[-\pi, -\delta] \cup [\delta, \pi]$ tende vers 0 quand n tend vers $+\infty$. On montre alors que la suite (f_n) converge uniformément vers f. On remarque enfin que f_n est un polynôme trigonométrique pour tout n. En effet si u_n est de degré p et s'écrit $u_n(x) = \displaystyle\sum_{k=-p}^{p} a_{k,n} e^{ikx}$. on a, pour tout réel x,

$$\begin{aligned} f_n(x) &= \int_0^{2\pi} f_n(t) u_n(x-t)\,dt = \sum_{k=-p}^{p} a_{k,n} e^{ikx} \int_0^{2\pi} f(t) e^{-ikt}\,dt \\ &= 2\pi \sum_{k=-p}^{p} a_{k,n} c_k(f) e^{ikx}. \end{aligned}$$

4.26. Théorème de Weierstrass trigonométrique

Soit E l'ensemble des applications de $\mathbb{R}$ dans $\mathbb{C}$ continues et 2π-périodiques. Pour f et g dans E, on considère l'application $f * g$ définie par $f * g(x) = \displaystyle\int_0^{2\pi} f(x-t)g(t)\mathrm{d}t$.

1. Montrer que $f * g$ est élément de E. Que dire si f est $\mathcal{C}^\infty$?

2. Pour $n \in \mathbb{N}$, on considère l'application $u_n : \mathbb{R} \to \mathbb{R}$ définie par $u_n(x) = c_n(1 + \cos x)^n$, où $c_n \in \mathbb{R}$ est choisi tel que $\displaystyle\int_0^{2\pi} u_n = 1$, et on pose $f_n = f * u_n$. Que peut-on dire de la suite (f_n) ?

(École polytechnique)

▷ **Solution.**

1. La fonction $\varphi : (x,t) \longmapsto f(x-t)g(t)$ est définie et continue sur $\mathbb{R}^2$. On en déduit, d'après le théorème de continuité sous le signe $\int$, que $f*g$ est définie et continue sur $\mathbb{R}$. Il est clair qu'elle est 2π-périodique. Elle est donc élément de E. Si f est de classe $\mathcal{C}^1$, la fonction φ est dérivable

par rapport à x et la fonction $\dfrac{\partial\varphi}{\partial x}(x,t) \longmapsto f'(x-t)g(t)$ est continue sur $\mathbb{R}^2$. Du théorème de dérivation sous le signe $\int$, on déduit que $f*g$ est de classe $\mathcal{C}^1$ et que $(f*g)'(x) = \int_0^{2\pi} f'(x-t)g(t)\mathrm{d}t$. Une démonstration par récurrence immédiate montre que si f est $\mathcal{C}^\infty$, $f*g$ est également $\mathcal{C}^\infty$ et que, pour tout $n \in \mathbb{N}$, $(f*g)^{(n)} = f^{(n)} * g$.

Du changement de variable $u = x - t$, on tire

$$f*g(x) = \int_{x-2\pi}^{x} f(u)g(x-u)\mathrm{d}u = \int_0^{2\pi} f(u)g(x-u)\mathrm{d}u = g*f(x).$$

On a donc $f*g = g*f$. Si g est $\mathcal{C}^\infty$, il en est de même de $f*g$.

2. Montrons que la suite (f_n) converge uniformément vers f. Puisque par définition on a, pour tout $n \in \mathbb{N}$, $\int_0^{2\pi} u_n(t)\mathrm{d}t = 1$, on peut écrire, pour tout $x \in \mathbb{R}$,

$$\begin{aligned}
|f_n(x) - f(x)| &= \left| \int_0^{2\pi} f(x-t)u_n(t)\mathrm{d}t - f(x)\int_0^{2\pi} u_n(t)\mathrm{d}t \right| \\
&= \left| \int_{-\pi}^{\pi} (f(x-t) - f(x))u_n(t)\mathrm{d}t \right| \\
&\leqslant \int_{-\pi}^{\pi} |f(x-t) - f(x)|u_n(t)\mathrm{d}t,
\end{aligned}$$

puisque u_n est positive.

Donnons-nous $\varepsilon > 0$. Comme f est continue et 2π-périodique, elle est uniformément continue. Il existe $\eta \in]0,\pi[$ tel que $|f(s) - f(t)| \leqslant \varepsilon$ si $|s-t| \leqslant \eta$. On coupe l'intégrale en trois morceaux, les intégrales sur $[-\pi,-\eta]$, sur $[-\eta,\eta]$ et sur $[\eta,\pi]$. On obtient

$$\int_{-\eta}^{\eta} |f(x-t) - f(x)|u_n(t)\mathrm{d}t \leqslant \varepsilon \int_{-\eta}^{\eta} u_n(t)\mathrm{d}t \leqslant \varepsilon \int_{-\pi}^{\pi} u_n(t)\mathrm{d}t = \varepsilon \quad \text{et}$$

$$\begin{aligned}
\int_{-\pi}^{-\eta} |f(x-t) - f(x)|u_n(t)\mathrm{d}t + \int_{\eta}^{\pi} |f(x-t) - f(x)|u_n(t)\mathrm{d}t \\
\leqslant 4\|f\|_\infty \int_{\eta}^{\pi} u_n(t)\mathrm{d}t,
\end{aligned}$$

puisque la fonction u_n est paire. Il reste à majorer cette dernière intégrale par une quantité tendant vers 0 quand n tend vers $+\infty$. Comme cos, la fonction u_n décroît sur $[\eta,\pi]$. On a donc

$$\int_{\eta}^{\pi} u_n(t)\mathrm{d}t \leqslant \pi u_n(\eta) \leqslant \pi c_n(1+\cos\eta)^n.$$

Mais on a, par définition,

$$\frac{1}{c_n} = \int_0^{2\pi} (1+\cos t)^n \mathrm{d}t = 2\int_0^{\pi} (1+\cos t)^n \mathrm{d}t \geqslant 2\int_0^{\pi} (1+\cos t)^n \sin t \mathrm{d}t$$
$$= 2\left[\frac{-(1+\cos t)^{n+1}}{n+1}\right]_0^{\pi} = 2\cdot\frac{2^{n+1}}{n+1}.$$

On obtient finalement $c_n \leqslant \dfrac{n+1}{2^{n+2}}$, $\displaystyle\int_\eta^\pi u_n(t)\mathrm{d}t \leqslant \frac{\pi(n+1)}{4}\left(\frac{1+\cos\eta}{2}\right)^n$

$$\text{et}\quad |f_n(x) - f(x)| \leqslant \varepsilon + \pi\|f\|_\infty (n+1)\left(\frac{1+\cos\eta}{2}\right)^n.$$

Puisque $\dfrac{1+\cos\eta}{2} < 1$, on a $\displaystyle\lim_{n\to+\infty}(n+1)\left(\frac{1+\cos\eta}{2}\right)^n = 0$. On en déduit que l'on a $|f_n(x) - f(x)| \leqslant 2\varepsilon$, pour tout réel x, si n est assez grand. Ceci démontre que la suite (f_n) converge uniformément vers f sur $\mathbb{R}$. ◁

Comme cela est expliqué avant l'énoncé de l'exercice on vérifie facilement que f_n est un polynôme trigonométrique. L'exercice a donc montré que toute fonction de E *est limite uniforme sur $\mathbb{R}$ d'une suite de polynômes trigonométriques (théorème de Weierstrass trigonométrique).*

Se pose alors le problème de la mesure de la qualité de l'approximation d'une fonction 2π-périodique continue par les polynômes trigonométriques. De manière précise, si $f : \mathbb{R} \to \mathbb{C}$ est 2π-périodique continue et si $\mathcal{T}_n$ est l'espace des polynômes trigonométriques de degré $\leqslant n$ on cherche à étudier à quelle vitesse la suite $\delta_n(f) = \inf_{\mathrm{P}\in\mathcal{T}_n} \|f - \mathrm{P}\|_\infty$ converge vers 0. Le théorème de Jackson qui fait l'objet de l'exercice ci-après redémontre le théorème de Weierstrass trigonométrique et fournit une première réponse à cette question.

4.27. Théorème de Jackson (1911)

On pose $\mathrm{J}_n(t) = c_n\left(\dfrac{\sin\dfrac{(n+1)t}{2}}{\sin\dfrac{t}{2}}\right)^4$, la constante c_n étant choisie de telle sorte que $\displaystyle\int_{-\pi}^{\pi} \mathrm{J}_n = 1$ (J_n est appelé *le noyau de Jackson* d'indice n). On note E l'ensemble des applications de $\mathbb{R}$ dans $\mathbb{C}$ continues, 2π-périodique. Pour $g \in \mathrm{E}$ et $h > 0$, on pose

$$\omega_h(g) = \max_{|t-s|\leqslant h} |g(s) - g(t)| \quad \text{et} \quad \mathrm{T}_n g(x) = \int_{-\pi}^{\pi} \mathrm{J}_n(x-t)g(t)\mathrm{d}t.$$

1. Montrer qu'il existe $a > 0$ tel que $\int_{-\pi}^{\pi} |t|\mathrm{J}_n(t)\mathrm{d}t \leqslant \dfrac{a}{n+1}$.

2. Montrer qu'il existe $b > 0$ tel que, pour tout $g \in \mathrm{E}$ et $n \in \mathbb{N}^*$ on ait

$$\|\mathrm{T}_n g - g\|_\infty \leqslant b\,\omega_{\frac{2\pi}{n}}(g).$$

3. Conclure que la suite $(\mathrm{T}_n g)$ converge uniformément vers g.

(École normale supérieure)

▷ **Solution.**

1. La fonction f_n définie sur $[-\pi, \pi]$ par $f_n(t) = \left(\dfrac{\sin\frac{(n+1)t}{2}}{\sin\frac{t}{2}}\right)^4$ si $t \neq 0$ et $f_n(0) = (n+1)^4$ est continue sur $[-\pi, \pi]$ et paire. Par définition, on a $\dfrac{1}{c_n} = \int_{-\pi}^{\pi} f_n(t)\mathrm{d}t = 2\int_0^{\pi} f_n(t)\mathrm{d}t$ et donc

$$\int_{-\pi}^{\pi} |t|\mathrm{J}_n(t)\mathrm{d}t = c_n \int_{-\pi}^{\pi} |t| f_n(t)\mathrm{d}t = \frac{\int_0^{\pi} t f_n(t)\mathrm{d}t}{\int_0^{\pi} f_n(t)\mathrm{d}t}.$$

Nous allons majorer le numérateur N_n et minorer le dénominateur D_n de ce quotient, en partant de l'encadrement classique : $\dfrac{2}{\pi}t \leqslant \sin t \leqslant t$ pour $t \in \left[0, \dfrac{\pi}{2}\right]$. On obtient

$$\mathrm{N}_n \leqslant \int_0^{\pi} t \left(\frac{\sin\frac{(n+1)t}{2}}{\frac{t}{\pi}}\right)^4 \mathrm{d}t = \pi^4 \int_0^{\pi} \frac{\sin^4\frac{(n+1)t}{2}}{t^3}\mathrm{d}t$$

et

$$\mathrm{D}_n = \frac{1}{2c_n} \geqslant \int_0^{\pi} \left(\frac{\sin\frac{(n+1)t}{2}}{\frac{t}{2}}\right)^4 \mathrm{d}t.$$

Le changement de variable $u = \dfrac{(n+1)t}{2}$ conduit à

$$\mathrm{N}_n \leqslant \pi^4 \left(\frac{n+1}{2}\right)^2 \int_0^{\frac{n+1}{2}\pi} \frac{\sin^4 u}{u^3}\mathrm{d}u \leqslant k_1(n+1)^2,$$

où $k_1 = \dfrac{\pi^4}{4}\int_0^{+\infty} \dfrac{\sin^4 u}{u^3}\,\mathrm{d}u > 0$. Le même changement de variable donne

$$\mathrm{D}_n \geqslant 2(n+1)^3 \int_0^{(n+1)\frac{\pi}{2}} \frac{\sin^4 u}{u^4}\mathrm{d}u \geqslant k_2(n+1)^3$$

où $k_2 = 2\int_0^{\frac{\pi}{2}} \frac{\sin^4 u}{u^4}\,\mathrm{d}u > 0$. On note que les fonctions qui interviennent dans les intégrales précédentes sont toutes prolongeables par continuité en 0, ce qui assure leur intégrabilité sur les intervalles bornés. De plus, la fonction $u \longmapsto \frac{\sin^4 u}{u^3}$ est majorée par $\frac{1}{u^3}$, ce qui assure son intégrabilité sur $[1, +\infty[$.

Finalement, on obtient, pour tout $n \in \mathbb{N}$, $\int_{-\pi}^{\pi} |t|\mathrm{J}_n(t)\mathrm{d}t \leqslant \frac{\frac{k_1}{k_2}}{n+1}$, ce qui est le résultat voulu avec $a = \frac{k_1}{k_2}$.

