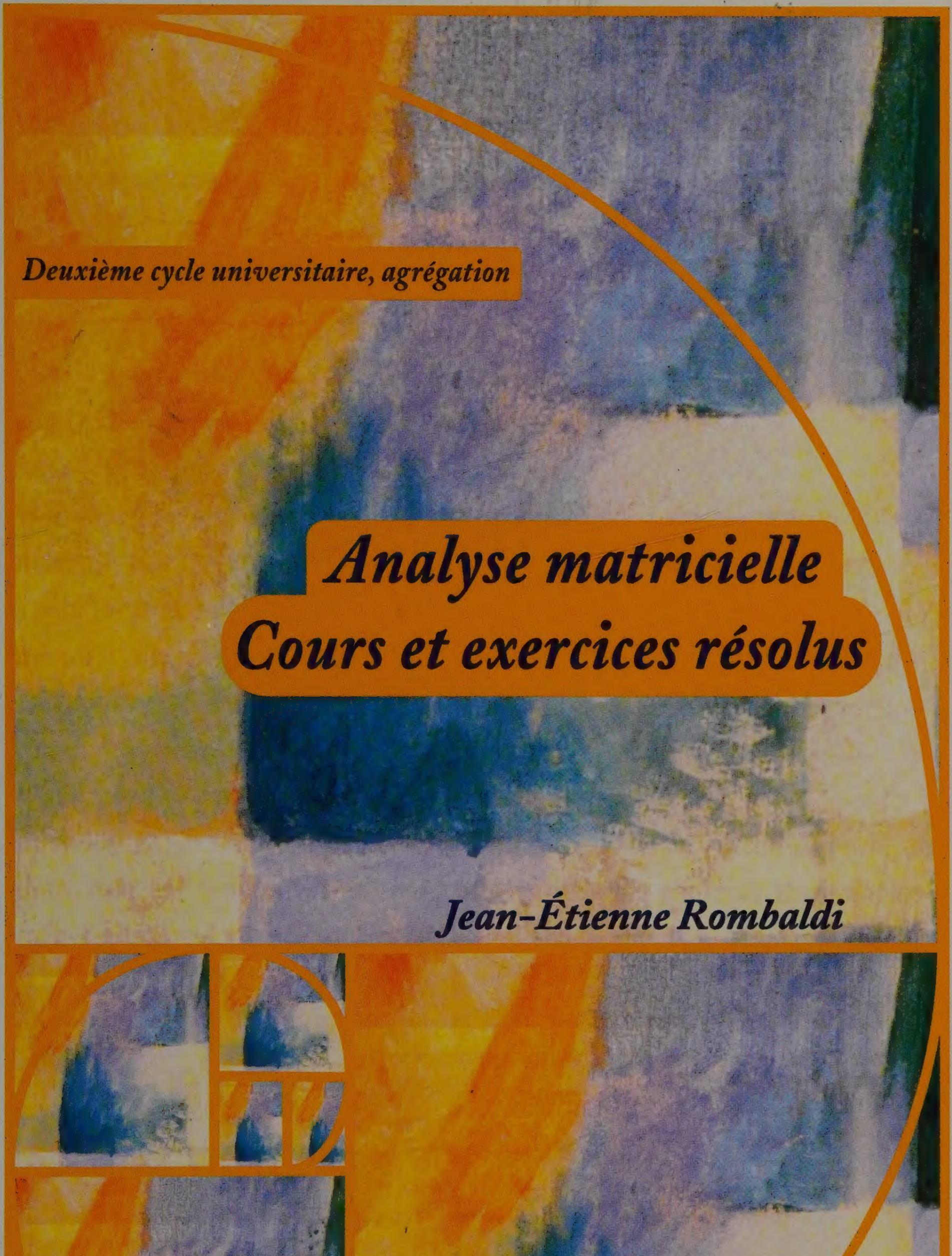

Deuxième cycle universitaire, agrégation

Analyse matricielle
Cours et exercices résolus

Jean-Étienne Rombaldi

EDP SCIENCES

Analyse matricielle
Cours et exercices résolus

(Deuxième cycle universitaire, agrégation)

Jean-Étienne Rombaldi

Professeur agrégé de mathématiques à l'université d'Aix-Marseille III

EDP SCIENCES

7, avenue du Hoggar
Parc d'activités de Courtabœuf, BP 112
91944 Les Ulis Cedex A, France

Couverture : Éric Bonnet

Composition sous LaTeX : Hypercom

ISBN : 2-86883-425-6

Avant-propos

Cet ouvrage, qui pourrait s'intituler « Matrices réelles et complexes, propriétés algébriques et topologiques, applications », est consacré à l'étude de l'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ des matrices carrées d'ordre n à coefficients réels ou complexes du point de vue algébrique et topologique. Cette étude est un préalable important à tout bon cours d'analyse numérique.

Des connaissances de base en algèbre linéaire sont amplement suffisantes pour la lecture de cet ouvrage.

Le public visé est celui des étudiants du deuxième cycle universitaire et des candidats à l'agrégation externe et interne de mathématiques.

La synthèse proposée est, je pense, un bon moyen de réviser ses connaissances sur les espaces vectoriels normés et l'algèbre linéaire. Les candidats à l'agrégation trouveront tout au long de cet ouvrage de nombreux exemples d'applications des résultats classiques souvent proposés dans les leçons d'oral. Par exemple, si dans une leçon sur le groupe orthogonal on pense à mentionner la compacité de $\mathcal{O}_n(\mathbb{R})$ il faut avoir réfléchi à quelques exemples d'applications de ce résultat. En suivant cette idée je me suis efforcé de faire suivre chaque résultat classique et important d'un certain nombre d'applications.

Chaque chapitre est suivi d'une liste d'exercices corrigés. Une bonne utilisation de ces exercices consiste bien évidemment à les chercher au préalable puis à confronter les résultats obtenus aux solutions proposées.

Le premier chapitre est consacré à l'étude des espaces vectoriels normés et particulièrement au cas de la dimension finie. L'étude des propriétés topologiques de l'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ et l'application aux méthodes itératives de résolution des systèmes linéaires et de recherche des valeurs et vecteurs propres utilisent quelques résultats de ce chapitre. Le théorème du point fixe de Banach est utilisé dans l'étude des systèmes différentiels linéaires.

Les chapitres 2 et 3 sont consacrés à l'étude des valeurs et vecteurs propres des matrices réelles ou complexes. Les résultats importants sont le théorème de décomposition des noyaux et les divers théorèmes de réduction à la forme triangulaire ou diagonale.

C'est au chapitre 4 qu'on aborde l'étude des propriétés topologiques de l'espace vectoriel $\mathcal{M}_n(\mathbb{K})$. On y introduit les notions de norme matricielle induite par une norme vectorielle et on démontre quelques résultats classiques de densité et de connexité.

Pour ce qui est des applications de ce chapitre, je me suis limité à l'analyse numérique linéaire. Pour une application aux groupes de Lie, le lecteur intéressé pourra consulter l'ouvrage de Mnéimné et Testard [11].

Les chapitres 5 et 6 sont deux chapitres importants de l'analyse numérique linéaire. On s'intéresse aux méthodes directes et itératives de résolution des systèmes linéaires et aux méthodes de calcul approché des valeurs et vecteurs propres d'une matrice carrée réelle ou complexe.

Enfin le chapitre 8 est une application à l'étude des systèmes différentiels linéaires à coefficients constants ou non et à l'exponentielle d'une matrice. L'exponentielle d'une matrice y est définie à partir de l'étude des systèmes différentiels linéaires à coefficients constants.

Un grand merci à René Adad qui a lu et critiqué une première version de ce travail et à Dominique Barbolosi pour ses remarques toujours judicieuses.

Mes remerciements vont aussi à toute l'équipe d'EDP Sciences pour la qualité de leur travail.

Jean-Étienne ROMBALDI

Notations

$\mathbb{R}$	corps des nombres réels.
$\mathbb{C}$	corps des nombres complexes.
$\Re(z)$	partie réelle du nombre complexe z.
$\Im(z)$	partie imaginaire du nombre complexe z.
$\mathbb{K}$	corps des nombres réels ou complexes.
$\mathbb{K}[X]$	algèbre des polynômes à une indéterminée à coefficients dans $\mathbb{K}$.
$C_n^p = \dfrac{n!}{p!\,(n-p)!}$	coefficient binomial.
$\mathbf{Vect}\,(X)$	sous-espace vectoriel engendré par une partie X d'un espace vectoriel.
E^*	dual algébrique de l'espace vectoriel E.
$\Vert\cdot\Vert$	norme sur un $\mathbb{K}$-espace vectoriel.
$\langle\cdot \mid \cdot\rangle$	produit scalaire sur un espace euclidien ou hermitien.
$(E, \Vert\cdot\Vert)$	espace vectoriel E sur $\mathbb{K}$ muni de la norme $\Vert\cdot\Vert$.
$\mathcal{B}(X, E)$	espace vectoriel des fonctions définies sur l'ensemble X, à valeurs dans l'espace vectoriel normé E et bornées.
$\mathcal{C}^0([a,b], E)$	espace vectoriel des applications continues définies sur l'intervalle $[a,b]$ $(a < b)$ et à valeurs dans l'espace vectoriel E.
$\mathrm{d}\varphi(x)$	vecteur gradient de la fonction différentiable φ en x.
$(x_n)_{n\in\mathbb{N}}$	suite à valeurs dans un espace vectoriel E.
$\mathcal{L}(E)$	espace vectoriel des endomorphismes de l'espace vectoriel E.
$\mathcal{O}(E)$	groupe des endomorphismes orthogonaux de l'espace vectoriel E.
Id	application linéaire identité.
$\ker(u)$	noyau de l'application linéaire u.
$\mathrm{Im}\,(u)$	image de l'application linéaire u.
$\det(u)$	déterminant de l'endomorphisme u.
π_u	polynôme minimal de l'endomorphisme u.
$\mathrm{sp}\,(u)$	ensemble des valeurs propres de l'endomorphisme u.
$\mathrm{Tr}\,(u)$	trace de l'endomorphisme u.
$\mathcal{M}_n(\mathbb{K})$	espace vectoriel des matrices à n lignes et n colonnes à coefficients dans $\mathbb{K}$.

$\mathcal{D}_n(\mathbb{K})$	ensemble des matrices diagonalisables à n lignes et n colonnes à coefficients dans $\mathbb{K}$.
$GL_n(\mathbb{K})$	groupe des matrices inversibles à n lignes et n colonnes à coefficients dans $\mathbb{K}$.
$\mathcal{O}_n(\mathbb{R})$	groupe des matrices orthogonales.
$\mathcal{O}_n^+(\mathbb{R})$	groupe des matrices de rotations.
$\mathcal{U}_n(\mathbb{C})$	groupe des matrices unitaires.
I_n	matrice unité d'ordre n.
${}^tA$	transposée de la matrice A.
A^*	adjointe de la matrice A.
$\det(A)$	déterminant de la matrice A.
π_A	polynôme minimal de la matrice A.
$\mathrm{sp}(A)$	ensemble des valeurs propres de la matrice A.
$\rho(A)$	rayon spectral de la matrice A.
$\mathrm{Tr}(A)$	trace de la matrice A.
R_A	quotient de Rayleigh-Ritz associé à la matrice A.
e^A	exponentielle de la matrice A.

Sommaire

1 Espaces vectoriels normés

Dans tout ce chapitre $\mathbb{K}$ désigne le corps des réels ou des complexes.

1. Normes sur un espace vectoriel réel ou complexe

Dans ce paragraphe E désigne un espace vectoriel sur $\mathbb{K}$ de dimension quelconque (finie ou non).

Définition 1.1 : *Une norme sur E est une application :*

$$\begin{array}{rccc} \|\cdot\| : & E & \to & \mathbb{R} \\ & x & \longmapsto & \|x\| \end{array}$$

qui vérifie les propriétés suivantes :

$$\begin{array}{lll} \forall x \in E, & (\|x\| = 0 \Rightarrow x = 0), & \\ \forall \lambda \in \mathbb{K}, & \forall x \in E, & \|\lambda x\| = |\lambda| \, \|x\|, \\ \forall x \in E, & \forall y \in E, & \|x + y\| \leq \|x\| + \|y\|. \end{array}$$

On notera $(E, \|\cdot\|)$ l'espace vectoriel E muni de la norme $\|\cdot\|$.

Remarque 1.1 *: Une norme sur E est nécessairement à valeurs dans $\mathbb{R}^+$. En effet pour tout x dans E on a :*

$$0 = \|x - x\| \leq 2 \|x\| .$$

Proposition 1.1 : *Pour toute norme sur E on a :*

$$\forall x \in E, \quad \forall y \in E, \quad |\|x\| - \|y\|| \leq \|x - y\| .$$

Démonstration : On a pour tous x, y dans E :

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\| ,$$
$$\|y\| = \|y - x + x\| \leq \|x - y\| + \|x\| ,$$

et en conséquence :

$$- \|x - y\| \leq \|x\| - \|y\| \leq \|x - y\| .$$ ■

Exemple 1.1 : *Soient E un espace vectoriel sur $\mathbb{K}$ de dimension finie $n \geq 1$ et $\mathcal{B} = \{e_1, \ldots, e_n\}$ une base de E. On peut définir sur E les trois normes classiques suivantes (exercice 1.1) :*

$$\forall x = \sum_{i=1}^{n} x_i e_i \in E, \begin{cases} \|x\|_\infty = \max\limits_{1 \leq i \leq n} |x_i|, \\ \|x\|_1 = \sum\limits_{i=1}^{n} |x_i|, \\ \|x\|_2 = \sqrt{\sum\limits_{i=1}^{n} |x_i|^2}. \end{cases}$$

Exemple 1.2 : *Soient E un espace vectoriel sur $\mathbb{K}$ de dimension finie $n \geq 1$ et $\mathcal{B} = \{e_1, \ldots, e_n\}$ une base de E. Pour tout réel $p \geq 1$ l'application :*

$$x = \sum_{i=1}^{n} x_i e_i \mapsto \|x\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p}$$

définit une norme sur E (exercice 1.2).

Remarque 1.2 : *On peut montrer (exercice 1.3) que pour tout x dans E on a, avec les notations des exemples 1.1 et 1.2, $\lim\limits_{p \to +\infty} \|x\|_p = \|x\|_\infty$.*

Exemple 1.3 : *Sur l'espace vectoriel $\mathcal{C}^0([a,b], \mathbb{K})$ des applications continues sur l'intervalle $[a,b]$ $(a < b)$ et à valeurs dans $\mathbb{K}$ les application suivantes :*

$$f \mapsto \|f\|_\infty = \sup_{[a,b]} |f|,$$

$$f \mapsto \|f\|_1 = \int_a^b |f(t)| \, \mathrm{d}t,$$

$$f \mapsto \|f\|_2 = \sqrt{\int_a^b |f(t)|^2 \, \mathrm{d}t}$$

définissent des normes (exercice 1.4).

Exemple 1.4 : *Sur l'espace vectoriel $\mathcal{C}^0([a,b], \mathbb{K})$ des applications continues sur l'intervalle $[a,b]$ $(a < b)$ et à valeurs dans $\mathbb{K}$, pour tout réel $p \geq 1$, l'application :*

$$f \mapsto \|f\|_p = \sqrt[p]{\int_a^b |f(x)|^p \, \mathrm{d}x}$$

définit une norme (exercice 1.5).

Remarque 1.3 : *On peut montrer (exercice 1.6) que pour toute fonction f dans $\mathcal{C}^0([a,b], \mathbb{K})$ on a $\lim\limits_{p \to +\infty} \|f\|_p = \|f\|_\infty$.*

Les normes $\|\cdot\|_2$ des exemples 1.1 et 1.3 sont des cas particuliers de normes déduites d'un produit scalaire (voir le paragraphe 3, page 82).

2. Topologie associée à une norme

Pour ce paragraphe, $(E, \|\cdot\|)$ est un espace vectoriel normé.

Une norme sur un espace vectoriel normé permet de définir une topologie c'est-à-dire une famille d'ouverts.

Définition 1.2 : *On dit qu'une partie $\mathcal{O}$ de $(E, \|\cdot\|)$ est un ouvert si cette partie est vide, ou si elle est non vide et si pour tout a dans $\mathcal{O}$ il existe un réel $r > 0$ tel que :*

$$\mathcal{B}(a, r) = \{x \in \mathcal{O} \mid \|x - a\| < r\} \subset \mathcal{O}.$$

On dit aussi que tout point de $\mathcal{O}$ est intérieur à $\mathcal{O}$.

Définition 1.3 : *On dit qu'une partie $\mathcal{F}$ de $(E, \|\cdot\|)$ est un fermé si son complémentaire dans E est un ouvert de E.*

Définition 1.4 : *On dit qu'une partie $\mathcal{A}$ de $(E, \|\cdot\|)$ est connexe si toute inclusion $\mathcal{A} \subset \mathcal{O}_1 \cup \mathcal{O}_2$ avec $\mathcal{O}_1, \mathcal{O}_2$ ouverts de E tels que $\mathcal{A} \cap \mathcal{O}_1 \cap \mathcal{O}_2 = \emptyset$ entraîne $\mathcal{A} \cap \mathcal{O}_1 = \emptyset$ (et $\mathcal{A} \subset \mathcal{O}_2$) ou $\mathcal{A} \cap \mathcal{O}_2 = \emptyset$ (et $\mathcal{A} \subset \mathcal{O}_1$).*

Il est facile de vérifier qu'une partie $\mathcal{A}$ de $(E, \|\cdot\|)$ est connexe si toute inclusion $\mathcal{A} \subset \mathcal{F}_1 \cup \mathcal{F}_2$ avec $\mathcal{F}_1, \mathcal{F}_2$ fermés de E tels que $\mathcal{A} \cap \mathcal{F}_1 \cap \mathcal{F}_2 = \emptyset$ entraîne $\mathcal{A} \cap \mathcal{F}_1 = \emptyset$ (et $\mathcal{A} \subset \mathcal{F}_2$) ou $\mathcal{A} \cap \mathcal{F}_2 = \emptyset$ (et $\mathcal{A} \subset \mathcal{F}_1$).

Définition 1.5 : *On dit qu'une partie non vide $\mathcal{A}$ de $(E, \|\cdot\|)$ est connexe par arcs si pour tous points x, y de $\mathcal{A}$ on peut trouver une application continue γ définie sur $[0, 1]$ et à valeurs dans $\mathcal{A}$ telle que $\gamma(0) = x$ et $\gamma(1) = y$.*

On peut vérifier qu'un ensemble connexe par arcs est en particulier connexe.

Définition 1.6 : *On dit qu'une partie non vide $\mathcal{A}$ de $(E, \|\cdot\|)$ est convexe si pour tous points x, y de $\mathcal{A}$ le segment :*

$$[x, y] = \{(1 - t)x + ty \mid 0 \leq t \leq 1\}$$

est contenu dans $\mathcal{A}$.

Une partie convexe de E est en particulier connexe par arcs.

On peut également définir les notions de convergence pour les suites et les séries ainsi que la notion de continuité pour les fonctions définies sur E et à valeurs dans $\mathbb{K}$ ou dans un autre espace vectoriel normé.

Dans le cas d'un espace vectoriel normé de dimension finie, ces notions ne dépendent pas du choix d'une norme comme on va le voir dans le paragraphe 5, page 25.

Définition 1.7 : *On dit qu'une suite $(x_n)_{n \in \mathbb{N}}$ d'éléments de $(E, \|\cdot\|)$ est convergente s'il existe x dans E tel que $\lim\limits_{n \to +\infty} \|x_n - x\| = 0$. Une suite non convergente est dite divergente.*

Si une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de $(E, \|\cdot\|)$ est convergente alors sa limite est unique et on peut noter $x = \lim_{n\to+\infty} x_n$.

Cette notion dépend *a priori* de la norme choisie sur E. Considérons par exemple la suite de fonctions $(f_n)_{n\in\mathbb{N}}$ définie sur $[0,1]$ par :

$$f_n(x) = \begin{cases} -n^3 x + n & \text{si } x \in \left[0, \dfrac{1}{n^2}\right], \\ 0 & \text{si } x \in \left[\dfrac{1}{n^2}, 1\right]. \end{cases}$$

On a $\lim_{n\to+\infty} \|f_n\|_1 = 0$ et $\lim_{n\to+\infty} \|f_n\|_\infty = +\infty$, c'est-à-dire que cette suite converge vers la fonction nulle pour la norme $\|\cdot\|_1$ et diverge pour la norme $\|\cdot\|_\infty$.

Définition 1.8 : *On dit qu'une partie non vide $\mathcal{B}$ de $(E, \|\cdot\|)$ est bornée s'il existe une constante $\lambda > 0$ telle que $\|x\| \leq \lambda$ pour tout x dans $\mathcal{B}$.*

Il est facile de vérifier qu'une suite convergente est bornée.

La démonstration du théorème suivant est également élémentaire.

Théorème 1.1 : *Soit $\mathcal{F}$ un sous-ensemble non vide de $(E, \|\cdot\|)$. Cet ensemble est fermé si et seulement si pour toute suite d'éléments de $\mathcal{F}$ qui converge vers x dans E on a $x \in \mathcal{F}$.*

Définition 1.9 : *On dit qu'une suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de $(E, \|\cdot\|)$ est une suite de Cauchy si pour tout réel $\epsilon > 0$ on peut trouver un entier n_ϵ tel que pour tous entiers p, q on a :*

$$p \geq n_\epsilon, \quad q \geq n_\epsilon, \quad \Rightarrow \quad \|x_p - x_q\| < \epsilon.$$

En utilisant l'inégalité triangulaire on déduit immédiatement que toute suite convergente dans un espace vectoriel normé est de Cauchy. Mais en général la réciproque est fausse (exercice 1.11).

Définition 1.10 : *On dit qu'un espace vectoriel normé $(E, \|\cdot\|)$ est complet, ou que c'est un espace de Banach, si toute suite de Cauchy dans E est convergente.*

On rappelle que $(\mathbb{K}, |\cdot|)$ est complet. De ce résultat on déduira que tout espace vectoriel normé de dimension finie est complet.

Un exemple important d'espace de Banach est donné par le résultat suivant.

Théorème 1.2 : *Soient X un ensemble non vide, $(E, \|\cdot\|)$ un espace vectoriel normé et $\mathcal{B}(X, E)$ l'espace vectoriel des fonctions bornées de X dans E normé par $\|f\|_\infty = \sup_{x\in X} \|f(x)\|$. L'espace vectoriel normé $\mathcal{B}(X, E)$ est complet si et seulement si E est complet.*

Démonstration : Supposons E complet. Soit $(f_n)_{n\in\mathbb{N}}$ une suite de Cauchy dans l'espace vectoriel normé $\left(\mathcal{B}(X,E), \|\cdot\|_\infty\right)$. Pour tout $\epsilon > 0$, il existe un entier n_ϵ tel que :

$$\forall n \geq n_\epsilon, \forall m \geq n_\epsilon, \forall x \in X, \quad \|f_n(x) - f_m(x)\| < \epsilon. \tag{1.1}$$

Pour x fixé dans X, la suite $\left(f_n(x)\right)_{n\in\mathbb{N}}$ est de Cauchy dans l'espace complet E et converge donc vers un élément $f(x)$ dans E. En faisant tendre m vers l'infini dans (1.1), on déduit que :

$$\forall x \in X, \forall n \geq n_\epsilon, \quad \|f_n(x) - f(x)\| \leq \epsilon.$$

C'est-à-dire que la suite $(f_n)_{n\in\mathbb{N}}$ converge uniformément vers f sur X.

En écrivant que :

$$\|f(x)\| \leq \|f(x) - f_{n_\epsilon}(x)\| + \|f_{n_\epsilon}(x)\| \leq \epsilon + \|f_{n_\epsilon}\|_\infty$$

pour tout x dans X, on déduit que f est dans $\mathcal{B}(X,E)$. Ce qui achève de prouver que $\left(\mathcal{B}(X,E), \|\cdot\|_\infty\right)$ est complet.

Réciproquement supposons que $\left(\mathcal{B}(X,E), \|\cdot\|_\infty\right)$ soit complet. Si $(y_n)_{n\in\mathbb{N}}$ est une suite de Cauchy dans E alors la suite de fonctions constantes $(f_n)_{n\in\mathbb{N}}$ définies par $f_n(x) = y_n$ est de Cauchy dans $\mathcal{B}(X,E)$ et elle converge uniformément vers une fonction f nécessairement constante. La valeur prise par f est alors limite de la suite $(y_n)_{n\in\mathbb{N}}$ dans E. C'est-à-dire que E est complet. ∎

Définition 1.11 : *Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $(E, \|\cdot\|)$. On appelle sous-suite ou suite extraite de $(x_n)_{n\in\mathbb{N}}$ toute suite $\left(x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ où φ est une application strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$.*

Si la suite $\left(x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ converge on dit alors que sa limite est une valeur d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$.

Définition 1.12 : *On dit qu'une partie non vide $\mathcal{C}$ de $(E, \|\cdot\|)$ est compacte si de toute suite de points de $\mathcal{C}$ on peut extraire une sous-suite qui converge vers un élément de $\mathcal{C}$ (propriété de Bolzano-Weierstrass).*

Théorème 1.3 : *Toute partie compacte d'un espace vectoriel normé $(E, \|\cdot\|)$ est fermée et bornée.*

Démonstration : Soient $\mathcal{C}$ une partie non vide de E et $(x_n)_{n\in\mathbb{N}}$ une suite d'éléments de $\mathcal{C}$ qui converge vers $x \in E$. Si $\mathcal{C}$ est compacte on peut alors extraire de $(x_n)_{n\in\mathbb{N}}$ une suite qui converge vers un élément de $\mathcal{C}$ et, comme toute suite extraite de $(x_n)_{n\in\mathbb{N}}$ converge nécessairement vers x, on déduit que x est dans $\mathcal{C}$. En conséquence $\mathcal{C}$ est fermée.

Si $\mathcal{C}$ est non bornée on peut alors trouver pour tout entier n positif un élément x_n dans $\mathcal{C}$ tel que $\|x_n\| \geq n$. Mais alors pour toute fonction φ strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$ on a :

$$\forall n \in \mathbb{N}, \quad \left\|x_{\varphi(n)}\right\| \geq \varphi(n) \geq n$$

et $\lim\limits_{n\to+\infty} \left\|x_{\varphi(n)}\right\| = +\infty$. En conséquence $\mathcal{C}$ n'est pas compacte. ∎

Remarque 1.4 : *On verra qu'en dimension finie la réciproque du théorème 1.3 est vraie. Mais en dimension infinie une partie fermée bornée n'est pas nécessairement compacte (exercice 1.12).*

On rappelle un résultat important sur $\mathbb{R}$.

Théorème 1.4 (Bolzano-Weierstrass) : *De toute suite bornée de nombres réels on peut extraire une sous-suite convergente.*

La démonstration peut se faire en utilisant le principe de dichotomie. Si $[a_0, b_0]$ est un intervalle réel qui contient tous les éléments de la suite $(x_n)_{n\in\mathbb{N}}$, on le coupe en deux parties égales et on garde une de ces parties qui contient des x_n pour une infinité d'indices n. En réitérant ce procédé on construit deux suites adjacentes $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ et une application φ strictement croissante de $\mathbb{N}$ dans $\mathbb{N}$ telles que chaque intervalle $[a_n, b_n]$ contient un terme $x_{\varphi(n)}$. La suite $\left(x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ est alors convergente.

De ce théorème on déduit immédiatement le corollaire suivant.

Corollaire 1.1 : *Les compacts de $\mathbb{K}$ sont les fermés bornés.*

Ce corollaire s'exprime en disant que $\mathbb{K}$ est localement compact.

Définition 1.13 : *Soient C une partie non vide d'un espace vectoriel normé $(E, \|\cdot\|)$ et f une fonction définie sur C à valeurs dans un espace vectoriel normé $\left(F, \|\cdot\|'\right)$. On dit que f est continue en $x_0 \in C$ si :*

$$\forall \epsilon > 0, \quad \exists \eta > 0 \mid \forall x \in C, \quad \|x - x_0\| \leq \eta \Rightarrow \|f(x) - f(x_0)\|' \leq \epsilon.$$

On dit que f est continue sur C si elle est continue en tout point de C.

Cette notion de continuité dépend du choix des normes sur E et F (exercice 1.13).

Une définition équivalente est donnée par la proposition suivante.

Proposition 1.2 : *Soient C une partie non vide d'un espace vectoriel normé $(E, \|\cdot\|)$ et f une fonction définie sur C à valeurs dans un espace vectoriel normé $\left(F, \|\cdot\|'\right)$. La fonction f est continue sur C si et seulement si pour toute suite $(x_n)_{n\in\mathbb{N}}$ d'éléments de C qui converge vers $x \in C$ on a $\lim\limits_{n\to+\infty} f(x_n) = f(x)$ dans $\left(F, \|\cdot\|'\right)$.*

Dans le cas particulier des fonctions définies sur un compact et à valeurs réelles on a le résultat important suivant.

Théorème 1.5 : *Soient C un compact d'un espace vectoriel normé $(E, \|\cdot\|)$ et f une fonction définie sur C à valeurs dans $\mathbb{R}$. Si f est continue alors f est bornée et atteint ses bornes. C'est-à-dire qu'il existe x_1 et x_2 dans C tels que :*

$$f(x_1) = \inf_{x\in C} f(x), \quad f(x_2) = \sup_{x\in C} f(x).$$

Démonstration : Soit $(y_n)_{n\in\mathbb{N}}$ une suite dans $f(\mathcal{C})$ avec $y_n = f(x_n)$ pour tout $n \in \mathbb{N}$. De la suite $(x_n)_{n\in\mathbb{N}}$ dans le compact $\mathcal{C}$ on peut extraire une suite $\left(x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ qui converge vers un élément x de $\mathcal{C}$. Avec la continuité de f on a alors :

$$\lim_{n\to+\infty} y_{\varphi(n)} = \lim_{n\to+\infty} f\left(x_{\varphi(n)}\right) = f(x).$$

En conséquence $f(\mathcal{C})$ est compacte. En particulier $f(\mathcal{C})$ est une partie non vide bornée de $\mathbb{R}$ et donc admet une borne inférieure et une borne supérieure. Notons :

$$m = \inf_{x\in\mathcal{C}} f(x), \quad M = \sup_{x\in\mathcal{C}} f(x).$$

Il reste à montrer que m et M sont dans $f(\mathcal{C})$.

Par définition de la borne inférieure m, pour tout entier $n > 0$ on peut trouver x_n dans $\mathcal{C}$ tel que :

$$m \leq f(x_n) < m + \frac{1}{n}\cdot$$

De la suite $(x_n)_{n\in\mathbb{N}}$ ainsi définie dans le compact $\mathcal{C}$ on peut extraire une sous-suite $\left(x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ qui converge vers un élément x_1 de $\mathcal{C}$. On a donc pour tout entier $n > 0$:

$$m \leq f\left(x_{\varphi(n)}\right) < m + \frac{1}{\varphi(n)},$$

avec $\lim\limits_{n\to+\infty} \varphi(n) = +\infty$. On a donc, avec la continuité de f :

$$f(x_1) = \lim_{n\to+\infty} f\left(x_{\varphi(n)}\right) = m.$$

On procède de manière analogue pour la borne supérieure. ∎

Pour tout intervalle réel $[a,b]$ on désigne par $\mathcal{C}^0([a,b],E)$ l'espace vectoriel des fonctions continues définies sur $[a,b]$ et à valeurs dans E. On munit cet espace vectoriel de la norme de la convergence uniforme définie par :

$$\forall f \in \mathcal{C}^0([a,b],E), \quad \|f\|_\infty = \sup_{t\in[a,b]} \|f(t)\|.$$

Théorème 1.6 : *Si E est un espace de Banach, alors pour tout intervalle réel $[a,b]$ l'espace vectoriel normé $\left(\mathcal{C}^0([a,b],E), \|\cdot\|_\infty\right)$ est complet.*

Démonstration : Pour montrer que $\left(\mathcal{C}^0([a,b],E), \|\cdot\|_\infty\right)$ est complet il suffit de montrer que c'est un fermé de l'espace vectoriel normé complet $\left(\mathcal{B}([a,b],E), \|\cdot\|_\infty\right)$ (théorème 1.2).

Soit donc $(f_n)_{n\in\mathbb{N}}$ une suite de fonctions continues de $[a,b]$ dans E qui converge uniformément vers $f \in \mathcal{B}([a,b],E)$. Pour tout $\epsilon > 0$ on peut trouver un entier n_ϵ tel que :

$$\forall n \geq n_\epsilon, \quad \|f_n - f\|_\infty < \epsilon.$$

Pour x, x_0 dans $[a,b]$ on a alors :

$$\begin{aligned}\|f(x)-f(x_0)\| &\leq \|f(x)-f_{n_\epsilon}(x)\| + \|f_{n_\epsilon}(x)-f_{n_\epsilon}(x_0)\| \\ &\quad + \|f_{n_\epsilon}(x_0)-f(x_0)\| \\ &\leq 2\|f-f_{n_\epsilon}\|_\infty + \|f_{n_\epsilon}(x)-f_{n_\epsilon}(x_0)\| \\ &\leq 2\epsilon + \|f_{n_\epsilon}(x)-f_{n_\epsilon}(x_0)\|.\end{aligned}$$

La fonction f_{n_ϵ} étant continue on peut trouver un réel $\eta > 0$ tel que :

$$x \in [a,b] \text{ et } |x-x_0| < \eta \Rightarrow \|f_{n_\epsilon}(x)-f_{n_\epsilon}(x_0)\| < \epsilon.$$

On déduit alors que :

$$x \in [a,b] \text{ et } |x-x_0| < \eta \Rightarrow \|f(x)-f(x_0)\| < 3\epsilon.$$

C'est-à-dire que la fonction f est continue.

On a donc ainsi prouvé que $\mathcal{C}^0([a,b],E)$ est un fermé de l'espace complet $(\mathcal{B}([a,b],E), \|\cdot\|_\infty)$, et qu'il est donc lui aussi complet. ■

Définition 1.14 : *Soient $(E, \|\cdot\|)$ et $(F, \|\cdot\|')$ deux espaces vectoriels normés. On dit qu'une application f définie sur E et à valeurs dans F est un homéomorphisme si elle est continue bijective d'inverse f^{-1} bijective.*

3. Le théorème du point fixe de Banach

Définition 1.15 : *Soit $(E, \|\cdot\|)$ un espace vectoriel normé et F une partie non vide de E. On dit qu'une application $\varphi : F \longrightarrow E$ est strictement contractante s'il existe une constante $\lambda \in [0,1[$ telle que :*

$$\forall (x,y) \in F \times F, \quad \|\varphi(x)-\varphi(y)\| \leq \lambda \|x-y\|.$$

On dit aussi que φ est λ-contractante.

Définition 1.16 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé, F une partie non vide de E et φ une application de F dans F. On dit qu'un élément x de F est un point fixe de φ si $\varphi(x) = x$.*

Théorème 1.7 : *Si $(E, \|\cdot\|)$ est un espace vectoriel normé complet, F une partie fermée de E et $\varphi : F \to F$ une application λ-contractante avec $\lambda \in [0,1[$, alors φ admet un unique point fixe $y \in F$ et ce point fixe est la limite de toute suite $(x_n)_{n\in\mathbb{N}}$ d'approximations successives définie par :*

$$\begin{cases} x_0 \in F, \\ \forall n \in \mathbb{N}, \quad x_{n+1} = \varphi(x_n). \end{cases}$$

Démonstration : Pour l'existence du point fixe, on montre que la suite $(x_n)_{n\in\mathbb{N}}$ est de Cauchy dans $(E, \|\cdot\|)$.