2. Soit g un élément de E. Puisque, par hypothèse, $\int_{-\pi}^{\pi} \mathrm{J}_n(t)\mathrm{d}t = 1$, on peut écrire, pour tout réel x et $n \in \mathbb{N}^*$, $g(x) = \int_{-\pi}^{\pi} g(x)\mathrm{J}_n(t)\mathrm{d}t$. Le changement de variable $u = x - t$ et la périodicité de g et J_n permettent d'écrire

$$\mathrm{T}_n g(x) = \int_{x-\pi}^{x+\pi} \mathrm{J}_n(u)g(x-u)\mathrm{d}u = \int_{-\pi}^{\pi} \mathrm{J}_n(u)g(x-u)\mathrm{d}u.$$

On obtient donc

$$\mathrm{T}_n g(x) - g(x) = \int_{-\pi}^{\pi} \mathrm{J}_n(t)(g(x-t) - g(x))\mathrm{d}t$$

$$\text{puis} \quad |\mathrm{T}_n g(x) - g(x)| \leqslant \int_{-\pi}^{\pi} \mathrm{J}_n(t)|g(x-t) - g(x)|\mathrm{d}t.$$

On cherche une majoration de $|g(x-t) - g(x)|$ faisant intervenir $\omega_{\frac{2\pi}{n}}(g)$. Pour tout entier $\mathrm{N} \in \mathbb{N}^*$, on peut écrire

$$|g(x-t) - g(x)| \leqslant \sum_{i=1}^{\mathrm{N}} \left| g\left(x - i\frac{t}{\mathrm{N}}\right) - g\left(x - (i-1)\frac{t}{\mathrm{N}}\right)\right|.$$

Si N vérifie $\left|\frac{t}{\mathrm{N}}\right| \leqslant \frac{2\pi}{n}$, soit $\mathrm{N} \geqslant \frac{n}{2\pi}|t|$, on obtient

$$|g(x-t) - g(x)| \leqslant \mathrm{N}\omega_{\frac{2\pi}{n}}(g).$$

On prend $\mathrm{N} = \mathrm{E}\left(\frac{n}{2\pi}|t|\right) + 1$, ce qui donne $\mathrm{N} \leqslant \frac{n}{2\pi}|t| + 1$ et donc

$$|g(x-t) - g(x)| \leqslant \left(\frac{n}{2\pi}|t| + 1\right)\omega_{\frac{n}{2\pi}}(g).$$

On en déduit alors, en utilisant la majoration obtenue dans la question précédente et l'égalité $\int_{-\pi}^{\pi} \mathrm{J}_n(t)\mathrm{d}t = 1$, que, pour tout réel x,

$$\begin{aligned}|\mathrm{T}_n g(x) - g(x)| &\leqslant \omega_{\frac{n}{2\pi}}(g) \int_{-\pi}^{\pi} \left(\frac{n}{2\pi}|t| + 1\right) \mathrm{J}_n(t)\mathrm{d}t \\ &\leqslant \omega_{\frac{n}{2\pi}}(g) \left(\frac{n}{2\pi}\frac{a}{n+1} + 1\right) \\ &\leqslant \omega_{\frac{n}{2\pi}}(g) \left(\frac{a}{2\pi} + 1\right).\end{aligned}$$

Finalement, en posant $b = \dfrac{a}{2\pi} + 1 > 0$, on a

$$\|\mathrm{T}_n g - g\|_\infty \leqslant b\,\omega_{\frac{2\pi}{n}}(g).$$

On note que la constante b ne dépend ni de n, ni du choix de $g \in \mathrm{E}$.

3. Toute fonction g de E étant continue et 2π-périodique est uniformément continue sur $\mathbb{R}$. On en déduit que $\lim\limits_{n\to+\infty} \omega_{\frac{2\pi}{n}}(g) = 0$. En effet, pour tout $\varepsilon > 0$, considérons η un module d'uniforme continuité de f pour ε et un entier naturel n_0 tel que $\dfrac{2\pi}{n} \leqslant \eta$ pour $n \geqslant n_0$. Si $n \geqslant n_0$, alors pour $|s-t| \leqslant \dfrac{2\pi}{n}$, on a $|s-t| \leqslant \eta$ et donc $|g(s) - g(t)| \leqslant \varepsilon$. On en déduit que $\omega_{\frac{2\pi}{n}}(g) \leqslant \varepsilon$, ce qui démontre le résultat voulu. Comme la suite $(b\omega_{\frac{2\pi}{n}}(g))$ converge vers 0, la suite $(\mathrm{T}_n g)$ converge uniformément vers g sur $\mathbb{R}$. ◁

Il est facile de vérifier que J_n *est un polynôme trigonométrique de degré* $2n$. *En effet, on part de*

$$\frac{\sin(n+1)\frac{t}{2}}{\sin\frac{t}{2}} = \frac{e^{i(n+1)\frac{t}{2}} - e^{-i(n+1)\frac{t}{2}}}{e^{i\frac{t}{2}} - e^{-i\frac{t}{2}}} = e^{-ni\frac{t}{2}} \frac{e^{i(n+1)t} - 1}{e^{it} - 1} = e^{-ni\frac{t}{2}} \sum_{k=0}^{n} e^{ikt}.$$

En élevant au carré, on obtient

$$\begin{aligned}\left(\frac{\sin(n+1)\frac{t}{2}}{\sin\frac{t}{2}}\right)^2 &= e^{-nit}\left(\sum_{k=0}^{n} e^{ikt}\right)^2 = e^{-nit} \sum_{k=0}^{2n} (n+1-|k-n|)e^{ikt} \\ &= \sum_{k=-n}^{n} (n+1-|k|)e^{ikt}.\end{aligned}$$

En élevant de nouveau au carré, on obtient bien un polynôme trigonométrique de degré $2n$ *et ainsi* J_n *appartient à* $\mathcal{T}_{2n}$.

Prenons alors g *une fonction* 2π*-périodique et* α*-höldérienne (avec* $\alpha \in]0,1]$*), c'est-à-dire pour laquelle il existe* $\mathrm{K} > 0$ *tel que, pour tous réels* x *et* y, $|g(x) - g(y)| \leqslant \mathrm{K}|x-y|^\alpha$. *Reprenons les notations données dans l'introduction de l'exercice en posant* $\delta_n(f) = \inf\limits_{\mathrm{P}\in\mathcal{T}_n} \|g - \mathrm{P}\|_\infty$.

Comme $T_n g$ *est un polynôme trigonométrique de degré* $\leqslant 2n$, *la question 2. de l'exercice ci-dessus montre que* $\delta_{2n}(g) \leqslant b\mathrm{K}\left(\frac{2\pi}{n}\right)^{\alpha}$. *autrement dit que* $\delta_{2n}(g) = O(1/n^{\alpha})$. *Comme la suite* $\delta_n(g)$ *est décroissante, on a aussi* $\delta_n(g) = O(1/n^{\alpha})$. *C'est une mesure très précise de la qualité de l'approximation de* g. *Un an après les résultats de Jackson, Bernstein établit une réciproque de cela : si* $\delta_n(g) = O(1/n^{\alpha})$ *alors* g *est* α*-höldérienne. C'est l'objet de l'exercice suivant. Notons que c'est dans cette étude que Bernstein a prouvé l'inégalité qui porte son nom et qui a été vue dans l'exercice 4.3.*

4.28. Un théorème de Bernstein (1912)

On note $\mathcal{T}_n$ l'espace des polynômes trigonométriques de degré inférieur ou égal à n. Soit f une fonction continue et 2π-périodique de $\mathbb{R}$ dans $\mathbb{C}$. On pose $\delta_n(f) = \inf_{\mathrm{P}\in\mathcal{T}_n} \|f - \mathrm{P}\|_\infty$. Soit $\alpha \in]0,1[$. On suppose qu'il existe $\mathrm{C} > 0$ tel que, pour tout $n \in \mathbb{N}^*$, $\delta_n(f) \leqslant \mathrm{C}n^{-\alpha}$.

1. Montrer que $\delta_n(f)$ est atteint.

2. On considère, pour tout $n \in \mathbb{N}$, un polynôme P_n dans $\mathcal{T}_{2^n}$ tel que $\|\mathrm{P}_n - f\|_\infty = \delta_{2^n}(f)$. Montrer qu'il existe $\mathrm{A} > 0$ tel que, pour tout $n \in \mathbb{N}^*$, $\|\mathrm{P}_n'\|_\infty \leqslant \mathrm{A}2^{(1-\alpha)n}$ (on pourra utiliser *l'inégalité de Bernstein* $\|\mathrm{P}'\|_\infty \leqslant n\|\mathrm{P}\|_\infty$, si $\mathrm{P} \in \mathcal{T}_n$).

3. Montrer que f est α-höldérienne, c'est-à-dire qu'il existe un réel $\mathrm{K} > 0$ tel que, pour tous réels x et y, $|f(x) - f(y)| \leqslant \mathrm{K}|x - y|^{\alpha}$.

(**École normale supérieure**)

▷ **Solution.**

1. Cela résulte de ce que $\mathcal{T}_n$ est de dimension finie. Pour justifier que $\delta_n(f)$ est atteint, on montre qu'on peut se limiter à faire varier P dans un compact de $\mathcal{T}_n$, c'est-à-dire, puisque $\mathcal{T}_n$ est de dimension finie, dans un fermé borné.

Toute fonction continue et 2π-périodique est bornée sur $\mathbb{R}$. On a, pour tout $\mathrm{P} \in \mathcal{T}_n$, $\|f - \mathrm{P}\|_\infty \geqslant \|\mathrm{P}\|_\infty - \|f\|_\infty$. Ainsi, si $\|\mathrm{P}\|_\infty > \|f\|_\infty + \delta_n(f)$, on a $\|f - \mathrm{P}\|_\infty > \delta_n(f)$. En notant $\mathrm{B} = \{\mathrm{P} \in \mathcal{T}_n, \|\mathrm{P}\|_\infty \leqslant \|f\|_\infty + \delta_n(f)\}$, on obtient donc $\delta_n(f) = \inf_{\mathrm{P}\in\mathrm{B}} \|f - \mathrm{P}\|_\infty$. L'ensemble B étant compact et non vide, cette borne inférieure est atteinte.

2. L'existence de P_n résulte de la question 1. Suivant l'indication de l'énoncé, on utilise l'inégalité de Bernstein. On a, pour $n \in \mathbb{N}$,

$$\begin{aligned}\|\mathrm{P}_{n+1} - \mathrm{P}_n\|_\infty &\leqslant \|\mathrm{P}_{n+1} - f\|_\infty + \|\mathrm{P}_n - f\|_\infty \leqslant \delta_{2^{n+1}}(f) + \delta_{2^n}(f)\\ &\leqslant 2\delta_{2^n}(f) \leqslant \mathrm{C}2^{1-n\alpha},\end{aligned}$$

car $\delta_{2^{n+1}}(f) \leqslant \delta_{2^n}(f)$. On en déduit que

$$\|P'_{n+1} - P'_n\|_\infty \leqslant 2^{n+1} \cdot C2^{1-n\alpha} \leqslant 4C2^{n(1-\alpha)},$$

puis que

$$\|P'_n\|_\infty \leqslant \|P'_0\|_\infty + \sum_{k=0}^{n-1} \|P'_{k+1} - P'_k\|_\infty \leqslant \|P'_0\|_\infty + \sum_{k=0}^{n-1} 4C2^{k(1-\alpha)}$$
$$\leqslant \|P'_0\|_\infty + 4C\frac{(2^{1-\alpha})^n - 1}{2^{1-\alpha} - 1} \leqslant \|P'_0\|_\infty + \frac{4C}{2^{1-\alpha} - 1}2^{n(1-\alpha)}.$$

Puisque $2^{(1-\alpha)n}$ tend vers $+\infty$ quand n tend vers $+\infty$, la suite $\left(\frac{\|P'_0\|_\infty}{2^{(1-\alpha)n}}\right)$ est majorée (par B) et on a, pour tout $n \in \mathbb{N}$,

$$\|P'_n\|_\infty \leqslant A2^{(1-\alpha)n}, \quad \text{où} \quad A = B + \frac{4C}{2^{1-\alpha} - 1}.$$

3. Il s'agit de trouver $K > 0$ tel que $|f(x) - f(y)| \leqslant K|x - y|^\alpha$ pour tout couple $(x, y) \in \mathbb{R}^2$. Il faut donc majorer $\frac{|f(x) - f(y)|}{|x - y|^\alpha}$ quand $x \neq y$.