Pour $q > p \geq 1$, on a :

$$\|x_q - x_p\| \leq \|x_q - x_{q-1}\| + \cdots + \|x_{p+1} - x_p\|,$$

avec pour tout $k \geq 1$:

$$\begin{aligned}\|x_{k+1} - x_k\| &= \|\varphi(x_k) - \varphi(x_{k-1})\| \\ &\leq \lambda \|x_k - x_{k-1}\| \leq \cdots \leq \lambda^k \|x_1 - x_0\|.\end{aligned}$$

Ce qui donne :

$$\begin{aligned}\|x_q - x_p\| &\leq \left(\sum_{k=p}^{q-1} \lambda^k\right) \|x_1 - x_0\| = \frac{\lambda^p \left(1 - \lambda^{q-p}\right)}{1 - \lambda} \|x_1 - x_0\| \\ &\leq \frac{\lambda^p}{1-\lambda} \|x_1 - x_0\| \underset{p\to+\infty}{\to} 0\end{aligned}$$

si $\lambda \in [0, 1[$.

On a donc ainsi montré que la suite $(x_n)_{n\in\mathbb{N}}$ est de Cauchy dans E et donc qu'elle converge vers un élément $x \in F$ puisque E est complet et F fermée dans E.

Avec la continuité de φ on déduit que $\varphi(x) = x$, c'est-à-dire que x est un point fixe de φ.

Si $y \in F$ est un autre point fixe de φ alors :

$$\|x - y\| = \|\varphi(x) - \varphi(y)\| \leq \lambda \|x - y\|$$

et nécessairement $x = y$ si $\lambda \in [0, 1[$. ■

Remarque 1.5 : *Si $(E, \|\cdot\|)$ est un espace de Banach et F une partie fermée de E, alors une application $\varphi : F \to F$ telle que $\|\varphi(x) - \varphi(y)\| < \|x - y\|$ pour tous $x \neq y$ dans F n'a pas nécessairement de point fixe dans F. Par exemple, la fonction φ définie sur $F = [1, +\infty[$ par $\varphi(x) = x + \dfrac{1}{x}$ est telle que $0 \leq \varphi'(x) = 1 - \dfrac{1}{x^2} < 1$ pour tout $x \in F$. Avec le théorème des accroissements finis on déduit alors que $|\varphi(x) - \varphi(y)| < |x - y|$ pour tous $x \neq y$ dans F et l'équation $\varphi(x) = x$ n'a pas de solution dans F.*

Mais dans le cas où F est compacte on a un unique point fixe limite de toute suite d'approximations successives (exercice 1.15).

Lemme 1.1 : *Soient $p \in \mathbb{N} - \{0\}$ et $(x_n)_{n\in\mathbb{N}}$ une suite de points dans un espace vectoriel normé $(E, \|\cdot\|)$ telle que pour tout $r \in \{0, \ldots, p-1\}$ la suite extraite $(x_{jp+r})_{j\in\mathbb{N}}$ converge vers $x \in E$. La suite $(x_n)_{n\in\mathbb{N}}$ converge alors vers x.*

Démonstration : Soit $\epsilon > 0$. Pour tout $r \in \{0, \ldots, p-1\}$, on peut trouver un entier j_r tel que :

$$\forall j \geq j_r, \quad \|x_{jp+r} - x\| < \epsilon.$$

On note $j_\epsilon = \max_{0 \leq r \leq p-1} j_r$. Tout entier $n \in \mathbb{N}$ s'écrit de manière unique $n = jp + r$ avec $j \in \mathbb{N}$ et $r \in \{0, \ldots, p-1\}$ et pour $j \geq j_\epsilon$ on a $\|x_n - x\| < \epsilon$. On déduit donc que :

$$\forall n \geq (j_\epsilon + 1) p, \quad \|x_n - x\| < \epsilon.$$

On a ainsi prouvé que la suite $(x_n)_{n \in \mathbb{N}}$ converge vers x. ■

Théorème 1.8 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé complet et $\varphi : E \to E$ telle qu'il existe un entier $p \in \mathbb{N} - \{0\}$ pour lequel $\varphi^p = \underbrace{\varphi \circ \cdots \circ \varphi}_{p \text{ fois}}$ soit strictement contractante. L'application φ admet alors un unique point fixe et ce point fixe est la limite de toute suite d'approximations successives.*

Démonstration : L'application φ^p est strictement contractante sur l'espace complet E et admet donc un unique point fixe $x \in E$. De $\varphi^p(x) = x$ on déduit que $\varphi^p(\varphi(x)) = \varphi(\varphi^p(x)) = \varphi(x)$ et $\varphi(x)$ est aussi point fixe de φ^p. On a donc nécessairement $\varphi(x) = x$ du fait de l'unicité du point fixe de φ^p. L'application φ admet donc un point fixe dans E.

En considérant que tout point fixe de φ est aussi point fixe de φ^p (si $\varphi(x) = x$ alors $\varphi^p(x) = \varphi^{p-1}(\varphi(x)) = \varphi^{p-1}(x) = \ldots = x$) on déduit que φ admet un unique point fixe dans E.

Si $(x_n)_{n \in \mathbb{N}}$ est définie par $x_0 \in E$ et $x_{n+1} = \varphi(x_n)$ pour tout $n \in \mathbb{N}$, on a alors pour tout $r \in \{0, \ldots, p-1\}$ et tout $j \in \mathbb{N}$:

$$x_{jp+r} = \varphi^{jp+r}(x_0) = (\varphi^p)^j(\varphi^r(x_0)),$$

c'est-à-dire que $(x_{jp+r})_{j \in \mathbb{N}}$ est une suite d'approximations successives pour φ^p de valeur initiale $\varphi^r(x_0)$. On a donc $\lim_{j \to +\infty} x_{jp+r} = x$ pour tout $r \in \{0, \ldots, p-1\}$ ce qui équivaut à dire que $\lim_{n \to +\infty} x_n = x$. ■

4. Applications linéaires continues

L'étude de la continuité est plus simple dans le cas particulier des applications linéaires.

Théorème 1.9 : *Soient $(E, \|\cdot\|)$, $(F, \|\cdot\|')$ deux espaces vectoriels normés et u une application linéaire de E dans F. Les assertions suivantes sont équivalentes :*

(i) *u est continue en 0 ;*
(ii) *u est continue sur E ;*
(iii) *u est bornée sur la sphère [resp. boule] unité de $(E, \|\cdot\|)$;*
(iv) *il existe une constante réelle c telle que :*

$$\forall x \in E, \quad \|u(x)\|' \leq c\|x\|.$$

Démonstration :

(i) $\Rightarrow$ (ii) Si u est continue en 0, pour $\epsilon > 0$ donné on peut trouver $\eta > 0$ tel que :

$$x \in E, \quad \|x\| \leq \eta \quad \Rightarrow \quad \|u(x)\|' \leq \epsilon.$$

En utilisant la linéarité de u on déduit alors que pour x_0, x dans E tels que $\|x - x_0\| \leq \eta$ on a $\|u(x) - u(x_0)\|' \leq \epsilon$, ce qui prouve la continuité (et même la continuité uniforme) de f sur E.

(ii) $\Rightarrow$ (iii) Si u est continue sur E elle est en particulier continue en 0 et il existe un réel $\eta > 0$ tel que :

$$x \in E, \quad \|x\| \leq \eta \quad \Rightarrow \quad \|u(x)\|' \leq 1.$$

Pour tout x dans la sphère [resp. boule] unité de $(E, \|\cdot\|)$ on a $\|x\| = 1$ [resp. $\|x\| \leq 1$] de sorte que $\|\eta x\| = \eta$ [resp. $\|\eta x\| \leq \eta$] et avec la linéarité de u on déduit que $\|u(x)\|' \leq \frac{1}{\eta}$. On a donc ainsi prouvé que u est bornée sur la sphère [resp. boule] unité de $(E, \|\cdot\|)$.

(iii) $\Rightarrow$ (iv) Si u est bornée sur la sphère [resp. boule] unité de $(E, \|\cdot\|)$, il existe un réel $c > 0$ tel que $\|u(x)\|' \leq c$ pour tout $x \in E$ tel que $\|x\| = 1$ [resp. $\|x\| \leq 1$]. En remarquant que pour tout vecteur x non nul dans E le vecteur $\frac{1}{\|x\|}x$ est dans la sphère (et la boule) unité de $(E, \|\cdot\|)$ et en utilisant la linéarité de u on déduit que $\|u(x)\|' \leq c\|x\|$, cette inégalité étant aussi vérifiée pour $x = 0$.

(iv) $\Rightarrow$ (i) Cette implication est évidente. ■

Une application linéaire continue est donc caractérisée par le fait d'être bornée sur la sphère unité. Pour cette raison une telle application est également appelée opérateur borné. La norme d'un tel opérateur borné est alors définie par :

$$\|u\| = \sup_{\substack{x \in E \\ \|x\|=1}} \|u(x)\|'.$$

Corollaire 1.2 : *Soient $(E, \|\cdot\|)$ et $(F, \|\cdot\|')$ deux espaces vectoriels normés et u une application linéaire surjective de E dans F. Les assertions suivantes sont équivalentes :*

(i) *u est un homéomorphisme de E sur F ;*

(ii) *il existe des constantes réelles strictement positives α et β telles que :*

$$\forall x \in E, \quad \alpha\|x\| \leq \|u(x)\|' \leq \beta\|x\|;$$

(iii) *il existe des constantes réelles strictement positives α et β telles que :*

$$(x \in E, \quad \|x\| = 1) \quad \Rightarrow \quad \alpha \leq \|u(x)\|' \leq \beta.$$

Démonstration :

(i) $\Rightarrow$ (ii) Si u est un homéomorphisme de E sur F alors u et u^{-1} sont continues et il existe deux constantes $\beta > 0$ et $\gamma > 0$ telles que :

$$\begin{aligned} \forall x \in E, \quad & \|u(x)\|' \leq \beta \|x\|, \\ \forall y \in F, \quad & \left\|u^{-1}(y)\right\| \leq \gamma \|y\|'. \end{aligned}$$

Comme tout vecteur x de E s'écrit de manière unique $x = u^{-1}(y)$ avec y dans F, on déduit des inégalités précédentes que :

$$\forall x \in E, \quad \frac{1}{\gamma} \|x\| \leq \|u(x)\|' \leq \beta \|x\|.$$

(ii) $\Rightarrow$ (iii) Cette implication est évidente.

(iii) $\Rightarrow$ (i) Les inégalités $\|u(x)\|' \leq \beta$ pour x dans la sphère unité de E signifient que l'application linéaire u est continue. En remarquant que pour tout vecteur x non nul dans E le vecteur $\frac{1}{\|x\|}x$ est dans la sphère unité de $(E, \|\cdot\|)$ et en utilisant la linéarité de u on déduit des inégalités (iii) que :

$$\forall x \in E - \{0\}, \quad \alpha \|x\| \leq \|u(x)\|' \leq \beta \|x\|,$$

ces inégalités étant encore vérifiées pour $x = 0$. Si $u(x) = 0$ on a alors nécessairement $x = 0$, $(\alpha > 0)$, c'est-à-dire que l'application linéaire u est injective. Cette application étant supposée surjective, on déduit que c'est un isomorphisme de E sur F. Tout vecteur x de E s'écrivant de manière unique $x = u^{-1}(y)$ avec y dans F, les inégalités $\alpha \|x\| \leq \|u(x)\|'$ pour tout x dans E sont équivalentes à $\left\|u^{-1}(y)\right\| \leq \frac{1}{\alpha} \|y\|'$ pour tout y dans F, ce qui équivaut à la continuité de u^{-1}. ■

Définition 1.17 : *On dit que deux normes $\|\cdot\|$ et $\|\cdot\|'$ sont équivalentes sur un espace vectoriel E si l'application identité :*

$$\begin{array}{rccc} Id: & (E, \|\cdot\|) & \rightarrow & (E, \|\cdot\|') \\ & x & \mapsto & x \end{array}$$

est un homéomorphisme.

On définit ainsi une relation d'équivalence sur l'ensemble des normes de E.

Du corollaire 1.2 on déduit facilement le résultat suivant.

Théorème 1.10 : *Deux normes $\|\cdot\|$ et $\|\cdot\|'$ sur un espace vectoriel E sont équivalentes si et seulement si il existe deux constantes réelles α et β strictement positives telles que :*

$$\forall x \in E, \quad \alpha \|x\| \leq \|x\|' \leq \beta \|x\|.$$

De ce résultat on déduit que deux normes équivalentes sur E définissent la même topologie.

Dans ce qui suit on va voir qu'une application linéaire de rang fini est continue si et seulement si son noyau est fermé. De ce résultat, conséquence de la complétude de $\mathbb{K}$, on va déduire que sur un espace vectoriel de dimension finie toutes les normes sont équivalentes. Une autre démonstration, plus classique, de cette équivalence des normes en dimension finie est une conséquence de la compacité locale de $\mathbb{K}$.

On va tout d'abord caractériser les formes linéaires continues sur un espace vectoriel normé.

Lemme 1.2 : *Si φ est une forme linéaire non nulle sur un $\mathbb{K}$-espace vectoriel E, alors il existe un vecteur non nul a dans E tel que :*

$$E = \ker(\varphi) \oplus \mathbb{K}a.$$

Démonstration : La forme linéaire φ étant non nulle, on peut trouver un vecteur a dans E tel que $\varphi(a) \neq 0$. Ce vecteur a est nécessairement non nul. Pour tout vecteur x dans E, le vecteur $h = x - \dfrac{\varphi(x)}{\varphi(a)}a$ est dans le noyau de φ et en écrivant que $x = h + \dfrac{\varphi(x)}{\varphi(a)}a$ on déduit que $E = \ker(\varphi) + \mathbb{K}a$. Si x est dans $\ker(\varphi) \cap \mathbb{K}a$ on a alors $x = \lambda a$ et $\lambda\varphi(a) = \varphi(x) = 0$ avec $\varphi(a) \neq 0$, ce qui entraîne $\lambda = 0$ et $x = 0$. On a donc $\ker(\varphi) \cap \mathbb{K}a = \{0\}$ et $E = \ker(\varphi) \oplus \mathbb{K}a$. ∎

Théorème 1.11 : *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Une forme linéaire φ sur E est continue si et seulement si son noyau* $\ker(\varphi)$ *est fermé dans $(E, \|\cdot\|)$.*

Démonstration : Soient φ une forme linéaire continue sur E et $(x_n)_{n\in\mathbb{N}}$ une suite de points de $\ker(\varphi)$ qui converge vers $x \in E$. Avec la continuité de φ on a $\varphi(x) = \lim\limits_{n\to+\infty} \varphi(x_n) = 0$, c'est-à-dire que $x \in \ker(\varphi)$. On a donc ainsi montré que si φ est continue alors son noyau est fermé. On peut aussi dire que $\ker(\varphi)$ est un fermé comme image réciproque du fermé $\{0\}$ par l'application continue φ.

Supposons réciproquement que $\ker(\varphi)$ soit fermé dans $(E, \|\cdot\|)$. Dire que φ est non continue équivaut à dire qu'elle n'est pas bornée sur la sphère unité de $(E, \|\cdot\|)$. Dans ce cas on peut trouver une suite $(x_n)_{n\in\mathbb{N}}$ de points de E telle que :

$$\forall n \in \mathbb{N}, \quad \|x_n\| = 1, \quad |\varphi(x_n)| \geq n.$$

On considère une décomposition $E = \ker(\varphi) \oplus \mathbb{K}a$ avec $\varphi(a) \neq 0$ et on écrit, pour tout entier n, $x_n = y_n + \lambda_n a$ avec $y_n \in \ker(\varphi)$ et $\lambda_n = \dfrac{\varphi(x_n)}{\varphi(a)} \in \mathbb{K}$. Pour $n \geq 1$ on a $|\varphi(x_n)| \geq n > 0$ et on peut écrire $a = \dfrac{1}{\lambda_n}x_n + z_n$ avec $z_n = -\dfrac{1}{\lambda_n}y_n \in \ker(\varphi)$. Mais on a alors pour tout entier $n \geq 1$:

$$\|a - z_n\| = \frac{\|x_n\|}{|\lambda_n|} = \frac{1}{|\lambda_n|} = \frac{|\varphi(a)|}{|\varphi(x_n)|} \leq \frac{|\varphi(a)|}{n}$$

et a n'appartenant pas à $\ker(\varphi)$ est limite d'une suite de points de $\ker(\varphi)$, ce qui est en contradiction avec $\ker(\varphi)$ fermé. On a donc ainsi montré que si $\ker(\varphi)$ est fermé dans $(E, \|\cdot\|)$ alors φ est continue. ∎

Lemme 1.3 : *Soient E, F deux $\mathbb{K}$-espaces vectoriels et u une application linéaire de E dans F. L'application u est de rang r si et seulement si il existe des formes linéaires sur E, $\varphi_1, \ldots, \varphi_r$ linéairement indépendantes et des vecteurs de F, $a_1, \ldots, a_r$ linéairement indépendants tels que :*

$$\forall x \in E, \quad u(x) = \sum_{i=1}^{r} \varphi_i(x)\, a_i.$$

Démonstration : Si $u \in \mathcal{L}(E, F)$ est de rang r, alors son image $\mathrm{Im}(u)$ est un sous-espace vectoriel de F de dimension r et, en notant $\{a_1, \ldots, a_r\}$ une base de $\mathrm{Im}(u)$, pour tout x dans E, on peut trouver des scalaires uniquement déterminés $\varphi_1(x), \ldots, \varphi_r(x)$ tels que $u(x) = \sum_{i=1}^{r} \varphi_i(x)\, a_i$. De la linéarité de u et de l'unicité de l'écriture d'un vecteur dans une base on déduit que les applications φ_i sont linéaires.

Supposons le système $\{\varphi_1, \ldots, \varphi_r\}$ dans E^* (dual algébrique de E) lié avec, par exemple, $\varphi_1 = \sum_{i=2}^{r} \lambda_i \varphi_i$. On a alors, pour tout x dans E :

$$u(x) = \left(\sum_{i=2}^{r} \lambda_i \varphi_i(x) \right) a_1 + \sum_{i=2}^{r} \varphi_i(x)\, a_i = \sum_{i=2}^{r} \varphi_i(x)\, (\lambda_i a_1 + a_i)$$

et le système de $r-1$ vecteurs $\{\lambda_i a_1 + a_i \mid 2 \leq i \leq r\}$ engendre $\mathrm{Im}(u)$, ce qui est en contradiction avec $\dim(\mathrm{Im}(u)) = r$. Le système $\{\varphi_1, \ldots, \varphi_r\}$ est donc libre.