Pour les « grandes» valeurs de $|x - y|$, cela ne pose pas de problème. Si, par exemple $|x - y| \geqslant 1$, on a

$$\frac{|f(x) - f(y)|}{|x - y|^\alpha} \leqslant |f(x) - f(y)| \leqslant 2\|f\|_\infty.$$

Supposons désormais $0 < |x - y| < 1$. On utilise le résultat de la question précédente, l'inégalité des accroissements finis pour majorer $|P_n(x) - P_n(y)|$, et la proximité de f et P_n. Pour tout $n \in \mathbb{N}$, on a

$$|f(x) - f(y)| \leqslant |f(x) - P_n(x) + |f(y) - P_n(y)| + |P_n(x) - P_n(y)|$$
$$\leqslant 2\|P_n - f\|_\infty + \|P'_n\|_\infty|x - y|$$
$$\leqslant 2\delta_{2^n}(f) + A2^{(1-\alpha)n}|x - y| \leqslant C2^{1-n\alpha} + A2^{(1-\alpha)n}|x - y|.$$

On en déduit que

$$\frac{|f(x) - f(y)|}{|x - y|^\alpha} \leqslant 2C\left(\frac{1}{2^n|x - y|}\right)^\alpha + A(2^n|x - y|)^{1-\alpha}.$$

Choisissons $n \in \mathbb{N}$ tel que $2^n|x - y| \in \left[\frac{1}{2}, 1\right[$, *i.e.* $\frac{1}{2^{n+1}} \leqslant |x - y| < \frac{1}{2^n}$. C'est possible car $|x - y| \in]0, 1[$. On a alors

$$\frac{|f(x) - f(y)|}{|x - y|^\alpha} \leqslant 2C2^\alpha + A.$$

Finalement, on obtient, pour tout $(x, y) \in \mathbb{R}^2$,

$$|f(x) - f(y)| \leqslant K|x - y|^\alpha,$$

avec $K = \max(2\|f\|_\alpha, C2^{1+\alpha} + A)$. Ceci démontre que la fonction f est α-höldérienne. $\lhd$

L'exercice suivant est à rapprocher des exercices 4.26 et 4.27. On convole f avec le noyau de Poisson qui dépend d'un paramètre réel et plus d'un paramètre discret comme les noyaux de Fejér ou de Jackson. Il ne s'agit plus non plus d'un polynôme trigonométrique. On utilise essentiellement ce noyau de Poisson pour relier l'analyse de Fourier à la théorie des fonctions analytiques.

4.29. Noyau de Poisson

Soit $f : \mathbb{R} \longmapsto \mathbb{C}$ continue par morceaux et 2π-périodique. On note $(c_n)_{n\in\mathbb{Z}}$ la suite de ses coefficients de Fourier exponentiels.

Pour $r \in [0, 1[$, on pose $f_r(t) = \sum_{n=-\infty}^{+\infty} r^{|n|} c_n e^{int}$.

1. Déterminer φ_r telle que l'on ait

$$\forall t \in \mathbb{R}, f_r(t) = \frac{1}{2\pi} \int_0^{2\pi} f(u)\varphi_r(t-u)\mathrm{d}u.$$

La fonction φ_r est appelée *noyau de Poisson*.

2. Montrer que φ_r converge uniformément vers 0 lorsque r tend vers 1^- sur tout segment inclus dans $]0, 2\pi[$.

3. Montrer que, pour tout réel x,

$$\lim_{r\to 1^-} f_r(x) = \frac{1}{2}(f(x^+) + f(x^-))$$

(où $f(x^+)$ (resp. $f(x^-)$) désigne la limite à droite (resp. à gauche) en x).

4. Montrer que si f est continue, f_r converge uniformément vers f lorsque r tend vers 1^-.

(École polytechnique)

▷ **Solution.**

1. La fonction f_r est bien définie sur $\mathbb{R}$ et continue car la série qui la définit est normalement convergente. On a, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} f_r(t) &= \sum_{n=-\infty}^{+\infty} r^{|n|} e^{int} \frac{1}{2\pi} \int_0^{2\pi} f(u)e^{-inu}\mathrm{d}u \\ &= \frac{1}{2\pi} \sum_{n=-\infty}^{+\infty} \int_0^{2\pi} f(u) r^{|n|} e^{in(t-u)}\mathrm{d}u. \end{aligned}$$

La fonction f est continue par morceaux donc bornée sur $[0, 2\pi]$ et donc sur $\mathbb{R}$, puisqu'elle est 2π-périodique. On en déduit que, pour tout $n \in \mathbb{Z}$ et $(t.u) \in \mathbb{R}^2$,

$$|f(u)r^{|n|}e^{in(t-u)}| \leqslant \|f\|_\infty r^{|n|}.$$

La série $\sum f(u)r^{|n|}e^{in(t-u)}$ converge normalement donc uniformément par rapport à u et sa somme est continue par morceaux. On peut donc intervertir intégration et sommation. On obtient, pour tout $t \in \mathbb{R}$,

$$f_r(t) = \frac{1}{2\pi}\int_0^{2\pi} f(u) \sum_{n=-\infty}^{+\infty} e^{in(t-u)}r^{|n|}\mathrm{d}u.$$

En posant pour tout $t \in \mathbb{R}$,

$$\boxed{\varphi_r(t) = \sum_{n=-\infty}^{\infty} r^{|n|}e^{int},}$$

série qui converge normalement car, pour tout $n \in \mathbb{N}$, $|r^{|n|}e^{int}| = r^{|n|}$, on obtient, pour tout $t \in \mathbb{R}$,

$$f_r(t) = \frac{1}{2\pi}\int_0^{2\pi} f(u)\varphi_r(t-u)\mathrm{d}u.$$

2. Transformons l'expression de φ_r. Pour tout $t \in \mathbb{R}$, on a

$$\begin{aligned}\varphi_r(t) &= \sum_{n=0}^{+\infty} r^n e^{int} + \sum_{n=0}^{+\infty} r^n e^{-nit} - 1 = \frac{1}{1-re^{it}} + \frac{1}{1-re^{-it}} - 1\\ &= \frac{2-2r\cos t - 1 - r^2 + 2r\cos t}{1+r^2-2r\cos t} = \frac{1-r^2}{1+r^2-2r\cos t}\\ &= \frac{1-r^2}{(r-\cos t)^2+\sin t^2} > 0.\end{aligned}$$

Soit $[a, b] \subset]0, 2\pi[$. La fonction cos atteint sa borne supérieure $\lambda < 1$ sur le segment $[a, b]$. On a donc, pour tout $t \in [a, b]$,

$$0 \leqslant \varphi_r(t) \leqslant \frac{1-r^2}{1+r^2-2\lambda r} \xrightarrow[r\to 1^-]{} 0.$$

Ceci montre que φ_r converge uniformément vers 0 sur $[a, b]$.

3. En faisant le changement de variable $t = x - u$, on obtient, pour tout $x \in \mathbb{R}$,

$$\begin{aligned}f_r(x) &= \frac{1}{2\pi}\int_0^{2\pi} f(u)\varphi_r(x-u)\mathrm{d}u = \frac{1}{2\pi}\int_{x-2\pi}^{x} f(x-t)\varphi_r(t)\mathrm{d}t\\ &= \frac{1}{2\pi}\int_0^{2\pi} f(x-t)\varphi_r(t)\mathrm{d}t,\end{aligned}$$

car φ_r et f sont 2π-périodiques.

Soit $\varepsilon > 0$. On a

$$\lim_{t\to 0^+} f(x-t) = f(x_-) \text{ et } \lim_{t\to 2\pi^-} f(x-t) = \lim_{t\to 0^-} f(x-t) = f(x_+).$$

On peut donc trouver $\alpha \in]0,\alpha[$ tel que

$$\begin{cases} 0 < t \leqslant \alpha & \Longrightarrow \quad |f(x-t) - f(x_-)| \leqslant \varepsilon \\ 2\pi - \alpha \leqslant t < 2\pi & \Longrightarrow \quad |f(x-t) - f(x_+)| \leqslant \varepsilon. \end{cases}$$

On note que $\displaystyle\int_0^{2\pi} \varphi_r(t)\mathrm{d}t = \sum_{n=-\infty}^{+\infty} r^{|n|} \int_0^{2\pi} e^{int}\mathrm{d}t = 2\pi$, car la série converge normalement. On a d'autre part, puisque φ_r converge uniformément vers 0 sur $[\alpha, 2\pi - \alpha]$ quand r tend vers 1^-, on a $\displaystyle\lim_{r\to 1^-} \int_\alpha^{2\pi-\alpha} \varphi_r(t) = 0$. Enfin comme φ_r est paire et 2π-périodique, on a $\displaystyle\int_0^\alpha \varphi_r(t)\mathrm{d}t = \int_{2\pi-\alpha}^{2\pi} \varphi_r(t)\mathrm{d}t$. On déduit de tout ceci que

$$\lim_{r\to 1^-} \int_0^\alpha \varphi_r(t)\mathrm{d}t = \lim_{r\to 1^-} \int_{2\pi-\alpha}^{2\pi} \varphi_r(t)\mathrm{d}t = \pi.$$

On écrit alors

$$f_r(x) - \frac{f(x_+)+f(x_-)}{2} = \frac{1}{2\pi}\left(\int_0^{2\pi} f(x-t)\varphi_r(t)\mathrm{d}t - \pi f(x_+) - \pi f(x_-)\right)$$

$$= \frac{1}{2\pi}\left(\int_0^\alpha (f(x-t) - f(x_-))\varphi_r(t)\mathrm{d}t + \int_{2\pi-\alpha}^{2\pi} (f(x-t) - f(x_+))\varphi_r(t)\mathrm{d}t \right.$$
$$+ f(x_-)\left(\int_0^\alpha \varphi_r(t) - \pi\right)\mathrm{d}t + f(x_+)\left(\int_{2\pi-\alpha}^{2\pi} \varphi_r(t)\mathrm{d}t - \pi\right)$$
$$\left. + \int_\alpha^{2\pi-\alpha} f(x-t)\varphi_r(t)\mathrm{d}t\right).$$

On majore les différents termes. On obtient

$$\left|\int_0^\alpha (f(x-t) - f(x_-))\varphi_r(t)\mathrm{d}t\right| \leqslant \varepsilon \int_0^\alpha \varphi_r \leqslant \varepsilon\pi,$$

car φ_r est positive et $\displaystyle\int_0^\pi \varphi_r = \pi$, et de même,

$$\left|\int_{2\pi-\alpha}^{2\pi} (f(x-t) - f(x_+))\varphi_r(t)\mathrm{d}t\right| \leqslant \varepsilon\pi.$$

On en déduit que

$$\left|f_r(x) - \frac{1}{2}(f(x_+) + f(x_-))\right| \leqslant \varepsilon + \frac{\|f\|_\infty}{2\pi}\left(\left|\int_0^\alpha \varphi_r - \pi\right| + \right.$$
$$\left.\left|\int_{2\pi-\alpha}^{2\pi} \varphi_r - \pi\right| + \int_\alpha^{2\pi-\alpha} \varphi_r\right).$$

D'après ce qui précède, le terme de droite de l'inégalité tend vers ε lorsque r tend vers 1^-. Il existe donc $r_0 \in]0,1[$ tel que, pour $r_0 \leqslant r \leqslant 1$ on a

$$\left| f_r(x) - \frac{1}{2}\left(f(x^+) + f(x^-)\right)\right| \leqslant 2\varepsilon.$$

Conclusion. $\boxed{\lim_{r\to 1^-} f_r(x) = \frac{1}{2}\left(f(x^+) + f(x^-)\right).}$

4. Supposons f continue et reprenons la démonstration faite dans la question précédente. La fonction f étant continue sur $\mathbb{R}$, 2π-périodique, elle est donc uniformément continue sur $\mathbb{R}$. Soit $\varepsilon > 0$ et $\alpha \in]0,\pi[$ un module d'uniforme continuité de f pour ε.

On a alors, pour tout réel x,

$$\begin{cases} 0 < t \leqslant \alpha & \Longrightarrow \ |f(x-t) - f(x)| \leqslant \varepsilon \\ 2\pi - \alpha \leqslant t < 2\pi & \Longrightarrow \ |f(x-t) - f(x)| = |f(x-t) - f(x-2\pi)| \leqslant \varepsilon. \end{cases}$$

On peut reprendre les inégalités précédentes avec $f(x_+) = f(x_-) = f(x)$. Cette fois, α est indépendant de x. On obtient, pour tout $x \in \mathbb{R}$,

$$|f_r(x) - f(x)| \leqslant \varepsilon + \frac{\|f\|_\infty}{2\pi}\left(\left|\int_0^\alpha \varphi_r - \pi\right| + \left|\int_{2\pi-\alpha}^{2\pi} \varphi_r - \pi\right| + \int_\alpha^{2\pi-\alpha} \varphi_r\right).$$

On a donc, pour $r_0 \leqslant r \leqslant 1$ et tout $x \in \mathbb{R}$,

$$|f_r(x) - f(x)| \leqslant 2\varepsilon.$$

Conclusion. f_r converge uniformément vers f lorsque r tend vers 1^-. ◁

Le noyau de Poisson est utilisé dans l'exercice qui suit pour montrer qu'une fonction continue dont tous les coefficients de Fourier exponentiels sont réels et positifs est automatiquement développable en série de Fourier.

4.30. Fonctions à coefficients de Fourier positifs

Soit $f \in \mathcal{C}(\mathbb{R},\mathbb{C})$ une fonction 2π-périodique. On suppose que la suite $(c_n(f))_{n\in\mathbb{Z}}$ de ses coefficients de Fourier est réelle positive. Montrer que f est somme de sa série de Fourier. On pourra introduire $\sum_{n\in\mathbb{Z}} c_n(f) r^{|n|}$ pour $0 < r < 1$.