La réciproque est évidente. ∎

Théorème 1.12 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé et F un sous-espace vectoriel fermé de $(E, \|\cdot\|)$. Pour tout sous-espace vectoriel G de dimension finie dans E, le sous-espace vectoriel $H = F + G$ est fermé dans $(E, \|\cdot\|)$.*

Démonstration : On procède par récurrence sur la dimension $p \geq 1$ de G. Pour $p = 1$ il s'agit de montrer que pour tout vecteur non nul a de E, $H = F + \mathbb{K}a$ est fermé. Si $a \in F$ alors $H = F$ est fermé. On suppose donc que $a \notin F$ et on a alors $H = F \oplus \mathbb{K}a$. Tout vecteur x de H s'écrit de manière unique $x = y + \varphi(x)\, a$ avec $y \in F$ et $\varphi(x) \in \mathbb{K}$. De l'unicité d'une telle écriture on déduit que φ est une forme linéaire sur H. De plus le noyau de φ étant $\ker(\varphi) = F$ fermé dans $(E, \|\cdot\|)$, c'est donc aussi un fermé de $(H, \|\cdot\|)$ et φ est continue de H dans $\mathbb{K}$. De cette continuité et de la complétude de $\mathbb{K}$ on va déduire que H est fermé. Soit donc $(x_n)_{n \in \mathbb{N}}$ une suite de points de H qui converge vers $x \in E$. Tout vecteur x_n s'écrit $x_n = y_n + \varphi(x_n)\, a$ où $(y_n)_{n \in \mathbb{N}}$ est une suite de points de F. La suite $(x_n)_{n \in \mathbb{N}}$ étant convergente dans E, elle est de Cauchy et, avec la continuité de φ, on déduit que la suite $(\varphi(x_n))_{n \in \mathbb{N}}$ est de Cauchy dans $\mathbb{K}$. Le corps $\mathbb{K}$ étant complet, on déduit que la suite $(\varphi(x_n))_{n \in \mathbb{N}}$ converge vers un scalaire λ. On déduit alors que la suite $(y_n)_{n \in \mathbb{N}}$ converge vers $y = x - \lambda a$. L'espace F étant fermé, on a $y \in F$ et $x = y + \lambda a$ est dans H. On a donc ainsi montré que H est fermé.

Supposons maintenant le résultat acquis pour tous les sous-espaces vectoriels de E de dimension $p \geq 1$ et soit G de dimension $p+1$. En notant $\{a_1, \ldots, a_{p+1}\}$ une base de G on peut écrire :

$$H = F + G = \left(F + \bigoplus_{j=1}^{p} \mathbb{K} a_j \right) + \mathbb{K} a_{p+1}$$

et on conclut facilement avec l'hypothèse de récurrence et l'étude du cas $p = 1$. ■

Corollaire 1.3 : *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Tout sous-espace vectoriel de dimension finie de E est fermé dans E.*

Démonstration : Il suffit de prendre $F = \{0\}$ dans le théorème 1.12. ■

Théorème 1.13 : *Soient $(E, \|\cdot\|)$, $(F, \|\cdot\|')$ deux $\mathbb{K}$-espaces vectoriels normés et u une application linéaire de E dans F de rang fini. L'application u est continue si et seulement si son noyau $\ker(u)$ est fermé dans $(E, \|\cdot\|)$.*

Démonstration : Si u est continue alors $\ker(u)$ est fermé dans $(E, \|\cdot\|)$ comme image réciproque du fermé $\{0\}$ de $(F, \|\cdot\|')$ par l'application continue u.

Réciproquement, supposons que $\ker(u)$ soit fermé dans $(E, \|\cdot\|)$. L'application linéaire u étant de rang fini, il existe des formes linéaires $\varphi_1, \ldots, \varphi_r$, linéairement indépendantes dans E^* et un système libre $\{a_1, \ldots, a_r\}$ dans F tels que $u = \sum_{i=1}^{r} \varphi_i a_i$. On a alors $\ker(u) = \bigcap_{j=1}^{r} \ker(\varphi_j) \subset \ker(\varphi_j)$ pour tout j compris entre 1 et r. On peut alors écrire que $\ker(\varphi_j) = \ker(u) \oplus H_j$ et la restriction de u à H_j est injective ($u(x) = 0$ et $x \in H_j$ équivaut à $x \in \ker(u) \cap H_j$) de H_j dans $\operatorname{Im}(u)$ qui est de dimension finie. En définitive, pour tout j compris entre 1 et r le sous-espace vectoriel H_j est de dimension finie dans E et, avec le théorème 1.12, on déduit que $\ker(\varphi_j) = \ker(u) \oplus H_j$ est fermé et donc que la forme linéaire φ_j est continue. La continuité de $u = \sum_{i=1}^{r} \varphi_i a_i$ en résulte alors immédiatement. ■

Corollaire 1.4 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé de dimension finie et $(F, \|\cdot\|')$ un espace vectoriel normé. Toute application linéaire de E dans F est continue.*

Démonstration : Si E est de dimension finie alors toute application linéaire u de E dans F est de rang fini. De plus, avec $\dim(E) = \dim(\ker(u)) + \dim(\operatorname{Im}(u))$, on déduit que $\ker(u)$ est un sous-espace vectoriel de dimension finie de E. Il est donc fermé et u est continue. ■

5. Espaces vectoriels normés de dimension finie

Lemme 1.4 : *Soient E un espace vectoriel de dimension $n \geq 1$, $\{e_1, \ldots, e_n\}$ une base de E et $\|\cdot\|_\infty$ la norme définie sur E par :*

$$\forall x = \sum_{i=1}^{n} x_i e_i \in E, \quad \|x\|_\infty = \max_{1 \leq i \leq n} |x_i| .$$

La boule unité B_∞ et la sphère unité S_∞ sont compactes dans $(E, \|\cdot\|_\infty)$.

Démonstration : Soit $\left(x^{(k)}\right)_{k\in\mathbb{N}}$ une suite dans B_∞ [resp. S_∞] avec, pour tout $k \in \mathbb{N}$, $x^{(k)} = \left(x_j^{(k)}\right)_{1\le j\le n}$. Pour tout entier j compris entre 1 et n on a :

$$\left|x_j^{(k)}\right| \le \left\|x^{(k)}\right\|_\infty \le 1.$$

De la suite bornée $\left(x_1^{(k)}\right)_{k\in\mathbb{N}}$ on peut alors extraire une sous-suite $\left(x_1^{\left(\varphi_1(k)\right)}\right)_{k\in\mathbb{N}}$ qui converge vers un scalaire x_1 vérifiant $|x_1| \le 1$. Puis de la suite bornée $\left(x_2^{\left(\varphi_1(k)\right)}\right)_{k\in\mathbb{N}}$ on peut extraire une sous-suite $\left(x_2^{\left(\varphi_2(k)\right)}\right)_{k\in\mathbb{N}}$ qui converge vers un scalaire x_2 vérifiant $|x_2| \le 1$. Et en continuant ainsi on extrait une suite $\left(x^{\left(\varphi(k)\right)}\right)_{k\in\mathbb{N}}$ telle que :

$$\forall j \in \{1, 2, \ldots, n\}, \quad \lim_{k\to+\infty} x_j^{\left(\varphi(k)\right)} = x_j,$$

avec $|x_j| \le 1$. On a alors $\lim\limits_{k\to+\infty} \left\|x^{\left(\varphi(k)\right)} - x\right\|_\infty = 0$ où $x = (x_j)_{1\le j\le n}$ est dans B_∞ [resp. S_∞], c'est-à-dire que la suite $\left(x^{\left(\varphi(k)\right)}\right)_{k\in\mathbb{N}}$ converge vers x dans B_∞ [resp. S_∞]. On a donc ainsi montré que B_∞ [resp. S_∞] est compacte dans $(E, \|\cdot\|_\infty)$. ■

Lemme 1.5 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé et F un sous-espace vectoriel fermé de E distinct de E. Pour tout x dans E on note :*

$$d(x, F) = \inf_{y\in F} \|x - y\|.$$

Pour tout réel $\epsilon > 0$ il existe un vecteur x dans la sphère unité de $(E, \|\cdot\|)$ tel que $d(x, F) \ge 1 - \epsilon$.

Démonstration : Pour $\epsilon \ge 1$ le résultat est évident. On suppose donc que $\epsilon \in \left]0, 1\right[$. Si F est fermé dans E alors pour tout $y \in E - F$ on a $d(y, F) > 0$ (exercice 1.18). Pour $\epsilon \in \left]0, 1\right[$ on a $\dfrac{1}{1-\epsilon} > 1$ et pour $y \in E - F$ on peut trouver $z \in F$ tel que :

$$0 < d(y, F) \le \|y - z\| < \frac{d(y, F)}{1 - \epsilon}.$$

Le vecteur $x = \dfrac{1}{\|y - z\|}(y - z)$ est alors dans la sphère unité de $(E, \|\cdot\|)$ et pour tout vecteur $t \in F$ on a :

$$\|x - t\| = \frac{1}{\|y - z\|} \left\|y - \left(z + \|y - z\| \, t\right)\right\|,$$

avec $u = z + \|y - z\| \, t \in F$, de sorte que :

$$\|x - t\| = \frac{1}{\|y - z\|} \|y - u\| \ge \frac{d(y, F)}{\|y - z\|} > 1 - \epsilon.$$

On a donc $d(x, F) = \inf\limits_{t\in F} \|x - t\| \ge 1 - \epsilon$. ■

Le théorème qui suit donne plusieurs caractérisations des $\mathbb{K}$-espaces vectoriels de dimension finie.

Théorème 1.14 : *Soit E un espace vectoriel sur $\mathbb{K}$. Les assertions suivantes sont équivalentes.*

(i) *E est de dimension finie sur $\mathbb{K}$.*
(ii) *Toutes les normes sur E sont équivalentes.*
(iii) *Quelle que soit la norme choisie sur E toute forme linéaire définie sur $(E, \|\cdot\|)$ est continue.*
(iv) *Quelle que soit la norme choisie sur E, la sphère [resp. boule] unité de $(E, \|\cdot\|)$ pour cette norme est compacte.*
(v) *Quelle que soit la norme choisie sur E, les compacts de $(E, \|\cdot\|)$ sont les fermés bornés.*

Démonstration :

(i) $\Rightarrow$ (ii) On suppose que E est de dimension finie et on se donne deux normes $\|\cdot\|$ et $\|\cdot\|'$ sur E. L'application linéaire $Id : (E, \|\cdot\|) \to (E, \|\cdot\|')$ [resp. $Id : (E, \|\cdot\|') \to (E, \|\cdot\|)$] étant de rang fini à noyau fermé, elle est donc continue. L'application $Id : (E, \|\cdot\|) \to (E, \|\cdot\|')$ est donc un homéomorphisme, c'est-à-dire que les normes $\|\cdot\|$ et $\|\cdot\|'$ sont équivalentes sur E.

(ii) $\Rightarrow$ (iii) On suppose que toutes les normes sur E sont équivalentes. Si $\|\cdot\|$ est une norme sur E et φ une forme linéaire sur E alors l'application :

$$N : x \mapsto \|x\| + |\varphi(x)|$$

définit une norme sur E. Elle est donc équivalente à $\|\cdot\|$ et il existe une constante $\alpha > 0$ telle que :

$$\forall x \in E, \quad N(x) \leq \alpha \|x\|.$$

On a donc :

$$\forall x \in E, \quad |\varphi(x)| \leq (\alpha - 1) \|x\|,$$

ce qui équivaut à la continuité de la forme linéaire φ.

(iii) $\Rightarrow$ (iv) On suppose que quelle que soit la norme choisie sur E toute forme linéaire sur E est continue. Si $(E, \|\cdot\|)$ est de dimension infinie on peut trouver un système libre infini dénombrable $\{e_n \mid n \in \mathbb{N}\}$ tel que $\|e_n\| = 1$ pour tout $n \in \mathbb{N}$. En désignant par G un supplémentaire dans E de $H = \mathrm{Vect}\{e_n \mid n \in \mathbb{N}\}$ on définit la forme linéaire φ sur E par :

$$\begin{cases} \forall x \in G, & \varphi(x) = 0, \\ \forall n \in \mathbb{N}, & \varphi(e_n) = n. \end{cases}$$

L'application linéaire ainsi définie n'est pas bornée sur la sphère unité de $(E, \|\cdot\|)$ et en conséquence n'est pas continue, ce qui est en contradiction avec l'hypothèse de départ. L'espace vectoriel E est donc de dimension finie.

On a donc ainsi montré que les assertions (i), (ii) et (iii) sont équivalentes. En particulier si (iii) est vérifiée alors toutes les normes sur E sont équivalentes et la compacité de la sphère [resp. boule] unité de $(E, \|\cdot\|)$ résulte de la compacité de la sphère [resp. boule] unité de $(E, \|\cdot\|_\infty)$.

(iv) ⇒ (v) On suppose que quelle que soit la norme choisie sur E, la sphère [resp. boule] unité de E pour cette norme est compacte. On sait déjà que toute partie compacte d'un espace vectoriel normé $(E, \|\cdot\|)$ est fermée et bornée (théorème 1.3). Réciproquement, soit $\mathcal{C}$ une partie non vide fermée et bornée dans $(E, \|\cdot\|)$. Il existe une constante $\lambda > 0$ telle que $\|x\| \leq \lambda$ pour tout x dans $\mathcal{C}$ et pour toute suite $(x_n)_{n\in\mathbb{N}}$ de points de $\mathcal{C}$, la suite $\left(\frac{1}{\lambda}x_n\right)_{n\in\mathbb{N}}$ est dans la boule unité de $(E, \|\cdot\|)$. Cette boule étant compacte, on peut extraire une sous-suite $\left(\frac{1}{\lambda}x_{\varphi(n)}\right)_{n\in\mathbb{N}}$ qui converge vers $y \in E$. La suite $(x_{\varphi(n)})_{n\in\mathbb{N}}$ converge alors vers $x = \lambda y$ et $x \in \mathcal{C}$ puisque $\mathcal{C}$ est fermée. On a donc ainsi montré que $\mathcal{C}$ est compacte dans $(E, \|\cdot\|)$.

(v) ⇒ (i) On suppose que quelle que soit la norme choisie sur E, les compacts de $(E, \|\cdot\|)$ sont les fermés bornés. Si E est de dimension infinie on peut trouver un système libre infini dénombrable $\{e_n \mid n \in \mathbb{N}\}$. Pour tout entier n, on désigne par E_n le sous-espace vectoriel de E engendré par $\{e_k \mid 1 \leq k \leq n\}$. On a alors $E_p \subsetneq E_q$ pour $q > p$. De plus chaque E_n est de dimension finie dans E, donc fermé dans E et aussi dans E_{n+1}. En utilisant le lemme 1.5, on peut construire une suite $(x_n)_{n\geq 1}$ dans la sphère unité de $(E, \|\cdot\|)$ telle que $x_n \in E_n$ et $d(x_n, E_{n-1}) \geq \frac{1}{2}$ pour tout $n \geq 1$. Mais alors, on a :

$$\forall q > p, \quad \|x_q - x_p\| \geq d(x_q, E_{q-1}) \geq \frac{1}{2}$$

et il est impossible d'extraire de la suite $(x_n)_{n\geq 1}$ une sous-suite convergente, ce qui est en contradiction avec la compacité de la sphère unité (c'est un fermé borné) de $(E, \|\cdot\|)$. L'espace vectoriel E est donc nécessairement de dimension finie. ∎

***Remarque 1.6** : L'équivalence (i) ⇔ (iv) est le théorème de Riesz.*

On peut donc conclure que sur un espace vectoriel de dimension finie on a une seule topologie compatible avec la structure d'espace vectoriel.

Une autre démonstration de l'équivalence des normes en dimension finie, basée sur la compacité locale de $\mathbb{K}$, peut se faire comme suit.

On désigne par E un espace vectoriel de dimension finie sur $\mathbb{K}$, par $\{e_1, \ldots, e_n\}$ une base de E et on note $\|\cdot\|_\infty$ la norme définie sur E par :

$$\forall x = \sum_{i=1}^{n} x_i e_i \in E, \quad \|x\|_\infty = \max_{1\leq i\leq n} |x_i|.$$

On sait déjà que la boule unité et la sphère unité de $(E, \|\cdot\|_\infty)$ sont compactes (lemme 1.4).

Lemme 1.6 : *Pour toute norme $\|\cdot\|$ sur E l'application :*

$$\begin{array}{ccc} (E, \|\cdot\|_\infty) & \to & \mathbb{R} \\ x & \mapsto & \|x\| \end{array}$$

est continue.

Démonstration : Pour tout $x = \sum_{j=1}^{n} x_j e_j$ dans E on a :

$$\|x\| \leq \beta \|x\|_\infty ,$$

où $\beta = \sum_{j=1}^{n} \|e_j\| > 0$. On a donc pour x et y dans E :

$$|\|x\| - \|y\|| \leq \|x - y\| \leq \beta \|x - y\|_\infty$$

et la continuité en résulte. ■

On retrouve alors le résultat suivant.

Théorème 1.15 : *Toutes les normes sur E sont équivalentes.*

Démonstration : Il suffit de montrer que toute norme $\|\cdot\|$ sur E est équivalente à $\|\cdot\|_\infty$.

On a déjà vu que pour tout $x = \sum_{i=1}^{n} x_i e_i$ dans E on a :

$$\|x\| \leq \left(\sum_{j=1}^{n} \|e_j\|\right) \|x\|_\infty = \beta \|x\|_\infty .$$

On a également vu que la sphère unité

$$S_\infty = \{x \in E \mid \|x\|_\infty = 1\}$$

est compacte dans $(E, \|\cdot\|_\infty)$ et que l'application $x \mapsto \|x\|$ est continue de $(E, \|\cdot\|_\infty)$ dans $\mathbb{R}$. On peut donc poser :

$$\alpha = \inf_{x \in S_\infty} \|x\| = \lim_{n \to +\infty} \|x_n\| ,$$

où $(x_n)_{n \in \mathbb{N}}$ est une suite de points de S_∞.

La sphère unité S_∞ étant compacte, on peut extraire de $(x_n)_{n \in \mathbb{N}}$ une sous-suite $(x_{\varphi(n)})_{n \in \mathbb{N}}$ qui converge vers x dans S_∞. On a alors $\alpha = \|x\| > 0$.

Enfin, pour tout x dans $E - \{0\}$, on a $\left\|\frac{x}{\|x\|_\infty}\right\| \geq \alpha$ soit $\alpha \|x\|_\infty \leq \|x\|$.

D'où l'équivalence des normes. ■

Corollaire 1.5 : *Un espace vectoriel normé de dimension finie est complet.*

Démonstration : En reprenant les notations qui précèdent il suffit de montrer que $(E, \|\cdot\|_\infty)$ est complet, ce qui se déduit immédiatement du fait que $\mathbb{K}$ est complet. ■

Remarque 1.7 : *Avec l'équivalence des normes sur un espace vectoriel de dimension finie on peut retrouver le fait que si $(E, \|\cdot\|)$ est un espace vectoriel normé de dimension finie et $(F, \|\cdot\|')$ un espace vectoriel normé de dimension quelconque alors toute application linéaire de E dans F est continue (corollaire 1.4). En effet si $\{e_1, \ldots, e_n\}$ est une base de E pour toute application linéaire u de E dans F on a, pour tout $x = \sum_{i=1}^{n} x_i e_i \in E$:*

$$\|u(x)\|' \leq \sum_{i=1}^{n} |x_i| \|e_i\| \leq \left(\max_{1 \leq i \leq n} \|e_i\| \right) \|x\|_1 \leq \left(\max_{1 \leq i \leq n} \|e_i\| \right) \alpha \|x\|$$

(les normes $\|\cdot\|_1$ et $\|\cdot\|$ étant équivalentes, il existe donc $\alpha > 0$ tel que, pour tout x dans E, on ait $\|x\|_1 \leq \alpha \|x\|$) et la continuité de u en résulte.

6. Exercices

Exercice 1.1 : *Vérifier que les applications $x \mapsto \|x\|_k$ ($k = \infty, 1, 2$) de l'exemple 1.1 définissent bien des normes sur E (espace vectoriel de dimension finie sur $\mathbb{K}$).*

Solution : Il est clair que pour $k = \infty$ et $k = 1$ les applications $x \mapsto \|x\|_k$ définissent bien des normes sur E.

Pour $k = 2$, on utilise le produit scalaire hermitien défini sur le $\mathbb{C}$-espace vectoriel E par :

$$\forall (x, y) \in E^2, \quad \langle x \mid y \rangle = \sum_{j=1}^{n} x_j \overline{y_j}$$

et l'inégalité de Cauchy-Schwarz pour ce produit scalaire :

$$\forall (x, y) \in E^2, \quad |\langle x \mid y \rangle| \leq \sum_{j=1}^{n} |x_j| |y_j| \leq \sqrt{\sum_{j=1}^{n} |x_j|^2} \sqrt{\sum_{j=1}^{n} |y_j|^2} = \|x\|_2 \|y\|_2$$

(voir le paragraphe 7, page 96). En écrivant pour tous vecteurs x, y dans E que :

$$\|x + y\|_2^2 = \|x\|_2^2 + 2\Re(\langle x \mid y \rangle) + \|y\|_2^2,$$

avec $\Re(\langle x \mid y \rangle) \leq |\langle x \mid y \rangle| \leq \|x\|_2 \|y\|_2$, on déduit que :

$$\|x + y\|_2^2 \leq (\|x\|_2 + \|y\|_2)^2.$$

On a donc l'inégalité triangulaire pour $\|\cdot\|_2$. Les autres propriétés d'une norme se vérifient facilement.

On peut aussi utiliser le résultat de l'exercice 1.2 avec $p = q = 2$.

Exercice 1.2 : *Soient p et q deux réels strictement positifs tels que* $\frac{1}{p}+\frac{1}{q}=1$.

1. En utilisant la concavité de la fonction ln *montrer que :*

$$\forall \lambda \in [0,1], \quad \forall u \in \mathbb{R}^+, \quad \forall v \in \mathbb{R}^+, \quad u^\lambda v^{1-\lambda} \leq \lambda u + (1-\lambda)\, v. \tag{1.2}$$

2. Soient x et y dans $\mathbb{C}^n$. En utilisant (1.2) avec

$$\lambda = \frac{1}{p}, \quad u_i = \frac{|x_i|^p}{\sum\limits_{j=1}^n |x_j|^p}, \quad v_i = \frac{|y_i|^q}{\sum\limits_{j=1}^n |y_j|^q},$$

pour x et y non nuls montrer l'inégalité de Hölder :

$$\left|\sum_{i=1}^n x_i \overline{y_i}\right| \leq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |y_i|^q\right)^{\frac{1}{q}}.$$

3. Montrer que pour tous x, y dans $\mathbb{C}^n$ on a :

$$\sum_{i=1}^n |x_i|\, |x_i+y_i|^{p-1} \leq \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^n |x_i+y_i|^p\right)^{\frac{1}{q}}.$$

4. Déduire de ce qui précède que pour tout réel $p \geq 1$ l'application

$$x \mapsto \|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}}$$

définit une norme sur $\mathbb{C}^n$.

Solution :

1. La fonction ln est indéfiniment dérivable avec $\ln''(t) = -\frac{1}{t^2} < 0$ pour tout réel $t > 0$. On déduit donc que cette fonction est concave sur $]0,+\infty[$, c'est-à-dire que pour tout réel $\lambda \in [0,1]$ et tous réels u, v strictement positifs on a :

$$\ln\big(\lambda u + (1-\lambda)\, v\big) \geq \lambda \ln(u) + (1-\lambda) \ln(v),$$

ce qui peut aussi s'écrire $\ln\big(\lambda u + (1-\lambda)\, v\big) \geq \ln\left(u^\lambda v^{1-\lambda}\right)$, ou encore, du fait que la fonction exp est strictement croissante sur $]0,+\infty[$, $u^\lambda v^{1-\lambda} \leq \lambda u + (1-\lambda)\, v$. Cette inégalité est également vérifiée pour $u=0$ ou $v=0$.

2. Pour tous vecteurs x, y dans $\mathbb{C}^n$ on note :

$$\|x\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{\frac{1}{p}} \quad \text{et} \quad \|y\|_q = \left(\sum_{i=1}^n |y_i|^q\right)^{\frac{1}{q}}.$$

En prenant $\lambda = \frac{1}{p}$ dans (1.2), on a $1 - \lambda = \frac{1}{q}$ et :

$$\forall u \geq 0, \quad \forall v \geq 0, \quad u^{\frac{1}{p}} v^{\frac{1}{q}} \leq \frac{1}{p} u + \frac{1}{q} v.$$

Si x et y sont deux vecteurs non nuls dans $\mathbb{C}^n$, en prenant pour tout entier i compris entre 1 et n :

$$u_i = \frac{|x_i|^p}{\sum_{j=1}^{n} |x_j|^p} = \frac{|x_i|^p}{\|x\|_p^p}, \quad v_i = \frac{|y_i|^q}{\sum_{j=1}^{n} |y_j|^q} = \frac{|y_i|^q}{\|y\|_q^q},$$

dans l'inégalité précédente, on obtient :

$$\frac{|x_i|}{\|x\|_p} \frac{|y_i|}{\|y\|_q} \leq \frac{1}{p} \frac{|x_i|^p}{\|x\|_p^p} + \frac{1}{q} \frac{|y_i|^q}{\|y\|_q^q}.$$

En faisant la somme de toutes ces inégalités on obtient :

$$\frac{1}{\|x\|_p \|y\|_q} \sum_{i=1}^{n} |x_i| |y_i| \leq \frac{1}{p} \sum_{i=1}^{n} \frac{|x_i|^p}{\|x\|_p^p} + \frac{1}{q} \sum_{i=1}^{n} \frac{|y_i|^q}{\|y\|_q^q}.$$

Puis avec $\sum_{i=1}^{n} \frac{|x_i|^p}{\|x\|_p^p} = \sum_{i=1}^{n} \frac{|y_i|^q}{\|y\|_q^q} = 1$ et $\frac{1}{p} + \frac{1}{q} = 1$ on déduit que :

$$\left| \sum_{i=1}^{n} x_i \overline{y_i} \right| \leq \sum_{i=1}^{n} |x_i| |y_i| \leq \|x\|_p \|y\|_q,$$

cette inégalité étant encore réalisée pour $x = 0$ ou $y = 0$.

Remarque 1.8 : *L'inégalité de Hölder s'écrit, en notant $\langle x \mid y \rangle$ le produit scalaire hermitien canonique de $\mathbb{C}^n$,*

$$|\langle x \mid y \rangle| \leq \|x\|_p \|y\|_q.$$

Pour $p = q = 2$ on retrouve l'inégalité de Cauchy-Schwarz et cette inégalité est encore valable pour $p = 1$ et $q = +\infty$.

3. Pour x, y dans $\mathbb{C}^n$ on a, en utilisant l'inégalité de Hölder :

$$\sum_{i=1}^{n} |x_i| |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |x_i + y_i|^{(p-1)q} \right)^{\frac{1}{q}},$$

soit, avec $(p-1) q = p$:

$$\sum_{i=1}^{n} |x_i| |x_i + y_i|^{p-1} \leq \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}} \left(\sum_{i=1}^{n} |x_i + y_i|^p \right)^{\frac{1}{q}}.$$

4. On vérifie facilement que $x \mapsto \|x\|_1$ définit une norme sur $\mathbb{C}^n$. On suppose donc que $p > 1$ et on note q le réel défini par $\dfrac{1}{p} + \dfrac{1}{q} = 1$. Pour x, y dans $\mathbb{C}^n$ on a :

$$\sum_{i=1}^{n} |x_i + y_i|^p \leq \sum_{i=1}^{n} |x_i + y_i|^{p-1} |x_i| + \sum_{i=1}^{n} |x_i + y_i|^{p-1} |y_i| .$$

Ce qui donne avec l'inégalité précédente :

$$\left(\|x+y\|_p\right)^p \leq \left(\|x\|_p + \|y\|_p\right) \left(\|x+y\|_p\right)^{\frac{p}{q}},$$

soit, avec $p - \dfrac{p}{q} = 1$:

$$\|x+y\|_p \leq \|x\|_p + \|y\|_p .$$

On a donc l'inégalité triangulaire pour $\|\cdot\|_p$. Les autres propriétés d'une norme se vérifient facilement.

Exercice 1.3 : *Montrer que pour tout x dans $\mathbb{C}^n$ on a $\lim\limits_{p \to +\infty} \|x\|_p = \|x\|_\infty$.*

Solution : Pour x dans $\mathbb{C}^n$ et tout entier i compris entre 1 et n on a $|x_i|^p \leq \|x\|_\infty^p$. Il en résulte que $\|x\|_p \leq n^{\frac{1}{p}} \|x\|_\infty$. D'autre part il existe un entier k compris entre 1 et n tel que $\|x\|_\infty = |x_k|$ et, avec $|x_k|^p \leq \sum\limits_{i=1}^{n} |x_i|^p$, on déduit que $\|x\|_\infty \leq \|x\|_p$. On a donc :

$$\forall x \in \mathbb{C}^n, \quad \|x\|_\infty \leq \|x\|_p \leq n^{\frac{1}{p}} \|x\|_\infty .$$

Ce qui montre au passage que les normes $\|\cdot\|_\infty$ et $\|\cdot\|_p$ sont équivalentes et par transitivité que pour tous réels $p \geq 1$ et $q \geq 1$ les normes $\|\cdot\|_p$ et $\|\cdot\|_q$ sont équivalentes sur $\mathbb{C}^n$.

Avec $\lim\limits_{p \to +\infty} n^{\frac{1}{p}} = 1$ pour tout entier $n \geq 1$ on déduit que $\lim\limits_{p \to +\infty} \|x\|_p = \|x\|_\infty$.

Exercice 1.4 : *Vérifier que les applications $f \mapsto \|f\|_k$ ($k = \infty, 1, 2$) de l'exemple 1.3 définissent bien des normes sur E (espace vectoriel des applications continues sur l'intervalle $[a, b]$ ($a < b$) et à valeurs dans $\mathbb{K}$).*

Solution : Il est clair que pour $k = \infty$ et $k = 1$ les applications $f \mapsto \|f\|_k$ définissent bien des normes sur E. Pour $f \mapsto \|f\|_1$ on utilise le fait que l'application $\varphi \mapsto \int_a^b \varphi(t)\ dt$ est une forme linéaire positive avec $\int_a^b \varphi(t)\ dt = 0$ équivalant à $\varphi = 0$ si φ est une fonction continue à valeurs positives ou nulles.

Pour $k = 2$, on utilise l'inégalité de Cauchy-Schwarz dans E :

$$\forall (f, g) \in E^2, \quad \left| \int_a^b f(t)\, \overline{g(t)}\ \mathrm{d}t \right| \leq \int_a^b |f(t)|\, |g(t)|\ \mathrm{d}t \leq \|f\|_2\, \|g\|_2,$$

qui entraîne l'inégalité triangulaire pour $\|\cdot\|_2$. Les autres propriétés d'une norme se vérifient facilement.

On peut aussi utiliser le résultat de l'exercice 1.5 avec $p = q = 2$.

Exercice 1.5 : *Montrer que sur l'espace vectoriel $E = \mathcal{C}^0([a, b], \mathbb{K})$ des applications continues sur l'intervalle $[a, b]$ $(a < b)$ et à valeurs dans $\mathbb{K}$, pour tout réel $p \geq 1$, l'application*

$$f \mapsto \|f\|_p = \sqrt[p]{\int_a^b |f(x)|^p\ \mathrm{d}x}$$

définit une norme.

Solution : Pour $p = 1$, on sait déjà que $f \mapsto \|f\|_1$ définit une norme sur E. On suppose donc que $p > 1$ et on se donne deux fonctions f, g non identiquement nulles dans E.

On applique l'inégalité de convexité :

$$\forall u \geq 0, \quad \forall v \geq 0, \quad u^{\frac{1}{p}} v^{\frac{1}{q}} \leq \frac{1}{p} u + \frac{1}{q} v,$$

où q est tel que $\frac{1}{p} + \frac{1}{q} = 1$, à :

$$u = \frac{|f(t)|^p}{\|f\|_p^p}, \quad v = \frac{|g(t)|^q}{\|g\|_q^q},$$

pour tout t dans $[a, b]$, ce qui donne :

$$\frac{|f(t)|}{\|f\|_p} \frac{|g(t)|}{\|g\|_q} \leq \frac{1}{p} \frac{|f(t)|^p}{\|f\|_p^p} + \frac{1}{q} \frac{|g(t)|^q}{\|g\|_q^q}.$$

En intégrant sur $[a, b]$ on obtient :

$$\frac{1}{\|f\|_p \|g\|_q} \int_a^b |f(t)|\, |g(t)|\ \mathrm{d}t \leq \frac{1}{p} + \frac{1}{q} = 1.$$

On a donc l'inégalité de Hölder :

$$\left|\int_a^b f(t)\,\overline{g(t)}\,\mathrm{d}t\right| \leq \int_a^b |f(t)|\,|g(t)|\ \mathrm{d}t \leq \|f\|_p\,\|g\|_q,$$

cette inégalité étant encore réalisée pour $f=0$ ou $g=0$.

Pour f, g dans E on a, en utilisant l'inégalité de Hölder :

$$\int_a^b |f(t)|\,|f(t)+g(t)|^{p-1}\ \mathrm{d}t \leq \|f\|_p \left\|(f+g)^{p-1}\right\|_q,$$

soit, avec $(p-1)\,q=p$:

$$\int_a^b |f(t)|\,|f(t)+g(t)|^{p-1}\ \mathrm{d}t \leq \|f\|_p\,\|f+g\|_p^{\frac{p}{q}}.$$

On déduit alors que :

$$\|f+g\|_p^p = \int_a^b |f(t)+g(t)|^p\ \mathrm{d}t \leq \left(\|f\|_p + \|g\|_p\right)\|f+g\|_p^{\frac{p}{q}},$$

soit, avec $p-\dfrac{p}{q}=1$:

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

On a donc l'inégalité triangulaire pour $\|\cdot\|_p$. Les autres propriétés d'une norme se vérifient facilement.

Exercice 1.6 : *Montrer que pour toute fonction f dans $E = \mathcal{C}^0([a,b],\mathbb{K})$ on a :*

$$\lim_{p\to+\infty}\|f\|_p = \|f\|_\infty.$$

Solution : On suppose f non identiquement nulle. On a, pour tout réel $p \geq 1$:

$$\left(\forall x\in[a,b],\quad 0\leq |f(x)| \leq \|f\|_\infty\right) \Rightarrow \left(0\leq \|f\|_p \leq \|f\|_\infty (b-a)^{\frac{1}{p}}\right).$$

La fonction f étant continue sur le compact $[a,b]$, il existe donc un réel $x_0\in[a,b]$ tel que $|f(x_0)| = \|f\|_\infty$. Pour $\epsilon>0$ donné, il existe $\eta>0$ tel que $0\leq |f(x_0)|-|f(x)| < \epsilon$ pour tout $x\in I_0=[a,b]\cap[x_0-\eta,x_0+\eta]$. On a alors, en notant α la longueur de l'intervalle I_0 :

$$\begin{aligned}\int_a^b |f(x)|^p\ \mathrm{d}x &\geq \int_{I_0} |f(x)|^p\ \mathrm{d}x\\ &\geq \int_{I_0}\left(|f(x_0)|-\epsilon\right)^p\ \mathrm{d}x \geq \alpha\left(\|f\|_\infty-\epsilon\right)^p.\end{aligned}$$

Ce qui donne en définitive, pour tout $\epsilon > 0$:

$$\left(\|f\|_\infty - \epsilon\right)\alpha^{\frac{1}{p}} \leq \|f\|_p \leq \|f\|_\infty \left(b-a\right)^{\frac{1}{p}}.$$

Comme $\lim\limits_{p \to +\infty} \left(\|f\|_\infty - \epsilon\right)\alpha^{\frac{1}{p}} = \|f\|_\infty - \epsilon$ et $\lim\limits_{n \to +\infty} \|f\|_\infty \left(b-a\right)^{\frac{1}{p}} = \|f\|_\infty$, on déduit qu'il existe un entier p_0 tel que :

$$\forall p \geq p_0, \quad \|f\|_\infty - 2\epsilon \leq \|f\|_p \leq \|f\|_\infty + 2\epsilon.$$

On a donc ainsi prouvé que $\lim\limits_{p \to +\infty} \left(\|f\|_p\right) = \|f\|_\infty$.