(École normale supérieure)

▷ **Solution.**

Pour $r \in]0,1[$ et $n \in \mathbb{Z}$ on a $0 \leqslant r^{|n|}c_n(f) \leqslant r^{|n|}\|f\|_\infty$ donc la série $\sum_{n\in\mathbb{Z}} r^{|n|}c_n(f)$ converge. On pose $\mathrm{S}(r) = \sum_{n=-\infty}^{+\infty} r^{|n|}c_n(f)$. Nous allons montrer que S est majorée sur $]0,1[$, d'où nous déduirons la convergence de la série $\sum c_n(f)$ et donc la convergence normale de la série de Fourier de f.

On a, pour $r \in]0,1[$,

$$\begin{aligned}\mathrm{S}(r) &= \sum_{n=-\infty}^{+\infty} r^{|n|}c_n(f) = \frac{1}{2\pi}\sum_{n=-\infty}^{+\infty} r^{|n|}\int_0^{2\pi} f(t)e^{-int}\,\mathrm{d}t \\ &= \frac{1}{2\pi}\sum_{n=-\infty}^{+\infty}\int_0^{2\pi} r^{|n|}f(t)e^{-int}\,\mathrm{d}t.\end{aligned}$$

On a, pour tout $t \in [0,2\pi]$, $|r^{|n|}f(t)e^{-int}| \leqslant r^{|n|}\|f\|_\infty$, donc la série de fonctions $\sum r^{|n|}f(t)e^{-int}$ est normalement convergente sur $[0,2\pi]$. On peut donc échanger l'intégration et la sommation. On obtient

$$\mathrm{S}(r) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\sum_{n=-\infty}^{+\infty} r^{|n|}e^{-int}\,\mathrm{d}t.$$

La somme de cette série est calculée dans la question 2 de l'exercice précédent :

$$\begin{aligned}\sum_{n=-\infty}^{+\infty} r^{|n|}e^{-int} &= \sum_{n=0}^{+\infty} r^n e^{-int} + \sum_{n=1}^{+\infty} r^n e^{int} = \frac{1}{1-re^{-it}} + \frac{re^{it}}{1-re^{it}} \\ &= \frac{1-r^2}{|1-re^{it}|^2} = \frac{1-r^2}{1+r^2-2r\cos t}.\end{aligned}$$

On en déduit

$$\mathrm{S}(r) = \frac{1}{2\pi}\int_0^{2\pi} f(t)\frac{1-r^2}{1+r^2-2r\cos t}\,\mathrm{d}t.$$

Par hypothèse les coefficients $c_n(f)$ sont positifs donc $\mathrm{S}(r)$ est positif. D'autre part, pour tout $t \in [0,2\pi]$, $\dfrac{1-r^2}{1+r^2-2r\cos t} = \dfrac{1-r^2}{|1-re^{it}|^2} \geqslant 0$. On en déduit que

$$\begin{aligned}\mathrm{S}(r) &= \frac{1}{2\pi}\left|\int_0^{2\pi} f(t)\frac{1-r^2}{1+r^2-2r\cos t}\,\mathrm{d}t\right| \\ &\leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(t)|\frac{1-r^2}{1+r^2-2r\cos t}\,\mathrm{d}t \\ &\leqslant \frac{\|f\|_\infty}{2\pi}\int_0^{2\pi}\frac{1-r^2}{1+r^2-2r\cos t}\,\mathrm{d}t.\end{aligned}$$

Pour calculer l'intégrale il suffit d'utiliser la formule donnant $S(r)$ avec la fonction f constante égale à 1. Il est clair que dans ce cas $S(r) = 1$. Ainsi, pour tout $r \in]0,1[$,

$$\frac{1}{2\pi}\int_0^{2\pi} \frac{1-r^2}{1+r^2-2r\cos t}\,\mathrm{d}t = 1$$

et on a

$$S(r) \leqslant \|f\|_\infty.$$

Soit $N \in \mathbb{N}$. Comme la série est à termes positifs, on a, pour tout $r \in]0,1[$,

$$\sum_{n=-N}^{N} c_n(f) r^{|n|} \leqslant S(r) \leqslant \|f\|_\infty.$$

En faisant tendre r vers 1, on en déduit $\sum_{n=-N}^{N} c_n(f) \leqslant \|f\|_\infty$. Cela étant vrai pour tout $N \in \mathbb{N}$, on en déduit que la série à terme positifs $\sum c_n(f)$ est convergente, puis que la série de Fourier $\sum c_n(f)e^{nit}$ est absolument convergente. Sa somme est alors nécessairement égale à f (l'argument est classique et a été développé avant l'exercice 4.10). ◁

Le lecteur sait que si f est une fonction 2π-périodique et continue, la série de Fourier $S_n(f)$ ne converge pas forcément. Le mathématicien Fejér a été l'un des premiers à donner un exemple d'une telle fonction. Mais aussitôt il a obtenu un résultat positif qui se révèle très utile : la moyenne de Cesàro des sommes partielles de la série de Fourier de f, c'est-à-dire $\dfrac{S_1(f)+\cdots+S_n(f)}{n}$, converge uniformément vers cette fonction. Combiné au théorème de Cesàro, cela montre que si la suite $S_n(f)(x)$ converge en un réel x, sa limite ne peut être que $f(x)$. L'énoncé suivant propose la démonstration du théorème de Fejér (dans le cadre de fonctions 1-périodiques). L'idée est d'écrire la moyenne des sommes partielles sous la forme $f * K_n$, convolée de f et du noyau de Fejér K_n. Celui-ci (déjà rencontré dans l'exercice 4.2) a la propriété fondamentale d'être positif.

4.31. Théorème de Fejér (1904)

Soit f une application continue 1-périodique de $\mathbb{R}$ dans $\mathbb{C}$.

On pose, pour $n \in \mathbb{N}^*$, $K_n(x) = \sum_{k=-n}^{n} \left(1 - \dfrac{|k|}{n}\right) e^{2i\pi kx}$ (*noyau de Fejér d'indice n*).

On note S_n la somme partielle d'indice n de la série de Fourier de f et $f_n = \frac{1}{n}(S_0 + S_1 + \cdots + S_{n-1})$.

1. Montrer que, pour tout $n \in \mathbb{N}^*$, on a

$$f_n(x) = \int_0^1 K_n(x-y)f(y)dy.$$

2. Montrer que la suite (f_n) converge uniformément vers f.

(École normale supérieure)

▷ **Solution.**

1. Notons c_k, pour $k \in \mathbb{Z}$, les coefficients de Fourier exponentiels de f. On a, pour tout $n \in \mathbb{N}$,

$$S_n(x) = \sum_{k=-n}^{n} c_k e^{2i\pi kx} = \sum_{k=-n}^{n} \int_0^1 f(y)e^{2i\pi k(x-y)}\,dy$$
$$= \int_0^1 f(y)\left(\sum_{k=-n}^{n} e^{2i\pi k(x-y)}\right)dy.$$

On en déduit que

$$f_n(x) = \frac{1}{n}\int_0^1 f(y)\left(\sum_{k=0}^{n-1}\left(\sum_{\ell=-k}^{k} e^{2i\pi\ell(x-y)}\right)\right)dy.$$

Considérons la fonction L_n définie par $L_n(x) = \frac{1}{n}\sum_{k=0}^{n-1}\left(\sum_{\ell=-k}^{k} e^{2i\pi\ell x}\right)$. En échangeant les deux sommations, on obtient

$$L_n(x) = \frac{1}{n}\sum_{\ell=-(n-1)}^{n-1}(n-|\ell|)e^{2i\pi\ell x} = K_n(x).$$

On a donc bien $f_n(x) = \int_0^1 K_n(x-y)f(y)\,dy$.

2. Montrons que K_n est positive et vérifie $\int_0^1 K_n(x)dx = 1$.

En notant, pour tout $k \in \mathbb{N}$, $d_k(x) = \sum_{\ell=-k}^{k} e^{2i\pi\ell x}$, il résulte de la question 1 qu'on a, pour tout $n \in \mathbb{N}^*$, $K_n(x) = \frac{1}{n}\sum_{k=0}^{n-1} d_k(x)$. Supposons $e^{2i\pi x} \neq 1$. On a alors, pour tout $k \in \mathbb{N}^*$,

$$d_k(x) = e^{-ikx}\frac{e^{2i\pi(2k+1)x}-1}{e^{2i\pi}-1} = \frac{e^{2i\pi(k+1)x}-e^{-2i\pi kx}}{e^{2i\pi x}-1}.$$

On en déduit que

$$\begin{aligned}
\mathrm{K}_n(x) &= \frac{1}{n(e^{2i\pi x}-1)} \sum_{k=0}^{n-1} (e^{2i\pi(k+1)x} - e^{-2i\pi kx}) \\
&= \frac{1}{n(e^{2i\pi x}-1)} \left(\frac{e^{2i\pi x}(e^{2i\pi nx}-1)}{e^{2i\pi x}-1} - \frac{e^{-2i\pi nx}-1}{e^{-2i\pi x}-1} \right) \\
&= \frac{e^{2i\pi x}\,(e^{2i\pi nx} + e^{-2i\pi nx} - 2)}{n(e^{2i\pi x}-1)^2} = \frac{2\cos(2\pi nx) - 2}{n(2i\sin \pi x)^2} = \frac{\sin^2(\pi nx)}{n\sin^2(\pi x)}.
\end{aligned}$$

Par continuité de K_n, $\mathrm{K}_n(x) = n$ si $e^{2i\pi x} = 1$. La fonction K_n est bien positive.

Pour tout entier k, l'intégrale $\int_0^{2\pi} e^{2i\pi kx}\mathrm{d}x$ est nulle si $k \neq 0$ et vaut 1 sinon. On en déduit que $\int_0^1 \mathrm{K}_n(x)\mathrm{d}x = 1$.

Le changement de variable $t = x - y$ dans l'intégrale qui définit f_n conduit à

$$f_n(x) = \int_{x-1}^{x} \mathrm{K}_n(t) f(x-t)\,\mathrm{d}t = \int_0^1 \mathrm{K}_n(t) f(x-t)\,\mathrm{d}t,$$

car f et K_n sont 1-périodiques. On écrit alors

$$\begin{aligned}
f_n(x) - f(x) &= \int_0^1 \mathrm{K}_n(t) f(x-t)\mathrm{d}t - f(x) \int_0^1 \mathrm{K}_n(t)\mathrm{d}t \\
&= \int_0^1 \mathrm{K}_n(t)(f(x-t) - f(x))\mathrm{d}t.
\end{aligned}$$

Puisque K_n est positive, on en déduit que

$$|f_n(x) - f(x)| \leqslant \int_0^1 \mathrm{K}_n(t)|f(x-t) - f(x)|\mathrm{d}t.$$

La fonction f est uniformément continue sur $\mathbb{R}$, puisque continue et 2π-périodique. Pour $\varepsilon > 0$, soit $\eta \in \left]0, \frac{1}{2}\right[$ tel que $|f(x) - f(y)| \leqslant \varepsilon$ si $|x-y| \leqslant \eta$. On a alors

$$\int_0^\eta \mathrm{K}_n(t)|(f(x-t) - f(x))|\mathrm{d}t \leqslant \varepsilon \int_0^\eta \mathrm{K}_n(t)\mathrm{d}t \leqslant \varepsilon \int_0^1 \mathrm{K}_n(t)\mathrm{d}t \leqslant \varepsilon$$

$$\begin{aligned}
\int_{1-\eta}^1 \mathrm{K}_n(t)|(f(x-t) - f(x))|\mathrm{d}t &= \int_{1-\eta}^1 \mathrm{K}_n(t)|(f(x-t) - f(x-1))|\mathrm{d}t \\
&\leqslant \varepsilon \int_{1-\eta}^1 \mathrm{K}_n(t)\mathrm{d}t \leqslant \varepsilon.
\end{aligned}$$

On majore ensuite l'autre morceau d'intégrale. Pour $t \in [\eta, 1-\eta]$, on a

$$K_n(t) = \frac{\sin^2(\pi n x)}{n \sin^2(\pi t)} \leqslant \frac{1}{n \sin^2(\pi t)} \leqslant \frac{1}{n \sin^2(\pi \eta)}.$$

On en déduit que

$$\int_\eta^{1-\eta} K_n(t)|F(x-t) - F(x)|\,dt \leqslant \frac{2\|F\|_\infty}{n \sin^2(\pi\eta)}.$$

Cette dernière quantité tend vers 0 quand n tend vers $+\infty$. Elle est inférieure à ε si n est supérieur à un entier naturel n_0. On a alors, si $n \geqslant n_0$, $|F_n(x) - F(x)| \leqslant 3\varepsilon$, pour tout réel x. La suite $(F_n)_{n\in\mathbb{N}}$ converge donc uniformément vers F sur $\mathbb{R}$. ◁

Les exercices suivants s'intéressent à des questions de nature topologique sur l'espace des fonctions continues 2π-périodiques, les deux premiers à des problèmes de complétude.