Exercice 1.7 : *Soient f et g deux fonctions à valeurs réelles définies et continues sur l'intervalle $[a, b]$, la fonction g étant à valeurs strictement positives. Calculer $\lim\limits_{p \to +\infty} \left\|fg^{\frac{1}{p}}\right\|_p$.*

Solution : La fonction g étant continue et à valeurs strictement positives sur le compact $[a, b]$, on a $m = \inf\limits_{x \in [a,b]} g(x) > 0$ et $M = \sup\limits_{x \in [a,b]} g(x) > 0$. Puis avec :

$$m\,|f|^p \leq g\,|f|^p \leq M\,|f|^p,$$

on déduit que :

$$m^{\frac{1}{p}} \|f\|_p \leq \left\|g^{\frac{1}{p}} f\right\|_p \leq M^{\frac{1}{p}} \|f\|_p.$$

On en déduit alors, en utilisant le résultat de l'exercice 1.6, que $\lim\limits_{p \to +\infty} \left\|fg^{\frac{1}{p}}\right\|_p = \|f\|_\infty$.

Exercice 1.8 : *Soit $(x_n)_{n \in \mathbb{N}}$ une suite dans un espace vectoriel normé $(E, \|\cdot\|)$. Montrer que cette suite converge vers x dans $(E, \|\cdot\|)$ si et seulement si toute suite extraite converge vers x (une suite convergente a une unique valeur d'adhérence).*

Solution : Si $\lim\limits_{n \to +\infty} x_n = x$, alors pour tout réel $\epsilon > 0$ il existe un entier n_ϵ tel que $\|x_n - x\| < \epsilon$ pour tout $n \geq n_\epsilon$, et pour toute application $\varphi : \mathbb{N} \to \mathbb{N}$ strictement croissante on a :

$$n \geq n_\epsilon \quad \Rightarrow \quad \varphi(n) \geq n \geq n_\epsilon \quad \Rightarrow \quad \left\|x_{\varphi(n)} - x\right\| < \epsilon.$$

On a donc $\lim\limits_{n \to +\infty} x_{\varphi(n)} = x$, pour toute suite extraite de $(x_n)_{n \in \mathbb{N}}$.

Réciproquement, si toute suite extraite de $(x_n)_{n \in \mathbb{N}}$ converge vers x, alors en particulier la suite $(x_n)_{n \in \mathbb{N}}$ converge vers x.

Exercice 1.9 : *Soit $(x_n)_{n\in\mathbb{N}}$ une suite dans un espace vectoriel normé $(E, \|\cdot\|)$. Montrer que l'ensemble des valeurs d'adhérence de cette suite est un fermé dans $(E, \|\cdot\|)$.*

Solution : On désigne par $\mathcal{A}$ l'ensemble des valeurs d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$:

$$x \in \mathcal{A} \quad \Leftrightarrow \quad \forall \epsilon > 0,\ \forall n \in \mathbb{N},\ \exists p \geq n\ ;\ \|x_p - x\| < \epsilon$$

(tout voisinage de x contient des x_p pour une infinité d'indices p).

Soit $x \notin \mathcal{A}$, alors :

$$\exists \epsilon > 0,\ \exists n \in \mathbb{N} \mid\ \forall p \geq n,\ \|x_p - x\| \geq \epsilon$$

et $B\left(x, \frac{\epsilon}{2}\right) \subset E - \mathcal{A}$. En effet, pour tout $y \in B\left(x, \frac{\epsilon}{2}\right)$, on a :

$$\forall p \geq n,\ \|x_p - y\| \geq \|x_p - x\| - \|y - x\| \geq \epsilon - \frac{\epsilon}{2} = \frac{\epsilon}{2}$$

et donc $y \notin \mathcal{A}$. On a donc ainsi montré que $\mathcal{A}$ est fermé.

Exercice 1.10 : *Soit $(x_n)_{n\in\mathbb{N}}$ une suite dans un espace vectoriel normé de dimension finie $(E, \|\cdot\|)$. Montrer que cette suite converge dans $(E, \|\cdot\|)$ si et seulement si elle est bornée, elle n'admet qu'un nombre fini de valeurs d'adhérence et si $\lim\limits_{n\to+\infty} \|x_{n+1} - x_n\| = 0$.*

Solution : La condition nécessaire est une conséquence immédiate des définitions.

Soit $\{a_1, a_2, \ldots, a_n\}$ la suite finie des valeurs d'adhérence de la suite $(x_n)_{n\in\mathbb{N}}$. Pour $\epsilon = \frac{1}{4} \min\limits_{1\leq i\neq j\leq p} \|a_i - a_j\| > 0$, on peut trouver un entier k_ϵ tel que :

$$\forall k \geq k_\epsilon, \begin{cases} \|x_{k+1} - x_k\| < \epsilon, \\ x_k \in \bigcup\limits_{i=1}^{p} B(a_i, \epsilon), \end{cases}$$

où $B(a_i, \epsilon)$ désigne la boule fermée de centre a_i et de rayon ϵ. Si j est un entier tel que $x_{k_\epsilon} \in B(a_j, \epsilon)$, on a alors :

$$\forall k \geq k_\epsilon, \quad x_k \in B(a_j, \epsilon).$$

En effet le résultat est vrai pour k_ϵ. Supposons-le vrai pour $k \geq k_\epsilon$ et soit i tel que $x_{k+1} \in B(a_i, \epsilon)$. Si $i \neq j$, alors :

$$\|x_{k+1} - x_k\| \geq \|a_i - a_j\| - \|x_{k+1} - a_i\| - \|x_k - a_j\| \geq 4\epsilon - \epsilon - \epsilon = 2\epsilon,$$

ce qui est en contradiction avec $\|x_{k+1} - x_k\| < \epsilon$. On a donc $i = j$. On a donc ainsi montré que $\lim\limits_{k\to+\infty} x_k = a_j$.

Exercice 1.11 : *Pour tout réel $p \geq 1$, on note $E = \mathcal{C}^0([a,b], \mathbb{K})$ l'espace vectoriel des applications continues sur l'intervalle $[a,b]$ $(a < b)$ et à valeurs dans $\mathbb{K}$ et on munit cet espace de la norme :*

$$f \mapsto \|f\|_p = \sqrt[p]{\int_a^b |f(x)|^p \, \mathrm{d}x}.$$

Montrer que $\left(E, \|\cdot\|_p\right)$ n'est pas complet.

Solution : On désigne par $(f_n)_{n\geq 2}$ la suite de fonctions définies sur $[0,1]$ par :

$$f_n(x) = \begin{cases} 1 & \text{si } 0 \leq x < \dfrac{1}{2}, \\ n\left(\dfrac{1}{2} + \dfrac{1}{n} - x\right) & \text{si } \dfrac{1}{2} \leq x < \dfrac{1}{2} + \dfrac{1}{n}, \\ 0 & \text{si } \dfrac{1}{2} + \dfrac{1}{n} \leq x \leq 1. \end{cases}$$

Pour tout réel $p \geq 1$ et pour tout entier $m > n$, on a :

$$\begin{aligned} \|f_n - f_m\|_p^p &= (m-n)^p \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{m}} \left(x - \frac{1}{2}\right)^p \mathrm{d}x + \int_{\frac{1}{2}+\frac{1}{m}}^{\frac{1}{2}+\frac{1}{n}} n^p \left(\frac{1}{2} + \frac{1}{n} - x\right)^p \mathrm{d}x \\ &= \frac{1}{(p+1)\,m}\left(1 - \frac{n}{m}\right)^p + \frac{1}{(p+1)\,n}\left(1 - \frac{n}{m}\right)^{p+1} \\ &\leq \frac{1}{p+1}\left(\frac{1}{m} + \frac{1}{n}\right). \end{aligned}$$

On en déduit alors facilement que cette suite est de Cauchy dans $\left(E, \|\cdot\|_p\right)$.

D'autre part la suite de fonctions $(f_n)_{n\geq 2}$ converge simplement vers la fonction f définie sur $[0,1]$ par :

$$f(x) = \begin{cases} 1 & \text{si } 0 \leq x \leq \dfrac{1}{2}, \\ 0 & \text{si } \dfrac{1}{2} < x \leq 1. \end{cases}$$

Cette fonction n'appartient pas à E, mais étant continue par morceaux sur $[0,1]$ elle est quand même Riemann-intégrable. Si la suite $(f_n)_{n\geq 2}$ converge dans $\left(E, \|\cdot\|_p\right)$ vers une fonction g, on peut alors écrire :

$$\|f - g\|_p \leq \|f - f_n\|_p + \|f_n - g\|_p,$$

avec :

$$\|f - f_n\|_p^p = \int_{\frac{1}{2}}^{\frac{1}{2}+\frac{1}{n}} n^p \left(\frac{1}{2} + \frac{1}{n} - x\right)^p \mathrm{d}x$$
$$= \frac{1}{(p+1)\,n}.$$

Et en passant à la limite quand n tend vers l'infini on déduit que $\|f-g\|_p = 0$. Par continuité on déduit alors que $f = g$ sur $[0,1] - \left\{\frac{1}{2}\right\}$ avec f discontinue en $1/2$ et g continue en ce point, ce qui est impossible. On a donc ainsi montré que $\left(E, \|\cdot\|_p\right)$ n'est pas complet.

Exercice 1.12 : *En considérant la suite de fonctions $(f_n)_{n\in\mathbb{N}}$ définie sur $[a,b]$ par $f_n(x) = \cos(nx)$, montrer que la boule unité de $E = \left(\mathcal{C}^0([a,b],\mathbb{R}), \|\cdot\|_\infty\right)$ n'est pas compacte.*

Solution : Il est clair que pour tout entier n la fonction f_n est dans la boule unité de E. En remarquant que pour n assez grand on peut toujours trouver un entier k tel que $\frac{k\pi}{n} \in [a,b]$ (pour n assez grand on aura $n\frac{b-a}{\pi} > 1$ et il existe $k \in \left[n\frac{a}{\pi}, n\frac{b}{\pi}\right] \cap \mathbb{N}$) on déduit que $\|f_n\|_\infty = 1$ à partir d'un certain rang.

Il s'agit de montrer que la suite $(f_n)_{n\in\mathbb{N}}$ n'admet aucune sous-suite uniformément convergente (*i.e.* pour la norme $\|\cdot\|_\infty$) sur $[a,b]$.

Supposons que l'on puisse extraire une sous-suite $\left(f_{\varphi(n)}\right)_{n\in\mathbb{N}}$ qui converge uniformément sur $[a,b]$ vers une fonction f. La fonction f est alors continue sur $[a,b]$ et pour tout intervalle $[a,x] \subset [a,b]$ on a :

$$\int_a^x f(x)\,\mathrm{d}x = \lim_{n\to+\infty} \int_a^x f_{\varphi(n)}(t)\,\mathrm{d}t$$
$$= \lim_{n\to+\infty} \frac{1}{\varphi(n)}\left[\sin(\varphi(n)\,x) - \sin(\varphi(n)\,a)\right] = 0$$

et f est nécessairement la fonction nulle, ce qui est en contradiction avec $\left\|f_{\varphi(n)}\right\|_\infty = 1$ à partir d'un certain rang.

Exercice 1.13 : *Montrer que l'application $f \mapsto f(a)$ est continue de $\mathcal{C}^0([a,b],\mathbb{K})$ muni de $\|\cdot\|_\infty$ dans $(\mathbb{K}, |\cdot|)$, mais non continue sur $\mathcal{C}^0([a,b],\mathbb{K})$ muni de $\|\cdot\|_1$.*

Solution : La continuité de cette application linéaire, pour la norme $\|\cdot\|_\infty$, résulte de l'inégalité $|f(a)| \leq \|f\|_\infty$ vérifiée par toute fonction f dans $\mathcal{C}^0([a,b],\mathbb{K})$.

La suite de fonctions $(f_n)_{n\geq 1}$ définie sur $[0,1]$ par :

$$f_n(x) = \begin{cases} -nx+1 & \text{si } 0 \leq x \leq \dfrac{1}{n}, \\ 0 & \text{si } \dfrac{1}{n} \leq x \leq 1, \end{cases}$$

est telle que $\lim\limits_{n\to+\infty} \|f_n\|_1 = 0$ et $\lim\limits_{n\to+\infty} f_n(0) = 1$. Donc l'application $f \mapsto f(0)$ n'est pas continue pour $\|\cdot\|_1$.

Exercice 1.14 : *Montrer que l'application $f \mapsto f'(0)$ définie sur $\left(\mathcal{C}^1([0,1],\mathbb{K}),\|\cdot\|_\infty\right)$ et à valeurs dans $(\mathbb{K},|\cdot|)$ n'est pas continue.*

Solution : Pour tout entier $n \geq 1$ la fonction $f_n : x \mapsto \dfrac{nx}{1+nx}$ est dans la boule unité de $\left(\mathcal{C}^1([0,1],\mathbb{K}),\|\cdot\|_\infty\right)$ avec $f_n'(0) = n$. On en déduit que cette application linéaire n'est pas bornée sur la boule unité et donc qu'elle n'est pas continue.

Exercice 1.15 : *Soit C un compact d'un espace vectoriel normé $(E,\|\cdot\|)$ et $\varphi : C \to C$ telle que :*

$$\forall (x,y) \in C^2, \quad x \neq y \Rightarrow \|\varphi(x) - \varphi(y)\| < \|x - y\|.$$

1. Montrer que φ admet un unique point fixe dans C.

2. Ce point fixe est-il limite d'une suite d'approximations successives ?

3. Ce résultat est-il encore vrai si on suppose seulement C complet ?

Solution :

1. L'application $x \longmapsto \|\varphi(x) - x\|$ étant continue sur le compact C, il existe $\lambda \in C$ tel que :

$$\|\varphi(\lambda) - \lambda\| = \inf_{x\in K} \|\varphi(x) - x\|.$$

Si $\varphi(\lambda) \neq \lambda$ alors $\left\|\varphi(\lambda) - \varphi(\varphi(\lambda))\right\| < \|\lambda - \varphi(\lambda)\|$ avec $\varphi(\lambda) \in C$, ce qui est contradictoire avec la définition de λ. On a donc $\varphi(\lambda) = \lambda$.

Si $\mu \neq \lambda$ est un autre point fixe de φ, alors $\|\lambda - \mu\| = \|\varphi(\lambda) - \varphi(\mu)\| < \|\lambda - \mu\|$, ce qui est impossible. En conclusion, φ admet un unique point fixe dans C.

2. Soit $(x_n)_{n\in\mathbb{N}}$ une suite d'approximations successives définie par la donnée de $x_0 \in C$ et par la récurrence $x_{n+1} = \varphi(x_n)$. S'il existe $n_0 \in \mathbb{N}$ tel que $x_{n_0} = \lambda$, alors la suite est stationnaire. On suppose donc que $x_n \neq \lambda$ pour tout n. On a alors $0 < \|x_{n+1} - \lambda\| = \|\varphi(x_n) - \varphi(\lambda)\| < \|x_n - \lambda\|$ pour tout n. C'est-à-dire que la suite $(\|x_n - \lambda\|)_{n\in\mathbb{N}}$ est strictement décroissante minorée par 0, et qu'elle est donc convergente vers un réel $d \geq 0$. La suite $(x_n)_{n\in\mathbb{N}}$ étant dans le compact C, on peut en extraire une sous-suite $\left(x_{\psi(n)}\right)_{n\in\mathbb{N}}$ qui converge vers μ. On a alors :

$$\|\mu - \lambda\| = \lim_{n\to+\infty} \left\|x_{\psi(n)} - \lambda\right\| = d,$$
$$\|\varphi(\mu) - \lambda\| = \lim_{n\to+\infty} \left\|\varphi\left(x_{\varphi(n)}\right) - \lambda\right\| = \lim_{n\to+\infty} \left\|x_{\varphi(n)+1} - \lambda\right\| = d. \qquad (1.3)$$

Si $\mu \neq \lambda$, alors $\|\lambda - \varphi(\mu)\| = \|\varphi(\lambda) - \varphi(\mu)\| < \|\lambda - \mu\| = d$, ce qui est contradictoire avec (1.3), donc $\mu = \lambda$ et $d = 0$. C'est-à-dire que $\lim\limits_{n\to+\infty} x_n = \lambda$.

3. Le résultat du 1. est faux si on suppose seulement que C est complet, comme le montre l'exemple de la fonction $\varphi : x \longmapsto x + \dfrac{1}{x+1}$ sur $\mathbb{R}^+$ ($|\varphi(x) - \varphi(y)| < |x - y|$ avec les accroissements finis).

Exercice 1.16 : *Soient $(E, \|\cdot\|)$ un espace vectoriel normé complet et $\varphi : E \to E$ telle qu'il existe $\alpha, \beta \in]0,1[$ avec $\alpha + \beta < 1$ et :*

$$\forall (x,y) \in E^2, \quad \|\varphi(x) - \varphi(y)\| \leq \alpha \|\varphi(x) - x\| + \beta \|\varphi(y) - y\|.$$

Montrer que φ admet un unique point fixe dans E.

Solution : Soit $(x_n)_{n\in\mathbb{N}}$ la suite définie par $x_0 \in E$ et $x_{n+1} = \varphi(x_n)$. On a :

$$\forall n \in \mathbb{N}^*, \quad \|x_{n+1} - x_n\| \leq \alpha \|x_{n+1} - x_n\| + \beta \|x_n - x_{n-1}\|.$$

Soit :

$$\forall n \in \mathbb{N}^*, \quad \|x_{n+1} - x_n\| \leq \frac{\beta}{1-\alpha} \|x_n - x_{n-1}\|,$$

avec $\dfrac{\beta}{1-\alpha} \in]0,1[$. On en déduit alors que la suite $(x_n)_{n\in\mathbb{N}}$ est de Cauchy et qu'elle converge vers $x \in E$.

Puis avec :

$$\|x_{n+1} - \varphi(x)\| \leq \alpha \|x_{n+1} - x_n\| + \beta \|\varphi(x) - x\|$$

et en faisant tendre n vers l'infini, on déduit que :

$$0 \leq \|x - \varphi(x)\| \leq \beta \|\varphi(x) - x\|$$

et $\varphi(x) = x$. C'est-à-dire que x est point fixe de φ.

Si $y \in E$ est un autre point fixe de φ alors :

$$\|x - y\| = \|\varphi(x) - \varphi(y)\| \leq \alpha \|\varphi(x) - x\| + \beta \|\varphi(y) - y\| = 0$$

et $x = y$. Donc φ admet un unique point fixe dans E.

Exercice 1.17 : *Montrer que les normes $f \mapsto \|f\|_k$ ($k = \infty, 1, 2$) de l'exemple 1.3 (sur l'espace vectoriel E des applications continues sur l'intervalle $[a, b]$ ($a < b$) et à valeurs dans $\mathbb{K}$) ne sont pas équivalentes.*

Solution : On se place sur $[a, b] = [0, 1]$ et on considère la suite de fonctions $(f_n)_{n \geq 1}$ définie par :

$$f_n(x) = \begin{cases} -n^3 x + n & \text{si } x \in \left[0, \frac{1}{n^2}\right], \\ 0 & \text{si } x \in \left[\frac{1}{n^2}, 1\right]. \end{cases}$$

On a $\lim_{n \to +\infty} \|f_n\|_1 = 0$ et $\lim_{n \to +\infty} \|f_n\|_\infty = +\infty$. Donc les normes $\|\cdot\|_1$ et $\|\cdot\|_\infty$ ne sont pas équivalentes.

Soit $(f_n)_{n \geq 1}$ la suite de fonctions définie sur $[0, 1]$ par :

$$f_n(x) = \begin{cases} \sqrt{n} & \text{si } x \in \left[0, \frac{1}{n^{\frac{3}{2}}}\right], \\ 0 & \text{si } x \in \left[\frac{1}{n^{\frac{3}{2}}}, 1\right]. \end{cases}$$

On a $\lim_{n \to +\infty} \|f_n\|_2 = 0$ et $\lim_{n \to +\infty} \|f_n\|_\infty = +\infty$. Donc les normes $\|\cdot\|_2$ et $\|\cdot\|_\infty$ ne sont pas équivalentes.

Soit $(f_n)_{n \geq 1}$ la suite de fonctions définie sur $[0, 1]$ par :

$$f_n(x) = \begin{cases} -n^{\frac{5}{3}} x + n^{\frac{2}{3}} & \text{si } x \in \left[0, \frac{1}{n}\right], \\ 0 & \text{si } x \in \left[\frac{1}{n}, 1\right]. \end{cases}$$

On a $\lim_{n \to +\infty} \|f_n\|_1 = 0$ et $\lim_{n \to +\infty} \|f_n\|_2 = +\infty$. Donc les normes $\|\cdot\|_1$ et $\|\cdot\|_2$ ne sont pas équivalentes.

Exercice 1.18 : *Soit $(E, \|\cdot\|)$ un espace vectoriel normé. Pour toute partie non vide F de E, on note $d(x, F) = \inf_{y \in F} \|x - y\|$. Montrer que si F est fermée dans $(E, \|\cdot\|)$ alors pour tout $x \in E - F$ on a $d(x, F) > 0$.*

Solution : Pour $x \in E - F$, on note $\delta = d(x, F)$. Par définition de la borne inférieure, pour tout entier $n > 0$ on peut trouver un élément x_n de F tel que :

$$\delta \leq \|x - x_n\| < \delta + \frac{1}{n} \cdot$$

On a alors $\lim_{n \to +\infty} \|x - x_n\| = \delta$. Si $\delta = 0$ alors $x = \lim_{n \to +\infty} x_n \in F$, puisque F est fermée, ce qui contredit $x \in E - F$. On a donc $\delta > 0$.

Exercice 1.19 : *Soit $(E, \|\cdot\|)$ un $\mathbb{K}$–espace vectoriel normé. L'espace vectoriel produit $E \times E$ est muni de la norme :*

$$(x, y) \longmapsto \|(x, y)\| = \max(\|x\|, \|y\|).$$

Montrer qu'une application bilinéaire $u : E \times E \to E$ est continue si et seulement si il existe une constante $\lambda > 0$ telle que $\|u(x, y)\| \leq \lambda \|x\| \|y\|$ pour tous x, y dans E.

Solution : Si l'application u est bilinéaire et continue alors $u(0,0) = 0$ et u est continue en $(0,0)$. Il existe donc un réel $\eta > 0$ tel que :

$$(x, y) \in E \times E, \quad \|x\| \leq \eta, \quad \|y\| \leq \eta \quad \Rightarrow \quad \|u(x, y)\| \leq 1.$$

Pour x, y dans $E - \{0\}$, on a :

$$\left\| \frac{\eta}{\|x\|} x \right\| = \eta \text{ et } \left\| \frac{\eta}{\|y\|} y \right\| = \eta, \text{ donc } \left\| u\left(\frac{\eta}{\|x\|} x, \frac{\eta}{\|y\|} y \right) \right\| \leq 1,$$

et, avec la bilinéarité de u, on déduit que :

$$\forall (x, y) \in E \times E, \quad \|u(x, y)\| \leq \lambda \|x\| \|y\|, \tag{1.4}$$

où $\lambda = \dfrac{1}{\eta^2}$.

Réciproquement, supposons (1.4) vérifiée. Pour x_0, x, y_0 et y dans E, on a :

$$\begin{aligned}
\|u(x, y) - u(x_0, y_0)\| &= \|u(x - x_0, y) + u(x_0, y - y_0)\| \\
&\leq \|u(x - x_0, y)\| + \|u(x_0, y - y_0)\| \\
&\leq \lambda\big(\|x - x_0\| \|y\| + \|x_0\| \|y - y_0\|\big),
\end{aligned}$$

ce qui entraîne $\lim\limits_{(x,y) \to (x_0, y_0)} u(x, y) = u(x_0, y_0)$ et la continuité de u sur $E \times E$.

Exercice 1.20 : *Montrer qu'un sous-espace vectoriel strict de dimension finie d'un espace vectoriel normé est d'intérieur vide (l'intérieur d'une partie C d'un espace vectoriel normé est le plus grand ouvert contenu dans C).*

Solution : Soient $(E, \|\cdot\|)$ un $\mathbb{K}$–espace vectoriel normé et F un sous-espace vectoriel strict de E de dimension n. Si l'intérieur de F est non vide il existe alors un élément a de F et un réel $r > 0$ tels que :

$$\mathcal{B}(a, r) = \{x \in E \mid \|x - a\| \leq r\} \subset F.$$

La boule fermée de centre 0 et de rayon r est alors aussi contenue dans l'espace vectoriel F comme translatée de $\mathcal{B}(a, r)$. Le sous-espace vectoriel F, de dimension n, étant strictement contenu dans E, il existe un système libre $\{e_1, \ldots, e_{n+1}\}$ dans E constitué de $n+1$ vecteurs que l'on peut supposer tous de norme 1. Mais alors le système $\{re_1, \ldots, re_{n+1}\}$ est un système libre contenu dans $\mathcal{B}(0, r)$, donc dans F, ce qui est en contradiction avec F de dimension n. En conséquence, un sous-espace vectoriel strict de dimension finie d'un espace vectoriel normé est d'intérieur vide.

Exercice 1.21 : *Montrer qu'un compact dans un espace vectoriel normé de dimension infinie est d'intérieur vide.*

Solution : Soient $(E, \|\cdot\|)$ un $\mathbb{K}$-espace vectoriel normé et $\mathcal{C}$ un compact dans $(E, \|\cdot\|)$. Si l'intérieur de $\mathcal{C}$ est non vide il existe alors un élément a de $\mathcal{C}$ et un réel $r > 0$ tels que :

$$\mathcal{B}(a,r) = \{x \in E \mid \|x - a\| \leq r\} \subset \mathcal{C}.$$

La boule fermée de centre a et de rayon r est alors compacte comme fermé dans un compact. On déduit alors que la boule fermée de centre 0 et de rayon r est compacte comme image du compact $\mathcal{B}(a,r)$ par la translation de vecteur $-a$ (application continue), puis que la boule unité de $(E, \|\cdot\|)$ est compacte comme image du compact $\mathcal{B}(0,r)$ par l'homothétie de rapport $1/r$ (application continue). Mais on sait que la compacité de cette boule unité équivaut à dire que E est de dimension finie (théorème de Riesz). En conséquence un compact dans un espace vectoriel normé de dimension infinie est d'intérieur vide.

Exercice 1.22 : *Montrer qu'un hyperplan dans un espace vectoriel normé est dense ou fermé.*

Solution : Soient $(E, \|\cdot\|)$ un $\mathbb{K}$-espace vectoriel normé et H un hyperplan de E. L'adhérence $\overline{H}$ de H dans E est un sous-espace vectoriel de E qui contient H. On a alors deux possibilités : soit $\overline{H} = H$ et dans ce cas H est fermé, soit $\overline{H} = E$ et dans ce cas H est dense dans E.

Exercice 1.23 : *On note $E = \mathcal{C}^0([a,b], \mathbb{R})$ l'espace vectoriel des applications continues sur l'intervalle $[a,b]$ $(a < b)$ et à valeurs dans $\mathbb{R}$ et on munit cet espace de la norme :*

$$f \mapsto \|f\|_\infty = \sup_{x \in [a,b]} |f(x)|.$$

Soient f une fonction appartenant à $E - \{0\}$ et $\mathcal{B}(0, 2\|f\|)$ la boule fermée de centre 0 et de rayon $2\|f\|$. Montrer qu'il existe un polynôme P dans $\mathbb{R}_n[x] \cap \mathcal{B}(0, 2\|f\|)$ tel que :

$$\|f - P\|_\infty = \inf_{Q \in \mathbb{R}_n[x]} \|f - Q\|_\infty.$$

On dit que P est un polynôme de meilleure approximation de f dans $\mathbb{R}_n[x]$, pour la norme de la convergence uniforme.

Solution : L'ensemble $\mathcal{B}_{n,f} = \mathbb{R}_n[x] \cap \mathcal{B}(0, 2\|f\|)$ est la boule fermée de centre 0 et de rayon $2\|f\|$ dans l'espace vectoriel $\mathbb{R}_n[x]$ muni de la norme de la convergence uniforme. Cet espace vectoriel étant de dimension finie, égale à $n+1$, on sait que cette boule est compacte. L'application $Q \mapsto \|f - Q\|_\infty$ étant continue sur le compact $\mathcal{B}_{n,f}$ de $\mathbb{R}_n[x]$ est minorée et atteint sa borne inférieure, c'est-à-dire qu'il existe un polynôme P dans $\mathcal{B}_{n,f}$ tel que :

$$\delta = \inf_{Q \in \mathcal{B}_{n,f}} \|f - Q\|_\infty = \|f - P\|_\infty .$$

Pour tout polynôme Q dans $\mathbb{R}_n[x]$ on a soit $Q \in \mathcal{B}_{n,f}$ et alors $\|f - Q\|_\infty \geq \delta$, soit $Q \notin \mathcal{B}_{n,f}$ et alors :

$$\|f - Q\|_\infty \geq \|Q\|_\infty - \|f\|_\infty > 2\|f\|_\infty - \|f\|_\infty = \|f\|_\infty$$

et en remarquant que $0 \in \mathcal{B}_{n,f}$ on déduit que $\|f\|_\infty = \|f - 0\|_\infty \geq \delta$. En définitive on a bien $\|f - Q\|_\infty \geq \delta$ pour tout Q dans $\mathbb{R}_n[x]$ et donc :

$$\delta = \inf_{Q \in \mathbb{R}_n[x]} \|f - Q\|_\infty = \|f - P\|_\infty .$$

Remarque 1.9 : *Le polynôme P est uniquement déterminé par f (voir ref. [2], exercice 21.4).*

2 Polynômes minimal et caractéristique. Sous-espaces caractéristiques

On note $\mathbb{K}$ le corps des réels ou des complexes et on désigne par E un $\mathbb{K}$-espace vectoriel de dimension $n \geq 1$.

Pour toute partie non vide X de E on désigne par $\operatorname{Vect}(X)$ le sous-espace vectoriel de E engendré par X.

On note $\mathcal{L}(E)$ l'algèbre des endomorphismes de E et $\mathcal{M}_n(\mathbb{K})$ l'algèbre des matrices carrées d'ordre n à coefficients dans $\mathbb{K}$.

On note Id [resp. I_n] l'endomorphisme [resp. la matrice] identité.

Le choix d'une base de E permet de réaliser un isomorphisme d'algèbres de $\mathcal{L}(E)$ sur $\mathcal{M}_n(\mathbb{K})$.

Cet isomorphisme est réalisé de la façon suivante. Si $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ est une base de E, alors à tout endomorphisme u de E on associe la matrice $A = \left((a_{ij})\right)_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{K})$ définie par :

$$\forall j \in \{1, 2, \ldots, n\}, \quad u(e_j) = \sum_{i=1}^{n} a_{ij} e_i.$$

À toute matrice A dans $\mathcal{M}_n(\mathbb{K})$ est associé l'endomorphisme de $\mathbb{K}^n$, que nous noterons encore A :

$$\begin{aligned} A: \ & \mathbb{K}^n \to \mathbb{K}^n \\ & x \ \ \mapsto y = Ax. \end{aligned}$$

On désigne par $\mathbb{K}[X]$ l'algèbre des polynômes à coefficients dans $\mathbb{K}$.

Un polynôme non nul est dit unitaire si son coefficient dominant est égal à 1.

On rappelle que $\mathbb{K}[X]$ est un anneau euclidien, donc principal et factoriel. Un résultat qui nous sera utile est le théorème de Bézout qui dit que deux polynômes A et B sont premiers entre eux dans $\mathbb{K}[X]$ si et seulement si il existe deux polynômes U et V tels que $AU + BV = 1$.

Pour tout endomorphisme u de E, la sous-algèbre de $\mathcal{L}(E)$ engendrée par u est constituée des endomorphismes $v = P(u)$ où P est dans $\mathbb{K}[X]$. On note naturellement $\mathbb{K}[u]$ cette algèbre et il est facile de vérifier qu'elle est commutative. Précisément on a :

$$\forall (P, Q) \in \mathbb{K}[X]^2, \quad (PQ)(u) = P(u) \circ Q(u) = Q(u) \circ P(u).$$

1. Définitions et premières propriétés

L'espace vectoriel $\mathcal{L}(E)$ étant de dimension n^2, on déduit que pour tout endomorphisme u de E la famille $\left\{u^k \mid 0 \leq k \leq n^2\right\}$ est liée, ce qui se traduit en disant qu'il existe un polynôme P non nul tel que $P(u) = 0$. On déduit donc que l'ensemble :

$$I_u = \{P \in \mathbb{K}[X] \mid P(u) = 0\}$$

n'est pas réduit au polynôme nul. Cet ensemble étant le noyau du morphisme d'algèbres $P \mapsto P(u)$, c'est donc un idéal de $\mathbb{K}[X]$. L'anneau $\mathbb{K}[X]$ étant principal on peut donner la définition suivante.

Définition 2.1 : *Pour tout endomorphisme u de E, on appelle idéal annulateur de u l'idéal I_u, et polynôme minimal de u le générateur unitaire de cet idéal. On note π_u ce polynôme.*

On a donc :

$$I_u = \{P \in \mathbb{K}[X] \mid P(u) = 0\} = \mathbb{K}[X]\,\pi_u$$

et π_u est le polynôme unitaire de plus petit degré annulant u.

On définit de manière analogue le polynôme minimal d'une matrice $A \in \mathcal{M}_n(\mathbb{K})$.

Si $u \in \mathcal{L}(E)$ a pour matrice A dans une base de E, alors A et u ont même polynôme minimal.

Définition 2.2 : *Soit u un endomorphisme de E. On dit que λ dans $\mathbb{K}$ est une valeur propre de u s'il existe un vecteur non nul x dans E tel que $u(x) = \lambda x$.*

On dit alors que x est un vecteur propre de u associé à la valeur propre λ et que le sous-espace vectoriel de E $E_\lambda = \ker(u - \lambda Id)$ est le sous-espace propre associé à λ.

L'ensemble des valeurs propres de u est appelé le spectre de u et est noté $\mathrm{sp}(u)$.

Il est facile de vérifier que si λ est une valeur propre de l'endomorphisme u et si x est un vecteur propre associé, alors pour tout polynôme $P \in \mathbb{K}[X]$, on a $P(u)x = P(\lambda)x$. C'est-à-dire que x est un vecteur propre de $P(u)$ associé à la valeur propre $P(\lambda)$.

Réciproquement, pour $\mathbb{K} = \mathbb{C}$, toute valeur propre de $P(u)$ s'écrit $P(\lambda)$ où λ est une valeur propre de u (exercice 2.3).

Définition 2.3 : *Soit A une matrice d'ordre n à coefficients dans $\mathbb{K}$. On dit que λ dans $\mathbb{K}$ est valeur propre de A s'il existe un vecteur non nul x dans $\mathbb{K}^n$ tel que $Ax = \lambda x$.*

On dit alors que x est un vecteur propre de A associé à la valeur propre λ et que le sous-espace vectoriel de $\mathbb{K}^n$ $E_\lambda = \ker(A - \lambda I_n)$ est le sous-espace propre associé à λ.

L'ensemble des valeurs propres de A est appelé le spectre de A et est noté $\mathrm{sp}(A)$.

Un scalaire $\lambda \in \mathbb{K}$ est valeur propre de $u \in \mathcal{L}(E)$ [resp. $A \in \mathcal{M}_n(\mathbb{K})$] si et seulement si $u - \lambda Id$ [resp. $A - \lambda I_n$] est non inversible, ce qui équivaut à dire que le polynôme caractéristique de u [resp. A] $P_u(t) = \det(u - tId)$ [resp. $P_A(t) = \det(A - tI_n)$] est nul pour $t = \lambda$. On déduit alors que le spectre de u [resp. A] est une partie finie de $\mathbb{K}$ ayant au plus n éléments. Pour $\mathbb{K} = \mathbb{R}$ ce spectre peut être vide, mais pour $\mathbb{K} = \mathbb{C}$ le théorème de d'Alembert-Gauss nous dit qu'il est non vide.