4.32. Complétude pour la norme quadratique

Soit E l'ensemble des applications de $\mathbb{R}$ dans $\mathbb{C}$ continues et 2π-périodiques. On munit E du produit hermitien défini par

$$\langle f, g\rangle = \int_0^{2\pi} \overline{f(t)} g(t)\,dt.$$

E est-il complet ?

(École polytechnique)

▷ **Solution.**

On note $\|\ \|_2$ la norme issue de ce produit hermitien. La réponse est négative. Une suite de Cauchy de E peut, par exemple, converger au sens de la norme $\|\ \|_2$ vers une fonction non continue.

Soit f une fonction de $\mathbb{R}$ dans $\mathbb{C}$, 2π-périodique, continue par morceaux, et pour laquelle il existe au moins un point $x_0 \in]0, 2\pi[$ en lequel la limite à droite diffère de la limite à gauche. Notons S_n les sommes partielles de sa série de Fourier. On a, pour $0 < n < m$, $\|S_m - S_n\|_2^2 = \sum_{\substack{k\in\mathbb{Z} \\ n<|k|\leqslant m}} |c_k(f)|^2$. La série $\sum_{k\in\mathbb{Z}} |c_k(f)|^2$ étant convergente, on en déduit que (S_n) est une suite de Cauchy. On sait que (S_n) converge vers f au sens de la norme $\|\ \|_2$. Si la suite (S_n) convergeait vers une fonction g de E, on aurait $\|f - g\|_2 = 0$, c'est-à-dire $\int_0^{2\pi} |f(t) - g(t)|^2\,dt = 0$.

Cela impose que $f(t) = g(t)$ en tout point t où f est continue. Mais alors, les limites de f à gauche et à droite en x_0 valent toutes les deux $g(x_0)$ ce qui contredit l'hypothèse faite ci-dessus. ◁

Bien entendu il est possible de donner une solution de cet exercice n'utilisant pas du tout l'analyse de Fourier.

Dans l'énoncé qui suit on s'intéresse à l'espace des fonctions continues 2π-périodiques dont la série de Fourier converge absolument et on montre qu'il s'agit d'une algèbre de Banach.

4.33. Fonctions dont la série des coefficients de Fourier converge absolument

Soit $\mathcal{B}$ l'ensemble des applications continues 2π-périodiques de $\mathbb{R}$ dans $\mathbb{C}$ telle que la série $\sum c_n(f)$ soit absolument convergente (les $c_n(f)$ étant les coefficients de Fourier de f).

1. Montrer que $\mathcal{B}$ est un $\mathbb{C}$-espace vectoriel et que tout élément de $\mathcal{B}$ est somme de sa série de Fourier.

2. Pour $f \in \mathcal{B}$, on pose $\|f\| = \sum_{n\in\mathbb{Z}} |c_n(f)|$. Montrer que l'on définit ainsi une norme pour laquelle $\mathcal{B}$ est complet.

3. Montrer que $(f, g) \in \mathcal{B}^2$ implique $fg \in \mathcal{B}$ et $\|fg\| \leqslant \|f\|\|g\|$.

(École polytechnique)

▷ **Solution.**

1. Montrons que $\mathcal{B}$ est un sous-espace de l'espace vectoriel des fonctions continues, 2π-périodique de $\mathbb{R}$ dans $\mathbb{C}$. Soit f et g deux éléments de $\mathcal{B}$, λ un réel. On a, pour tout $n \in \mathbb{N}$,

$$c_n(f+g) = c_n(f) + c_n(g) \quad \text{et} \quad c_n(\lambda f) = \lambda c_n(f).$$

Par hypothèse, les séries $\sum c_n(f)$ et $\sum c_n(g)$ convergent absolument. Il en est de même de $\sum c_n(f) + c_n(g)$ et $\sum \lambda c_n(f)$. On en déduit que $f+g$ et λf sont dans $\mathcal{B}$. Ainsi $\mathcal{B}$ est un espace vectoriel.

Soit $f \in \mathcal{B}$. On pose, pour tout réel x, $g(x) = \sum_{n\in\mathbb{Z}} c_n(f)e^{inx}$. La série qui définit $g(x)$ est normalement convergente sur $\mathbb{R}$, par hypothèse. La fonction g est donc définie et continue sur $\mathbb{R}$. Calculons ses coefficients de Fourier. On obtient, pour tout $k \in \mathbb{Z}$,

$$\begin{aligned} c_k(g) &= \frac{1}{2\pi}\int_0^{2\pi} \sum_{n\in\mathbb{Z}} c_n(f)e^{i(n-k)x}\mathrm{d}x = \sum_{n\in\mathbb{Z}} \frac{1}{2\pi}\int_0^{2\pi} c_n(f)e^{i(n-k)x}\mathrm{d}x \\ &= c_k(f), \end{aligned}$$

l'interversion de l'intégrale et de la sommation étant justifiée par la convergence normale de la série. Les fonctions f et g ont mêmes coefficients de Fourier. Les coefficients de Fourier de $f-g$ sont donc nuls. La formule de Parseval donne donc $\int_0^{2\pi} |f(x)-g(x)|^2 \mathrm{d}x = 0$. La fonction $f-g$ étant continue, elle est nulle sur $[0,2\pi]$ et donc sur $\mathbb{R}$ tout entier. On a donc $f=g$ et la fonction f est égale à la somme de sa série de Fourier.

2. Pour f et g dans $\mathcal{B}$ et $\lambda \in \mathbb{R}$, on a

$$\|f+g\| = \sum_{n\in\mathbb{Z}} |c_n(f+g)| \leqslant \sum_{n\in\mathbb{Z}} |c_n(f)| + \sum_{n\in\mathbb{Z}} |c_n(g)| = \|f\| + \|g\| \text{ et}$$

$$\|\lambda f\| = \sum_{n\in\mathbb{Z}} |c_n(\lambda f)| = \sum_{n\in\mathbb{Z}} |\lambda||c_n(f)| = |\lambda|\|f\|.$$

Par ailleurs, $\|f\| = 0$ implique, pour tout $n \in \mathbb{Z}$, $c_n(f) = 0$ et donc, d'après la formule de Parseval, $\int_0^{2\pi} |f(x)|^2 \mathrm{d}x = 0$. La fonction f étant continue, elle est nulle sur $[0,2\pi]$ et donc sur $\mathbb{R}$. On a donc bien défini une norme sur $\mathcal{B}$.

Montrons que $\mathcal{B}$ est complet pour cette norme. Soit $(f_k)_{k\in\mathbb{N}}$ une suite de Cauchy de $\mathcal{B}$. On a, pour tout $n \in \mathbb{Z}$ et tout $(k,\ell) \in \mathbb{N}^2$,

$$|c_n(f_k) - c_n(f_\ell)| \leqslant \|f_k - f_\ell\|.$$

On en déduit que, pour tout $n \in \mathbb{Z}$, la suite $(c_n(f_k))_{k\in\mathbb{N}}$ est de Cauchy. Elle est donc convergente, puisque $\mathbb{C}$ est complet. On note c_n sa limite.

Montrons que $\sum |c_n|$ converge. La suite (f_k) est bornée puisqu'elle converge. Soit $\mathrm{M} > 0$ tel que $\|f_k\| \leqslant \mathrm{M}$, pour tout $k \in \mathbb{N}$. Pour $\mathrm{N} \in \mathbb{N}$ et $k \in \mathbb{N}$, on a $\sum_{|n|\leqslant \mathrm{N}} |c_n(f_k)| \leqslant \|f_k\| \leqslant \mathrm{M}$. En faisant tendre k vers l'infini, on obtient $\sum_{|n|\leqslant \mathrm{N}} |c_n| \leqslant \mathrm{M}$. Ceci étant vrai pour tout $\mathrm{N} \in \mathrm{N}$, on en déduit que $\sum |c_n|$ converge. Posons, pour tout $x \in \mathbb{Z}$, $f(x) = \sum_{n\in\mathbb{Z}} c_n e^{inx}$. On montre, comme dans la question 1 qu'on définit ainsi une fonction f continue sur $\mathbb{R}$ et dont les coefficients de Fourier sont les c_n.

Montrons que la suite $(f_k)_{k\in\mathbb{N}}$ converge vers f. Soit $\varepsilon > 0$. Il existe $k_0 \in \mathbb{N}$ tel que $\|f_k - f_\ell\| = \sum_{n\in\mathbb{Z}} |c_n(f_k) - c_n(f_\ell)| \leqslant \varepsilon$ si $k,\ell \geqslant k_0$. Pour tout $\mathrm{N} \in \mathbb{N}$, on a alors $\sum_{|n|\leqslant \mathrm{N}} |c_n(f_k) - c_n(f_\ell)| \leqslant \varepsilon$. En faisant tendre ℓ vers $+\infty$, on obtient $\sum_{|n|\leqslant \mathrm{N}} |c_n(f_k) - c_n| \leqslant \varepsilon$. Ceci étant vrai pour tout $\mathrm{N} \in \mathbb{N}$, on en déduit que $\sum_{n\in\mathbb{Z}} |c_n(f_k) - c_n(f_\ell)| \leqslant \varepsilon$, c'est-à-dire que $\|f_k - f\| \leqslant \varepsilon$, ceci pour tout $k \geqslant k_0$. Cela montre que (f_k) converge vers f.

Tout suite de Cauchy de $\mathcal{B}$ est convergente. L'espace vectoriel normé $\mathcal{B}$ est complet.

On peut rédiger cela bien plus rapidement si on sait que $\ell^1(\mathbb{Z},\mathbb{C})$ l'espace des suites sommables $u=(u_n)_{n\in\mathbb{Z}}$ de nombres complexes muni de la norme $\|u\|_1=\sum_{n\in\mathbb{Z}}|u_n|$ est un espace de Banach. En effet, l'application qui a $f\in\mathcal{B}$ associe la suite $(c_n(f))_{n\in\mathbb{Z}}$ est un isomorphisme d'espaces vectoriels normés.

3. Soit f et g deux éléments de $\mathcal{B}$. On a, pour tout $x\in\mathbb{R}$,

$$f(x)=\sum_{n\in\mathbb{Z}}c_n(f)e^{inx}\quad\text{et}\quad g(x)=\sum_{n\in\mathbb{Z}}c_n(g)e^{inx}.$$

Ces deux séries sont absolument convergentes. La série double $\sum a_{n,m}$ définie par $a_{n,m}=c_n(f)e^{inx}c_m(g)e^{imx}$ est donc sommable et sa somme est $f(x)g(x)$. D'après le théorème d'associativité, on peut sommer pour $m+n=p$ et écrire

$$\begin{aligned}f(x)g(x)&=\sum_{p\in\mathbb{Z}}\left(\sum_{m+n=p}c_n(f)e^{inx}c_m(g)e^{imx}\right)\\&=\sum_{p\in\mathbb{Z}}\left(\sum_{m+n=p}c_n(f)c_m(g)\right)e^{ipx}.\end{aligned}$$

Posons, pour $p\in\mathbb{Z}$, $a_p=\sum_{m+n=p}c_n(f)c_m(g)$. Montrons que les a_p sont les coefficients de Fourier de fg. Les séries $\sum|c_n(f)|$ et $\sum|c_m(g)|$ convergent. On en déduit que la série double $\sum|c_n(f)||c_m(g)|$ est sommable de somme $\|f\|\|g\|$. Comme précédemment, on peut écrire

$$\|f\|\|g\|=\sum_{p\in\mathbb{Z}}\left(\sum_{m+n=p}|c_n(f)||c_m(g)|\right).$$

Comme $|a_p|\leqslant\sum_{m+n=p}|c_n(f)||c_m(g)|$, on en déduit que la série $\sum_p|a_p|$ converge. Sachant que, pour tout $x\in\mathbb{R}$, on a

$$f(x)g(x)=\sum_{p\in\mathbb{Z}}a_pe^{ipx}$$

et que cette série converge normalement, on en déduit, comme dans la question 1, que les a_p sont les coefficients de Fourier de fg, puis que fg est dans $\mathcal{B}$.

Enfin, par définition, on a

$$\|fg\|=\sum_{p\in\mathbb{Z}}|a_p|\leqslant\sum_{p\in\mathbb{Z}}\left(\sum_{m+n=p}|c_n(f)||c_m(g)|\right)=\|f\|\|g\|.$$

La dernière question de l'exercice montre que $\mathcal{B}$ est une algèbre de Banach.

Il n'est pas du tout évident de caractériser l'appartenance d'une fonction à $\mathcal{B}$ par des conditions de régularité. L'exercice suivant donne une condition suffisante : une fonction α-höldérienne avec $\alpha > 1/2$ est dans $\mathcal{B}$. Ce résultat est dû à Bernstein.

4.34. Convergence de la série de Fourier d'une fonction α-höldérienne

Soit f une application 2π-périodique de $\mathbb{R}$ dans $\mathbb{C}$. On suppose qu'il existe $\mathrm{M} > 0$ et $\alpha > 0$ tel qu'on ait, pour tous réels x et y, $|f(x) - f(y)| \leqslant \mathrm{M}|x - y|^\alpha$ (on dit que f est α-höldérienne).

1. Montrer qu'au voisinage de $\pm\infty$, $c_n(f) = O\left(\dfrac{1}{|n|^\alpha}\right)$.

2. Montrer qu'il existe $\mathrm{K} > 0$ tel que, pour tout $h > 0$,

$$\sum_{n\in\mathbb{Z}} |c_n(f)|^2 \sin^2 nh \leqslant \mathrm{K}|h|^{2\alpha}.$$

3. On suppose $\alpha > \dfrac{1}{2}$. Montrer que la série de Fourier de f est absolument convergente. Quelle est sa somme ?

(**École polytechnique**)

▷ **Solution.**

1. Il résulte de l'hypothèse que la fonction f est continue sur $\mathbb{R}$. Soit $h > 0$. Considérons les coefficients de Fourier de la fonction f_h, définie par $f_h(x) = f(x + h) - f(x - h)$, qui est continue et 2π-périodique. On obtient, pour tout $n \in \mathbb{Z}$,

$$\begin{aligned} c_n(f_h) &= \frac{1}{2\pi}\left(\int_0^{2\pi} (f(x+h)e^{-inx}\mathrm{d}x - \int_0^{2\pi} (f(x-h)e^{-inx}\mathrm{d}x\right) \\ &= (e^{inh} - e^{-inh})c_n(f) = 2i\sin(nh)c_n(f). \end{aligned}$$

De l'hypothèse, on déduit alors que

$$\begin{aligned} 2|\sin nh||c_n(f)| &= \frac{1}{2\pi}\left|\int_0^{2\pi} (f(x+h) - f(x-h))e^{-inx}\mathrm{d}x\right| \\ &\leqslant \frac{1}{2\pi}\int_0^{2\pi} |f(x+h) - f(x-h)|\mathrm{d}x \\ &\leqslant \frac{1}{2\pi}\int_0^{2\pi} \mathrm{M}(2h)^\alpha = \mathrm{M}(2h)^\alpha. \end{aligned}$$

En particulier, en prenant $h = \dfrac{1}{|n|}$, on obtient, pour $n \in \mathbb{Z}^*$

$$|c_n(f)| \leqslant \frac{\mathrm{M}}{2\sin 1}\frac{2^\alpha}{|n|^\alpha},$$

ce qui démontre le résultat voulu.

2. On écrit la formule de Parseval pour f_h. On obtient

$$\sum_{n\in\mathbb{Z}} 4|c_n(f)|^2 \sin^2(nh) = \frac{1}{2\pi}\int_0^{2\pi} |f(x+h) - f(x-h)|^2 \mathrm{d}x.$$

Mais on a, par hypothèse

$$\frac{1}{2\pi}\int_0^{2\pi} |f(x+h) - f(x-h)|^2 \mathrm{d}x \leqslant \frac{1}{2\pi}\int_0^{2\pi} \mathrm{M}|2h|^{2\alpha}\mathrm{d}x = \mathrm{M}|2h|^{2\alpha}.$$

Finalement, on obtient

$$\sum_{n\in\mathbb{Z}} |c_n(f)|^2 \sin^2(nh) \leqslant \mathrm{M}2^{2\alpha-2}|h|^{2\alpha},$$

ce qui est l'inégalité voulue avec $\mathrm{K} = \mathrm{M}2^{2\alpha-2}$.

3. Il faut démontrer la convergence de $\sum |c_n(f)|$, en utilisant la série précédente pour une valeur bien choisie de h. Étant donné la présence de carrés, on songe à utiliser l'inégalité de Cauchy-Schwarz.

On a pour $\mathrm{N} \in \mathbb{N}^*$,

$$\begin{aligned}\left(\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}} |c_n(f)|\right)^2 &= \left(\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}} |c_n(f)|\sin(nh)\frac{1}{\sin(nh)}\right)^2\\ &\leqslant \sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}} |c_n(f)|^2\sin^2(nh)\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}}\frac{1}{\sin^2(nh)}\\ &\leqslant \mathrm{K}|h|^{2\alpha}\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}}\frac{1}{\sin^2(nh)},\end{aligned}$$

où h est choisi de telle manière que $\sin(nh) \neq 0$ si $|n| \leqslant \mathrm{N}$ et $n \neq 0$. Il faut ensuite majorer $\dfrac{1}{\sin^2(nh)}$. On peut par exemple prendre $h = \dfrac{\pi}{2\mathrm{N}}$, auquel cas nh est dans $]-\dfrac{\pi}{2}, \dfrac{\pi}{2}[$, pour tout n. Compte tenu de l'inégalité $\sin x \geqslant \dfrac{2}{\pi}x$, pour $x \in [0, \dfrac{\pi}{2}]$, on obtient

$$\left(\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}} |c_n(f)|\right)^2 \leqslant \mathrm{K}h^{2\alpha}\frac{\pi^2}{4}\sum_{\substack{|n|\leqslant \mathrm{N}\\ n\neq 0}}\frac{1}{n^2h^2} \leqslant \frac{\mathrm{K}\pi^2}{2}h^{2\alpha-2}\sum_{n=1}^{+\infty}\frac{1}{n^2}.$$

Quand N tend vers $+\infty$, h tend vers 0 et si $\alpha < 1$ le membre de droite de l'inégalité tend vers $+\infty$. On ne peut rien en conclure. Si $\alpha > 1$, il tend vers 0, ce qui signifie que $\sum_{n\in\mathbb{Z}^*} |c_n(f)| = 0$ et donc $c_n(f) = 0$ si $n \neq 0$. Ceci était prévisible car si on fait tendre y vers x dans la relation fournie par l'énoncé, on obtient $f'(x) = 0$. La fonction f est donc constante.

Nous allons améliorer la méthode proposée en faisant varier h avec n. Plus précisément, nous allons considérer $\sum_{2^N\leqslant|n|<2^{N+1}} |c_n(f)|$ et prendre alors $h = \frac{\pi}{2^{N+2}}$. On obtient, comme précédemment

$$\left(\sum_{2^N\leqslant|n|<2^{N+1}} |c_n(f)|\right)^2 \leqslant \frac{K\pi^2}{2} h^{2\alpha-2} \sum_{2^N\leqslant n<2^{N+1}} \frac{1}{n^2}.$$

La dernière somme contient 2^N termes inférieur à $\left(\frac{1}{2^N}\right)^2$. Elle est inférieure à $\frac{1}{2^N}$. En remplaçant h par sa valeur $\frac{\pi}{2^{N+2}}$, on obtient

$$\left(\sum_{2^N\leqslant|n|<2^{N+1}} |c_n(f)|\right)^2 \leqslant \frac{K\pi^2}{2} \frac{\pi^{2\alpha-2}}{2^{(N+2)(2\alpha-2)}} \frac{1}{2^N} \leqslant \frac{K'}{2^{N(2\alpha-1)}},$$

où $K' = \frac{K\pi^{2\alpha}}{2^{(4\alpha-3)}}$. La série $\sum \frac{1}{2^{N(2\alpha-1)}}$ est convergente, car $2\alpha - 1 > 0$; on en déduit que $\sum |c_n(f)|$ converge. La série de Fourier de f converge donc normalement.

Sa somme ne peut être alors que f. En effet, si on considère la fonction g définie par $g(x) = \sum_{n\in\mathbb{Z}} c_n(f)e^{inx}$, celle-ci est continue, car la convergence est normale et ses coefficients de Fourier sont les $c_n(f)$ (pour les calculer, on peut intégrer terme à terme, puisque la série converge normalement). Les coefficients de Fourier de $f - g$ sont donc nuls et la formule de Parseval montre que $\int_0^{2\pi} |f(x) - g(x)|^2 dx = 0$. Les fonctions f et g étant continues, on en déduit que $g = f$. ◁

Notons que ce résultat est en un sens optimal. Il existe en effet des fonctions $1/2$-höldériennes dont la série de Fourier n'est pas absolument convergente.

L'exercice suivant décrit les formes linéaires continues sur l'espace préhilbertien des fonctions continues 1-périodiques.

4.35. Dual topologique

Soit E l'espace des fonctions 1-périodiques, continues de $\mathbb{R}$ dans $\mathbb{C}$. Soit λ une forme linéaire sur E continue pour la norme $\| \ \|_2$. Montrer qu'il existe une suite complexe $(a_n)_{n\in\mathbb{Z}}$ telle que $\sum |a_n|^2$ converge et

$$\forall f \in \mathrm{E}, \ \ \lambda(f) = \sum_{-\infty}^{+\infty} a_n c_n(f).$$

(École normale supérieure)

▷ **Solution.**

Soit λ une forme linéaire sur E, continue pour $\| \ \|_2$. Pour tout $n \in \mathbb{Z}$, on note e_n la fonction $t \longmapsto e^{2i\pi nt}$ et $a_n = \lambda(e_n)$. On sait que la famille $(e_n)_{n\in\mathbb{Z}}$ est orthonormée.

Soit $f \in \mathrm{E}$. Pour $\mathrm{N} \in \mathbb{N}$, on pose $\mathrm{S_N} = \sum_{n=-\mathrm{N}}^{\mathrm{N}} c_n(f)e_n$. On sait que la suite $(\mathrm{S_N})_{\mathrm{N}\geqslant 0}$ converge au sens de la norme $\| \ \|_2$ vers f. On a donc, par continuité de λ,

$$\lambda(f) = \lim_{\mathrm{N}\to+\infty} \lambda(\mathrm{S_N}) = \lim_{\mathrm{N}\to+\infty} \sum_{n=-\mathrm{N}}^{\mathrm{N}} a_n c_n(f).$$

La série $\sum a_n c_n(f)$ converge car, pour $0 \leqslant m \leqslant n$, on a

$$\left|\sum_{m}^{n} a_k c_k(f)\right| = \left|\lambda\left(\sum_{k=m}^{n} c_k(f)e_k\right)\right| \leqslant |\!|\!|\lambda|\!|\!| \ \left\|\sum_{k=m}^{n} c_k(f)e_k\right\|_2$$
$$\leqslant |\!|\!|\lambda|\!|\!| \sqrt{\sum_{k=m}^{n} |c_k(f)|^2}$$

et $\sum_{k=m}^{n} |c_k(f)|^2$ tend vers 0 quand m et n tendent vers $+\infty$, car la série $\sum |c_n(f)|^2$ converge. Ainsi, la série $\sum a_k c_k(f)$ converge. On montre de même que la série $\sum a_{-k} c_{-k}(f)$ converge. On peut donc écrire

$$\lambda(f) = \sum_{n=-\infty}^{+\infty} a_n c_n(f).$$

Il reste à démontrer que $\sum |a_n|^2$ converge. Pour $\mathrm{N} \in \mathbb{N}$, on considère la fonction $f_\mathrm{N} = \sum_{n=-\mathrm{N}}^{\mathrm{N}} \overline{a_n} e_n$. On a, d'une part,

$$\lambda(f_N) = \sum_{n=-N}^{N} \overline{a_n}\lambda(e_n) = \sum_{n=-N}^{N} |a_n|^2 = |\lambda(f_N)|$$

et d'autre part

$$|\lambda(f_N)| \leqslant |||\lambda||| \, \|f_N\|_2 \leqslant |||\lambda|||\sqrt{\sum_{n=-N}^{N} |a_n|^2}.$$

On en déduit que

$$\sqrt{\sum_{n=-N}^{N} |a_n|^2} \leqslant |||\lambda|||.$$

Cela étant vrai pour tout $N \in \mathbb{N}$, $\sum |a_n|^2$ converge et

$$\sqrt{\sum_{n=-\infty}^{\infty} |a_n|^2} \leqslant |||\lambda|||. \quad \lhd$$

On peut en fait montrer que $|||\lambda||| = \sqrt{\sum_{n=-\infty}^{\infty} |a_n|^2} = \|(a_n)\|_2$ *(norme dans* $\ell_2(\mathbb{Z})$*). En effet, pour tout* $f \in \mathrm{E}$*, on a d'après l'inégalité de Cauchy-Schwarz et la formule de Parseval,*

$$\begin{aligned} |\lambda(f)| &\leqslant \sqrt{\sum_{n=-\infty}^{\infty} |a_n|^2}\sqrt{\sum_{n=-\infty}^{\infty} |c_n|^2} \\ &\leqslant \|(a_n)\|_2 \, \|f\|_2 \end{aligned}$$

et donc $|||\lambda||| \leqslant \|(a_n)\|_2$.

Réciproquement, toute suite (a_n) *de* $\ell_2(\mathbb{Z})$ *définit une application linéaire continue de* E*. On pose pour tout* $f \in \mathrm{E}$, $\lambda(f) = \sum_{n=-\infty}^{+\infty} a_n c_n(f)$*. Comme* $(c_n(f)) \in \ell_2(\mathbb{Z})$, $\sum |a_n c_n(f)|$ *converge d'après l'inégalité de Cauchy-Schwarz et* λ *est bien définie. On montre en reprenant la démonstration précédente que* λ *est continue et* $|||\lambda||| = \|(a_n)\|_2$*. L'application* $(a_n) \in \ell_2(\mathbb{Z}) \longmapsto \lambda \in \mathcal{L}_c(\mathrm{E}, \mathbb{C})$ *est linéaire et surjective. Elle conserve la norme donc elle est injective.*

Nous avons vu l'importance de la convolution dans les problèmes d'approximation (exercices 4.26, 4.27, 4.29 et 4.31). Le dernier exercice de ce chapitre étudie les propriétés des opérateurs de convolution $f \longmapsto k * f$*, définis sur l'espace vectoriel* E *des fonctions continues de*

$\mathbb{R}$ dans $\mathbb{C}$, 2π-périodiques, muni de la norme de la convergence quadratique : compacité de l'opérateur, propriété des éléments propres. Un bon cadre pour la définition des opérateurs compacts est celui des espaces de Hilbert (ce que n'est pas E, *d'après l'exercice 4.32 ; il faudrait se placer dans le complété de* E*). Les propriétés spectrales démontrées dans la question 5 sont générales. Si* L *est un opérateur compact entre deux espaces de Hilbert, l'ensemble des valeurs propres est soit fini, soit une suite convergeant vers* 0*, et tout sous-espace propre relatif à une valeur propre non nulle est de dimension finie.*

4.36. Opérateurs de convolution

On considère l'espace E des fonctions continues de $\mathbb{R}$ dans $\mathbb{C}$ et 2π-périodiques muni de la norme $\| \ \|_2$ de la convergence quadratique. Pour $k \in$ E, on pose, pour toute fonction f de E,

$$(\mathrm{K}f)(x) = \int_0^{2\pi} k(x-t)f(t)\mathrm{d}t.$$

1. Montrer que K est un endomorphisme continu de E.

2. Déterminer les coefficients de Fourier de $\mathrm{K}f$.

3. *Compacité de l'opérateur* : montrer que si $(f_n)_{n\in\mathbb{N}}$ est une suite bornée de E, on peut extraire de $(\mathrm{K}f_n)_{n\in\mathbb{N}}$ une suite convergente.

4. Montrer que tout espace propre autre que le noyau de K est de dimension finie.

5. Plus généralement, soit L un endomorphisme continu de E tel que pour toute suite bornée $(f_n)_{n\in\mathbb{N}}$ de E on puisse extraire de $(\mathrm{L}f_n)_{n\in\mathbb{N}}$ une suite convergente.

a. Montrer que tout sous-espace propre autre que le noyau de L est de dimension finie et que 0 est valeur d'adhérence de toute suite $(\lambda_n)_{n\in\mathbb{N}}$ de valeurs propres deux à deux distinctes.

b. Conclure que toute suite $(\lambda_n)_{n\in\mathbb{N}}$ de valeurs propres de L deux à deux distinctes converge vers 0.

(**École normale supérieure**)

▷ **Solution.**

1. On rappelle que la norme $\| \ \|_2$ est définie par $\|f\|_2 = \sqrt{\int_0^{2\pi} |f|^2}$. On reconnaît dans $\mathrm{K}f$ le produit de convolution de k et de f. Les propriétés en sont classiques .

De la continuité de la fonction $(x,t) \longmapsto k(x-t)f(t)$ sur $\mathbb{R}^2$, on déduit

que $\mathrm{K}f$ est continue sur $\mathbb{R}$; elle est aussi 2π-périodique car k l'est. La fonction $\mathrm{K}f$ est donc un élément de E.

Il est clair que l'application K est linéaire. Reste à montrer qu'elle est continue. En appliquant l'inégalité de Cauchy-Schwarz, on obtient, pour tout réel x,

$$|\mathrm{K}f(x)|^2 \leqslant \int_0^{2\pi} |k(x-t)|^2 \mathrm{d}t \int_0^{2\pi} |f(t)|^2 \mathrm{d}t.$$

Puisque k est 2π-périodique, on a

$$\int_0^{2\pi} |k(x-t)|^2 \mathrm{d}t = \int_{x-2\pi}^{x} |k(u)|^2 \mathrm{d}u = \int_0^{2\pi} |k(t)|^2 \mathrm{d}t = \|k\|_2^2 .$$

On a donc $|\mathrm{K}f(x)|^2 \leqslant \|k\|_2^2 \, \|f\|_2^2$, pour tout réel x et en intégrant,

$$\|\mathrm{K}f\|_2^2 \leqslant 2\pi \, \|k\|_2^2 \, \|f\|_2^2, \text{ c'est-à-dire } \quad \|\mathrm{K}f\|_2 \leqslant \sqrt{2\pi} \, \|k\|_2 \, \|f\|_2 .$$

L'endomorphisme K est continu et $\|\mathrm{K}\| \leqslant \sqrt{2\pi} \, \|k\|_2$.

2. Pour tout élément f de E, on note $c_n(f)$ $(n \in \mathbb{Z})$ les coefficients de Fourier de f. Pour calculer les coefficients de Fourier de $\mathrm{K}f$, on applique le théorème de Fubini. Pour tout $n \in \mathbb{Z}$, on a

$$\begin{aligned} c_n(\mathrm{K}f) &= \frac{1}{2\pi} \int_0^{2\pi} \mathrm{K}f(x) e^{-inx} \mathrm{d}x \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{2\pi} k(x-t) f(t) \mathrm{d}t \right) e^{-inx} \mathrm{d}x \\ &= \frac{1}{2\pi} \iint\limits_{[0,2\pi]^2} k(x-t) f(t) e^{-inx} \mathrm{d}x \mathrm{d}t \\ &= \frac{1}{2\pi} \int_0^{2\pi} \left(\int_0^{2\pi} k(x-t) e^{-inx} \mathrm{d}x \right) f(t) \mathrm{d}t. \end{aligned}$$

On transforme

$$\begin{aligned} \int_0^{2\pi} k(x-t) e^{-inx} \mathrm{d}x &= e^{-int} \int_0^{2\pi} k(x-t) e^{-in(x-t)} \mathrm{d}x \\ &= e^{-int} \int_0^{2\pi} k(x) e^{-inx} \mathrm{d}x = 2\pi e^{-int} c_n(k) \end{aligned}$$

et on obtient finalement $c_n(\mathrm{K}f) = \int_0^{2\pi} c_n(k) f(t) e^{-int} \mathrm{d}t = 2\pi c_n(k) c_n(f)$.

3. Remarquons que si la suite (f_n) converge vers f, alors chaque suite $(c_p(f_n))_{n\in\mathbb{N}}$ converge vers $c_p(f)$: en effet, on a d'après le théorème de Parseval

$$|c_p(f_n) - c_p(f)|^2 \leqslant \sum_{q\in\mathbb{Z}} |c_q(f_n) - c_q(f)|^2 = \frac{1}{2\pi} \|f_n - f\|_2^2 \xrightarrow[n\to+\infty]{} 0.$$

De plus, par continuité de K, la suite $(\mathrm{K}f_n)$ converge vers $\mathrm{K}f$ et ses coefficients de Fourier sont égaux à $2\pi c_p(k)c_p(f)$. Cela nous indique comment choisir les coefficients de la valeur d'adhérence de la suite $(\mathrm{K}f_n)$ dans le cas général.

Soit $(f_n)_{n\in\mathbb{N}}$ une suite à valeurs dans E, majorée en norme par M. D'après le théorème de Parseval, on a pour tout $n \in \mathbb{N}$, $\sum_{p\in\mathbb{Z}} |c_p(f_n)|^2 = \frac{1}{2\pi}\|f_n\|_2^2 \leqslant \frac{1}{2\pi}\mathrm{M}^2$. Pour tout $p \in \mathbb{Z}$, la suite $(c_p(f_n))_{n\in\mathbb{N}}$ est bornée donc possède une sous-suite convergente. Un procédé diagonal fournit une sous-suite $(f_{\varphi(n)})_{n\in\mathbb{N}}$ telle que pour tout p de $\mathbb{Z}$, la suite $(c_p(f_{\varphi(n)}))_{n\in\mathbb{N}}$ converge (le lecteur pourra se reporter à l'exercice 2.34 pour une présentation de ce procédé). Notons $c_p = \lim_{n\to+\infty} c_p(f_{\varphi(n)})$. On a alors, pour $(n, p_0) \in \mathbb{N}^2$,

$$\sum_{p=-p_0}^{p_0} |c_p(f_{\varphi(n)})|^2 \leqslant \sum_{p=-\infty}^{+\infty} |c_p(f_{\varphi(n)})|^2 \leqslant \frac{1}{2\pi}\mathrm{M}^2.$$

En faisant tendre n vers $+\infty$, on obtient $\sum_{p=-p_0}^{p_0} |c_p|^2 \leqslant \frac{1}{2\pi}\mathrm{M}^2$.

Posons, pour tout $p \in \mathbb{Z}$, $d_p = 2\pi c_p c_p(k)$. On obtient, pour tout $p_0 \in \mathbb{N}$, en appliquant l'inégalité de Cauchy-Schwarz,

$$\sum_{-p_0}^{p_0} |d_p| \leqslant 2\pi\sqrt{\sum_{-p_0}^{p_0} |c_p|^2 \sum_{-p_0}^{p_0} |c_p(k)|^2} \leqslant 2\pi\frac{\mathrm{M}}{\sqrt{2\pi}}\frac{\|k\|_2}{\sqrt{2\pi}} \leqslant \mathrm{M}\|k\|_2.$$

La série de terme général $|d_p|$ est donc convergente. On peut poser, pour tout $x \in \mathbb{R}$, $g(x) = \sum_{p\in\mathbb{Z}} d_p e^{-ipx}$. La série qui définit g converge normalement sur $\mathbb{R}$, donc g est définie, continue et 2π-périodique et ses coefficients de Fourier sont les d_p.

Montrons que la suite $(\mathrm{K}f_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers g. On a, pour $n \in \mathbb{N}$,

$$\begin{aligned}\left\|\mathrm{K}f_{\varphi(n)} - g\right\|_2^2 &= 2\pi\sum_{p\in\mathbb{Z}} |c_p(\mathrm{K}f_{\varphi(n)}) - c_p(g)|^2 \\ &= 2\pi\sum_{p\in\mathbb{Z}} |c_p(\mathrm{K}f_{\varphi(n)}) - d_p|^2 \\ &= (2\pi)^3\sum_{p\in\mathbb{Z}} |c_p(k)|^2|c_p(f_{\varphi(n)}) - c_p|^2.\end{aligned}$$

On sait que pour tout $(n, p) \in \mathbb{N}\times\mathbb{Z}$, $|c_p(f_n)|^2 \leqslant \frac{\mathrm{M}^2}{2\pi}$ et donc $|c_p|^2 \leqslant \frac{\mathrm{M}^2}{2\pi}$. On a alors, pour tout $p_0 \in \mathbb{N}$,

$$\left\|\mathrm{K}f_{\varphi(n)}-g\right\|_2^2 \leqslant (2\pi)^3 \sum_{p=-p_0}^{p_0} |c_p(k)|^2 |c_p(f_{\varphi(n)})-c_p|^2 + (2\pi)^3 \left(\sum_{\substack{p\in\mathbb{Z}\\ |p|>p_0}} |c_p(k)|^2\right) \frac{2\mathrm{M}^2}{2\pi}.$$

Étant donné $\varepsilon > 0$, on choisit $p_0 \in \mathbb{Z}$ tel que $8\pi^2\mathrm{M}^2 \sum_{\substack{p\in\mathbb{Z}\\ |p|>p_0}} |c_p(k)|^2 \leqslant \frac{\varepsilon^2}{2}$.

Puisque $\lim_{n\to+\infty} (2\pi)^3 \sum_{p=-p_0}^{p_0} |c_p(k)|^2 |c_p(f_{\varphi(n)})-c_p|^2 = 0$, il existe n_0 tel que pour $n \geqslant n_0$, $(2\pi)^3 \sum_{p=-p_0}^{p_0} |c_p(k)|^2 |c_p(f_{\varphi(n)})-c_p|^2 \leqslant \frac{\varepsilon^2}{2}$. On obtient enfin, pour $n \geqslant n_0$, $\|\mathrm{K}f_{\varphi(n)}-g\|_2 \leqslant \varepsilon$.

Conclusion. La suite $(\mathrm{K}f_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers g pour la norme $\|\ \|_2$. On a extrait de $(\mathrm{K}f_n)_{n\in\mathbb{N}}$ une suite convergente.

Un opérateur K *tel que, si* $(f_n)_{n\in\mathbb{N}}$ *est une suite bornée alors on peut extraire de* $(\mathrm{K}f_n)_{n\in\mathbb{N}}$ *une suite convergente est appelé un* opérateur compact.

4. Déterminons les valeurs propres et les sous-espaces propres de K afin de montrer que les sous-espaces propres autres que le noyau sont de dimension finie. Soit $\lambda \in \mathrm{Sp}(\mathrm{K})$ et $f \neq 0$ tel que $\mathrm{K}f = \lambda f$. On a alors pour tout $p \in \mathbb{Z}$, $c_p(\mathrm{K}f) = \lambda c_p(f) = 2\pi c_p(k)c_p(f)$. Si $\lambda \neq 2\pi c_p(k)$, alors $c_p(f) = 0$. Si pour tout $p \in \mathbb{Z}$, $\lambda \neq 2\pi c_p(k)$, alors $\|f\|_2^2 = \sum_{p\in\mathbb{Z}} |c_p(f)|^2 = 0$ et $f = 0$: λ n'est pas valeur propre. On en déduit que

$$\mathrm{Sp}(\mathrm{K}) \subset \{2\pi c_p(k).\, p \in \mathbb{Z}\}.$$

Soit une valeur propre λ non nulle de K. On considère l'ensemble $\mathrm{A}_\lambda = \left\{p \in \mathbb{Z}.\, c_p(k) = \frac{\lambda}{2\pi}\right\}$. D'après ce qui précède, il n'est pas vide. Comme $c_p(k)$ tend vers 0 lorsque $|p|$ tend vers l'infini, il est nécessairement fini.

Soit f un vecteur propre relatif à la valeur propre λ. Si $p \notin \mathrm{A}_\lambda$, alors $c_p(f) = 0$. Soit l'élément de E défini par $g(x) = \sum_{p\in\mathrm{A}_\lambda} c_p(f)e^{ipx}$ et $h = f-g$. En distinguant les cas $m \in \mathrm{A}_\lambda$ et $m \notin \mathrm{A}_\lambda$, on montre que, pour tout $m \in \mathbb{Z}$, $c_m(h) = 0$. On en déduit que $\|h\|_2 = 0$ et donc $h = 0$, c'est-à-dire $f = g$. La fonction f est égal à la fonction $x \longmapsto \sum_{p\in\mathrm{A}_\lambda} c_p(f)e^{ipx}$.

Réciproquement, toute fonction f de la forme $x \longmapsto \sum_{p\in\mathrm{A}_\lambda} c_p e^{ipx}$, où les c_p sont des nombres complexes quelconques, vérifie que $\mathrm{K}f = \lambda f$.

On conclut donc que λ est valeur propre de K et que le sous-espace propre correspondant est engendré par les fonctions $x \longmapsto e^{ipx}$ $(p \in \mathrm{A}_\lambda)$. Il est donc de dimension finie.

5. Dans cette question, L est un opérateur compact quelconque de E.

a. Soit λ une valeur propre non nulle de L et E_λ l'espace propre correspondant. L'espace vectoriel E est muni du produit scalaire défini par $(f,g) = \int_0^{2\pi} f\overline{g}$, dont est issue la norme $\| \ \|_2$. Si E_λ n'est pas de dimension finie, on peut trouver dans E_λ une suite $(f_n)_{n\in\mathbb{N}}$ de vecteurs formant une famille orthonormale. Les vecteurs f_n sont unitaires, donc la suite $(f_n)_{n\in\mathbb{N}}$ est bornée. On peut extraire de $(\mathrm{L}f_n)_{n\in\mathbb{N}}$ une suite $(\mathrm{L}f_{\varphi(n)})_{n\in\mathbb{N}}$ convergente. Par définition, on a pour tout $n \in \mathbb{N}$, $f_{\varphi(n)} = \dfrac{1}{\lambda}\mathrm{L}f_{\varphi(n)}$. On en déduit que la suite $(f_{\varphi(n)})_{n\in\mathbb{N}}$ est également convergente. Soit f sa limite. Par continuité de la norme on obtient $\|f\|_2 = 1$. Mais d'autre part, on a pour tout $(n,p) \in \mathbb{N}^2$, $(f_{\varphi(n)}, f_{\varphi(p)}) = 0$. Par continuité du produit scalaire, on en déduit, en faisant tendre n et p vers $+\infty$, $(f,f) = 0$ et donc $\|f\|_2 = 0$. Nous aboutissons à la contradiction voulue. Tout espace propre distinct du noyau est donc de dimension finie.

Naturellement cette démonstration générale s'appliquait dans la question précédente, mais dans ce cas particulier, il était possible de déterminer les espaces propres.

Soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite de valeurs propres deux à deux distinctes. Montrons que 0 est valeur d'adhérence de la suite $(\lambda_n)_{n\in\mathbb{N}}$. Pour tout n, considérons un vecteur propre f_n associé à la valeur propre λ_n. Posons $\mathrm{X}_n = \mathrm{Vect}(f_0, f_1, \ldots, f_n)$. Les valeurs propres étant toutes distinctes, $(f_0, f_1, \ldots, f_n)$ est une base de X_n. Soit $(e_n)_{n\in\mathbb{N}}$ la suite obtenue en appliquant à $(f_n)_{n\in\mathbb{N}}$ le procédé d'orthonormalisation de Schmidt. Si $n \in \mathbb{N}^*$, alors il existe $\alpha_n \in \mathbb{R}$ et $g_n \in \mathrm{X}_{n-1}$ tels que $e_n = \alpha_n f_n + g_n$. On a alors $\mathrm{L}e_n = \alpha_n\lambda_n f_n + \mathrm{L}g_n$; par construction $\mathrm{L}g_n$ est dans X_{n-1} et donc $\mathrm{L}e_n - \lambda_n e_n$ est dans X_{n-1}. Soit maintenant n et m tels que $0 < m < n$. On a $\mathrm{L}e_n - \lambda_n e_n \in \mathrm{X}_{n-1}$, $\mathrm{L}e_m - \lambda_m e_m \in \mathrm{X}_{m-1} \subset \mathrm{X}_{n-1}$ et $e_m \in \mathrm{X}_{n-1}$. On en déduit $\mathrm{L}e_n - \lambda_n e_n - \mathrm{L}e_m \in \mathrm{X}_{n-1}$. Ce vecteur est donc orthogonal à e_n, ce qui donne $\lambda_n = (\mathrm{L}e_n - \mathrm{L}e_m, e_n)$ et donc $|\lambda_n| \leqslant \|\mathrm{L}e_n - \mathrm{L}e_m\|_2$.

La suite $(e_n)_{n\in\mathbb{N}}$ étant bornée, on peut extraire de $(\mathrm{L}e_n)_{n\in\mathbb{N}}$ une suite convergente $(\mathrm{L}e_{\varphi(n)})_{n\in\mathbb{N}}$. Pour $0 < m < n$, on a

$$|\lambda_{\varphi(n)}| \leqslant \|\mathrm{L}e_{\varphi(n)} - \mathrm{L}e_{\varphi(m)}\|_2.$$

La suite $(\mathrm{L}e_{\varphi(n)})_{n\in\mathbb{N}}$ étant de Cauchy, la suite $(\lambda_{\varphi(n)})_{n\in\mathbb{N}}$ converge vers 0, qui est donc valeur d'adhérence de la suite $(\lambda_n)_{n\in\mathbb{N}}$.

b. Supposons que $(\lambda_n)_{n\in\mathbb{N}}$ est une suite de valeurs propres deux à deux distinctes qui ne converge pas vers 0. Il existe donc $\varepsilon > 0$ et une suite extraite $(\lambda_{\varphi(n)})$ telle que $|\lambda_{\varphi(n)}| \geqslant \varepsilon$ pour tout n. Mais alors 0 ne peut pas être valeur d'adhérence de cette suite $(\lambda_{\varphi(n)})$ ce qui est absurde d'après la question précédente. ◁

Le lecteur trouvera encore quelques belles applications de la théorie des séries de Fourier (par exemple l'étude de l'équation de la chaleur) dans le troisième tome d'analyse.

Index

Abel (transformation d'), 97, 99, 104, 175, 176, 179, 180, 182, 299
Abel-Dirichlet (théorème d'), 178, 182, 188, 189, 203, 220
Alembert-Gauss (théorème de d'), 87, 252
Apéry (théorème d'), 313
Ascoli (théorème d'), 162

Banach (algèbre de), 351
Bernoulli (nombres de), 310, 312
Bernstein
  inégalité de, 277, 280
  polynômes de, 125, 135
  théorème de, 237, 340, 354
Bertrand (séries de), 62
Bieberbach (théorème de), 227
Bolzano (courbe de), 117
Bolzano-Weierstrass (théorème de), 162
Borel-Lebesgue (théorème de), 156, 161

Cantor (procédé diagonal de), 163, 165, 166, 209
Cauchy
  critère de, 120, 163, 180
  formules de, 241
  inégalités de, 239
  produit de, 189, 191, 192, 197, 200, 202, 204, 205, 217, 219, 312
  règle de, 174
  suite de, 177, 257, 350, 352, 363
  tranche de, 101, 102, 104, 300
Cauchy-Lipschitz (théorème de), 209
Cauchy-Schwarz (inégalité de), 12, 23–26, 28, 112, 223, 226, 249, 265, 287, 323, 325, 355, 360, 361
Chudnovsky (théorème de), 131
convergence dominée (théorème de), 53, 59, 60, 62, 68, 89, 188, 203, 333

découpe taubériene, 106, 221
Dini (théorème de), 156, 157, 160
Dirichlet
  séries de, 102
  théorème de, 287, 299, 311, 314

équirépartition, 45–50
eulérien (développement), 142, 152, 291

Fejér
  noyau de, 277, 279, 280, 342, 347
  théorème de, 347
fonction presque périodique, 133
Fourier
  transformée de, 307
Fubini (théorème de), 264, 304, 360

Gauss
  méthode de, 32
  théorème de, 215
Gibbs (phénomène de), 331
Gronwall (lemme de), 76

Hardy (inégalités de), 24
Helly (théorème de sélection de), 166
Hilbert (inégalité de), 284–287
Hölder (inégalité de), 11

indice d'une courbe fermée, 84
inégalité
  de Cauchy-Schwarz, 12, 23–26, 28, 112, 223, 226, 249, 265, 287, 323, 325, 355, 360, 361
  de Hilbert, 284–287
  de Hölder, 11
  de Van der Corput, 14
  de Wirtinger, 27, 315, 316
  de Young, 8

Jackson
  noyau de, 336, 342
  théorème de, 336
Jacobi (fonction theta de), 310

Korovkin (théorème de), 136

Laguerre (approximation de), 138
Lambert (séries de), 218
Laplace (méthode de), 66–74
Laurent (série de), 229
Leibniz (formule de), 237

moyenne
  deuxième formule de la, 43
  première formule de la, 11, 71
méthode des trapèzes, 17, 19

noyau
  de Fejér, 279, 280, 342, 347
  de Jackson, 342
  de Poisson, 342

Parseval (formule de), 284, 285, 287,
    315, 318, 321, 323, 326, 352,
    355, 356, 360
partition d'un entier, 197, 200
Poisson
  formule sommatoire de, 306–310
  noyau de, 342
principe des zéros isolés, 27,
    238–240
principe du maximum, 1

Riemann (sommes de), 34
    219
Riemann-Lebesgue (lemme d
Rolle (théorème de), 33, 281

Stirling (formule de), 40, 41, 73, 1,
    236, 244, 246
suites équiréparties, 45–50

taubériens (théorèmes), 221
Taylor
  formule de avec reste intégral, 25,
    232, 237
  polynôme de, 69, 72
  série de, 232, 237
Taylor-Lagrange
  formule de, 38, 39, 71, 177
  inégalité de, 138
théorème du relèvement, 84

Van der Corput (inégalité de), 14

Wallis (intégrales de), 56, 74
Walsh (théorème de), 130
Weierstrass (théorème de), 31, 32,
    42, 49, 86, 125, 132, 134,
    137, 139, 258, 334
Weyl (critère de), 47
Wirtinger (inégalité de), 27, 315, 316

Young (inégalité de), 8

Imprimerie BARNÉOUD - BONCHAMP-LÈS-LAVAL
Dépôt légal : mars 2009 - N° d'imprimeur : 902115
Imprimé en France

Le recueil d'exercices résolus des oraux des Écoles normales supérieures et de l'École polytechnique de Serge Francinou, Hervé Gianella et Serge Nicolas comprendra six volumes : trois consacrés à l'algèbre et trois à l'analyse.

Le présent volume aborde le cœur du programme d'analyse des concours : intégration, suites et séries de fonctions, séries entières, séries de Fourier.

Les auteurs se sont attachés à dégager les idées qui se trouvent à la source des solutions fournies, sans pour autant omettre le détail des vérifications et des calculs. Comme dans les volumes précédents, à côté d'exercices techniques, le lecteur trouvera de nombreux énoncés destinés à établir un résultat mathématique significatif. Les auteurs les ont identifiés, et les resituent dans leur contexte naturel. Un soin tout particulier a été apporté au texte de présentation qui accompagne les exercices, groupés par thèmes. La présentation historique qui ponctue la succession des énoncés montrera au lecteur que l'élaboration des concepts de l'Analyse – qui apparaît aujourd'hui comme un édifice achevé – n'a pas été sans erreurs, hésitations, retours en arrière. D'autre part, certains points du programme parfois négligés par les candidats font l'objet d'utiles rappels.

Ce livre s'adresse naturellement aux élèves des classes préparatoires, mais il sera également très utile aux candidats à l'agrégation qui y trouveront de nombreux développements pour leur oral. Ces exercices constituent aussi un excellent complément à la préparation à l'écrit du CAPES.

Collection enseignement des mathématiques

ISBN 978-2-84225-141-3 Graphisme : Massin