Si $u \in \mathcal{L}(E)$ a pour matrice A dans une base de E, alors A et u ont même polynôme caractéristique et mêmes valeurs propres.

Lemme 2.1 : *Soit u un endomorphisme non nul de E. Si F est un sous-espace vectoriel de E stable par u, alors le polynôme caractéristique de la restriction de u à F divise celui de u.*

Démonstration : On désigne par $\mathcal{B}_1$ une base de F complétée en une base de E $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$. Dans cette base la matrice de u est $A = \begin{pmatrix} A_1 & A_2 \\ 0 & A_3 \end{pmatrix}$ où A_1 est la matrice, dans la base $\mathcal{B}_1$, de la restriction de u à F (F est stable par u) et le polynôme caractéristique de u s'écrit :

$$P_u(X) = \det(A_1 - XI_{n_1}) \det(A_3 - XI_{n_3}).$$

On en déduit alors que P_u est un multiple du polynôme caractéristique de la restriction de u à F. ∎

Théorème 2.1 : *Soient u un endomorphisme de E et $\lambda \in \mathbb{K}$ une valeur propre de u. Si λ a pour multiplicité α en tant que racine du polynôme caractéristique de u on a alors :*

$$1 \leq \dim\bigl(\ker(u - \lambda Id)\bigr) \leq \alpha.$$

Démonstration : Si λ est valeur propre de u alors le sous-espace vectoriel $E_\lambda = \ker(u - \lambda Id)$ n'est pas réduit au vecteur nul et sa dimension est supérieure ou égale à 1. Ce sous-espace vectoriel étant stable par u, le polynôme caractéristique P_λ de la restriction de u à E_λ divise le polynôme caractéristique P_u de u (lemme 2.1). En remarquant que $P_\lambda(X) = (\lambda - X)^\delta$ où δ est la dimension de E_λ, on déduit alors que :

$$P_u(X) = (\lambda - X)^\delta Q(X)$$

et la racine λ de P_u étant de multiplicité α, on a nécessairement $\delta \leq \alpha$. ∎

Théorème 2.2 : *Soit u un endomorphisme de E. Les valeurs propres de u sont les racines de son polynôme minimal.*

Démonstration : Si $\lambda \in \mathbb{K}$ est une valeur propre de u et x un vecteur propre (non nul) associé, de l'égalité :

$$0 = \pi_u(u)(x) = \pi_u(\lambda) x$$

avec $x \neq 0$ on déduit que $\pi_u(\lambda) = 0$, c'est-à-dire que λ est racine de π_u.

Réciproquement, si λ est racine de π_u alors π_u s'écrit $\pi_u(X) = (X - \lambda) Q(X)$ et, avec $\pi_u(u) = (u - \lambda Id) \circ Q(u) = 0$ et le caractère minimal de π_u, on déduit que $u - \lambda Id$ est non inversible, ce qui équivaut à dire que λ est une valeur propre de u. ∎

Définition 2.4 : *Soit u un endomorphisme de E. La multiplicité d'une valeur propre de u en tant que racine de son polynôme minimal est appelée l'indice de cette valeur propre.*

Définition 2.5 : *On dit qu'un endomorphisme u dans $\mathcal{L}(E)$ [resp. une matrice A dans $\mathcal{M}_n(\mathbb{K})$] est nilpotent [resp. nilpotente] s'il existe un entier p strictement positif tel que $u^{p-1} \neq 0$ et $u^p = 0$ [resp. $A^{p-1} \neq 0$ et $A^p = 0$]. On dit que p est l'ordre de nilpotence de u [resp. de A].*

Il est facile de vérifier que 0 est la seule valeur propre d'un endomorphisme nilpotent.

2. Localisation des valeurs propres

Le théorème qui suit donne un premier résultat qui permet de localiser les valeurs propres d'une matrice réelle ou complexe.

Si A est une matrice d'ordre n supérieur ou égal à 2 à coefficients réels ou complexes, on note :

$$L_i = \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \quad (1 \leq i \leq n), \quad L = \max_{1 \leq i \leq n} \{L_i + |a_{ii}|\},$$

$$C_j = \sum_{\substack{i=1 \\ i \neq j}}^{n} |a_{ij}| \quad (1 \leq j \leq n), \quad C = \max_{1 \leq j \leq n} \{C_j + |a_{jj}|\}.$$

Théorème 2.3 (Gerschgörin-Hadamard) : *Soient A dans $\mathcal{M}_n(\mathbb{K})$ et λ dans $\mathbb{C}$ une valeur propre de A. Il existe un indice $i \in \{1, 2, \ldots, n\}$ tel que :*

$$|\lambda - a_{ii}| \leq L_i.$$

Démonstration : Soient $\lambda \in \mathbb{C}$ une valeur propre de A et x un vecteur propre associé dans $\mathbb{C}^n$ avec $\|x\|_\infty = 1$. Pour $i \in \{1, 2, \ldots, n\}$ tel que $|x_i| = \|x\|_\infty$, on a :

$$\sum_{\substack{j=1 \\ j \neq i}}^{n} a_{ij} x_j = (\lambda - a_{ii}) x_i$$

et $|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| = L_i$. ∎

L'exercice 2.10 est une application du théorème de Gerschgörin-Hadamard au calcul des valeurs propres d'une matrice.

Corollaire 2.1 : *Pour toute valeur propre $\lambda \in \mathbb{C}$ de $A \in \mathcal{M}_n(\mathbb{K})$ on a :*

$$|\lambda| \leq \min\{L, C\}.$$

Démonstration : Le théorème de Gerschgörin-Hadamard nous dit que toute valeur propre de A est dans l'un des disques $|\lambda - a_{ii}| \leq L_i$. Ce même théorème, appliqué à la transposée de A qui admet les mêmes valeurs propres que A, nous dit que toute valeur propre de A est dans l'un des disques $|\lambda - a_{jj}| \leq C_j$.

On peut alors écrire que pour toute valeur propre λ de A on a :

$$|\lambda| \leq |\lambda - a_{ii}| + |a_{ii}| \leq L_i + |a_{ii}| \leq L.$$

De manière analogue, on a $|\lambda| \leq C$. On a donc bien $|\lambda| \leq \min\{L, C\}$, pour toute valeur propre λ de A. ■

Une généralisation du théorème de Gerschgörin-Hadamard est le théorème d'Ostrowski qui suit.

Lemme 2.2 : *Soit A dans $\mathcal{M}_n(\mathbb{K})$. S'il existe un réel $\alpha \in [0,1]$ tel que :*

$$\forall i \in \{1, 2, \ldots, n\}, \quad |a_{ii}| > L_i^{\alpha} C_i^{1-\alpha},$$

alors A est inversible.

Démonstration : Si A n'est pas inversible alors 0 est valeur propre et, avec le théorème de Gerschgörin-Hadamard, on déduit qu'il existe des indices i et j compris entre 1 et n tels que $|a_{ii}| \leq L_i$, $|a_{jj}| \leq C_j$. Ce qui démontre le résultat pour $\alpha = 0$ et $\alpha = 1$.

On suppose maintenant que $\alpha \in]0,1[$ et que A est non inversible.

On désigne par x un vecteur propre non nul associé à la valeur propre 0. Le vecteur x est alors solution non nulle du système linéaire :

$$\sum_{j=1}^{n} a_{ij} x_j = 0 \quad (1 \leq i \leq n)$$

et on a :

$$|a_{ii}| \, |x_i| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \, |x_j| \quad (1 \leq i \leq n).$$

Avec $L_i^{\alpha} C_i^{1-\alpha} < |a_{ii}|$, pour tout $i \in \{1, 2, \ldots, n\}$, on déduit que :

$$L_i^{\alpha} C_i^{1-\alpha} |x_i| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \, |x_j| \quad (1 \leq i \leq n),$$

l'inégalité étant stricte pour tous les indices i tels que $x_i \neq 0$. On utilise alors l'inégalité de Hölder (avec $p = \frac{1}{\alpha}$ et $q = \frac{1}{1-\alpha}$, exercice 1.2), ce qui donne pour $1 \leq i \leq n$:

$$\begin{aligned} L_i^{\alpha} C_i^{1-\alpha} |x_i| &\leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|^{\alpha} \left(|a_{ij}|^{1-\alpha} |x_j| \right) \\ &\leq \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \right)^{\alpha} \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| \, |x_j|^{\frac{1}{1-\alpha}} \right)^{1-\alpha}, \end{aligned}$$

soit :

$$L_i^{\alpha} C_i^{1-\alpha} |x_i| \leq L_i^{\alpha} \left(\sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| |x_j|^{\frac{1}{1-\alpha}} \right)^{1-\alpha} \qquad (1 \leq i \leq n).$$

Si $x_i \neq 0$, alors l'inégalité est stricte et $L_i > 0$. On déduit donc que :

$$C_i |x_i|^{\frac{1}{1-\alpha}} \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| |x_j|^{\frac{1}{1-\alpha}} \quad (1 \leq i \leq n),$$

l'inégalité étant stricte pour $x_i \neq 0$ et évidente pour $x_i = 0$.

En additionnant ces inégalités, on aboutit à :

$$S = \sum_{i=1}^{n} C_i |x_i|^{\frac{1}{1-\alpha}} < \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| |x_j|^{\frac{1}{1-\alpha}} = \sum_{j=1}^{n} C_j |x_j|^{\frac{1}{1-\alpha}} = S,$$

ce qui est impossible. La matrice A est donc nécessairement inversible. ∎

Théorème 2.4 (Ostrowski) : *Soit A dans $\mathcal{M}_n(\mathbb{K})$. Pour tout réel $\alpha \in [0,1]$ et toute valeur propre de A, $\lambda \in \mathbb{C}$, il existe $i \in \{1, 2, \dots, n\}$ tel que :*

$$|\lambda - a_{ii}| \leq L_i^{\alpha} C_i^{1-\alpha}.$$

Démonstration : Si λ est valeur propre de A alors $A - \lambda I_n$ est non inversible et, avec le lemme 2.2, on déduit que pour tout réel $\alpha \in [0,1]$ on peut trouver un indice $i \in \{1, 2, \dots, n\}$ tel que $|a_{ii} - \lambda| \leq L_i^{\alpha} C_i^{1-\alpha}$. ∎

Remarque 2.1 *: Pour $\alpha = 1$, on retrouve le théorème de Gerschgörin-Hadamard.*

Corollaire 2.2 : *Soit A dans $\mathcal{M}_n(\mathbb{K})$. Pour toute valeur propre de A, $\lambda \in \mathbb{C}$, il existe $i \in \{1, 2, \dots, n\}$ tel que :*

$$|\lambda|^2 \leq (L_i + |a_{ii}|)(C_i + |a_{ii}|).$$

Démonstration : En prenant $\alpha = \dfrac{1}{2}$ dans le théorème d'Ostrowski, on peut trouver $i \in \{1, 2, ..., n\}$ tel que $|a_{ii} - \lambda| \leq \sqrt{L_i C_i}$. On déduit alors que :

$$|\lambda| \leq |a_{ii}| + \sqrt{L_i C_i} \leq \sqrt{(|a_{ii}| + L_i)(|a_{ii}| + C_i)},$$

la dernière inégalité résultant de $2\sqrt{L_i C_i} \leq C_i + L_i$. ∎

Définition 2.6 : *Une matrice réelle ou complexe A est dite à diagonale strictement dominante si :*

$$\forall i \in \{1, 2, \dots, n\}, \quad |a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}|.$$

Les matrices à diagonale strictement dominante se rencontrent dans de nombreux problèmes, par exemple dans le problème de l'interpolation par des fonctions splines cubiques ou dans les problèmes de résolution d'équations aux dérivées partielles par des méthodes de discrétisation par différences finies.

Corollaire 2.3 : *Une matrice réelle ou complexe à diagonale strictement dominante a toutes ses valeurs propres non nulles dans $\mathbb{C}$. En conséquence elle est inversible.*

Démonstration : Soient A une matrice à diagonale strictement dominante et λ une valeur propre de A dans $\mathbb{C}$. Le théorème de Gerschgörin-Hadamard nous dit qu'il existe un indice $i \in \{1, 2, \ldots, n\}$ tel que :

$$|\lambda - a_{ii}| \leq \sum_{\substack{j=1 \\ j \neq i}}^{n} |a_{ij}| .$$

On ne peut donc avoir $\lambda = 0$ si A est à diagonale strictement dominante. ∎

3. Le théorème de Cayley-Hamilton

Lemme 2.3 : *Soient u un endomorphisme non nul de E et x un vecteur non nul dans E. On note E_x le sous-espace vectoriel de E engendré par $\left\{u^k(x) \mid k \in \mathbb{N}\right\}$ (sous-espace cyclique engendré par x) et p_x le plus petit entier strictement positif tel que le système $\mathcal{B}_x = \left\{u^k(x) \mid 0 \leq k \leq p_x - 1\right\}$ soit libre. Le système $\mathcal{B}_x$ est alors une base de E_x et en notant :*

$$\pi_x(t) = t^{p_x} - \sum_{k=0}^{p_x-1} a_k t^k$$

le polynôme défini par $u^{p_x}(x) = \sum_{k=0}^{p_x-1} a_k u^k(x)$, π_x est le polynôme minimal et $(-1)^{p_x}\pi_x$ le polynôme caractéristique de la restriction de u à E_x.

Démonstration : On a $F_x = \operatorname{Vect}(\mathcal{B}_x) \subset E_x$.

Par définition de l'entier p_x, le système $\mathcal{B}_x$ est une base de F_x et $u^{p_x}(x) \in F_x$. On déduit alors, par récurrence sur k, que $u^{p_x+k}(x) \in F_x$ pour tout entier naturel k. On a donc $F_x = E_x$.

On a $\pi_x(u)(x) = 0$ et, $\mathbb{K}[u]$ étant commutatif :

$$\pi_x(u)\left(u^k(x)\right) = 0$$

pour tout entier k. Donc $\pi_x\left(u_{|E_x}\right) = 0$.

Si $Q \in \mathbb{K}_{p_x-1}[t] - \{0\}$ annule u, alors $Q(u)(x) = 0$ et le système $\mathcal{B}_x$ est lié, ce qui contredit la définition de p_x. En conclusion, π_x est le polynôme minimal de $u_{|E_x}$.

Ensuite, en écrivant que la matrice de $u_{|E_x}$ dans la base $\mathcal{B}_x$ est :

$$A_x = \begin{pmatrix} 0 & \dots & 0 & a_0 \\ 1 & \ddots & \vdots & a_1 \\ \vdots & \ddots & 0 & \vdots \\ 0 & \dots & 1 & a_{p_x-1} \end{pmatrix},$$

on déduit que $(-1)^{p_x}\pi_x$ est le polynôme caractéristique de $u_{|E_x}$. En effet, en notant $P_{A_x} = P_{(a_0,\dots,a_{p_x-1})}$ ce polynôme caractéristique et en le développant par rapport à la première ligne, on a :

$$P_{(a_0,\dots,a_{p_x-1})}(t) = -t\,P_{(a_1,\dots,a_{p_x-1})}(t) + (-1)^{p_x+1}a_0$$

et par récurrence $P_{A_x}(t) = (-1)^{p_x}\left(t^{p_x} - \sum\limits_{k=0}^{p_x-1} a_k t^k\right)$. ∎

Théorème 2.5 (Cayley-Hamilton) : *Soit u un endomorphisme de E de polynôme caractéristique P_u. On a alors $P_u(u) = 0$.*

Démonstration : Pour tout $x \in E$ le sous-espace cyclique E_x étant stable par u, le polynôme caractéristique π_x de $u_{|E_x}$ divise celui de u. C'est-à-dire que $P_u = Q\,\pi_x$ et $P_u(u)(x) = Q(u) \circ \pi_x(u)(x) = 0$.

On a donc ainsi montré que pour tout $x \in E$, on a $P_u(u)(x) = 0$ et donc que $P_u(u) = 0$. ∎

Remarque 2.2 : *La démonstration précédente du théorème de Cayley-Hamilton est en fait valable pour tout corps commutatif $\mathbb{K}$.*

Corollaire 2.4 : *Soit u un endomorphisme de E de polynôme caractéristique P_u et de polynôme minimal π_u.*

Le polynôme π_u divise le polynôme P_u. Il est donc de degré inférieur ou égal à n.

Démonstration : Le polynôme minimal divisant tout polynôme annulateur de u, on déduit du théorème de Cayley-Hamilton que π_u divise le polynôme caractéristique P_u. ∎

Remarque 2.3 : *On retrouve le fait que les racines de π_u sont valeurs propres de u.*

Remarque 2.4 : *Le corps des nombres complexes étant algébriquement clos (théorème de d'Alembert-Gauss), pour $\mathbb{K} = \mathbb{C}$ le polynôme caractéristique de l'endomorphisme u s'écrit :*

$$P_u(X) = (-1)^n \prod_{k=1}^{p} (X - \lambda_k)^{\alpha_k},$$

avec $\alpha_k \in \mathbb{N} - \{0\}$ et les λ_k deux à deux distinctes (tout polynôme à coefficients complexes est scindé sur $\mathbb{C}$). Le polynôme minimal π_u étant un diviseur de P_u est également scindé sur $\mathbb{C}$ et il s'écrit :

$$\pi_u(X) = \prod_{k=1}^{p} (X - \lambda_k)^{\beta_k}$$

avec $1 \leq \beta_k \leq \alpha_k$.

***Remarque 2.5** : Dans le cas où l'endomorphisme u est inversible, le théorème de Cayley-Hamilton nous donne un moyen de calculer l'inverse de u si on connaît son polynôme caractéristique P_u.*

En effet, l'égalité $P_u(u) = 0$ avec $P_u(t) = \sum_{k=0}^{n} a_k t^k$ donne :

$$u^{-1} = -\frac{1}{a_0}\sum_{k=1}^{n} a_k u^{k-1} = -\frac{1}{\det(u)}\sum_{k=1}^{n} a_k u^{k-1}.$$

On peut aussi remarquer que l'inverse de u est un polynôme en u.

***Remarque 2.6** : Le théorème de Cayley-Hamilton permet également de calculer A^p pour tout entier p supérieur ou égal à n en fonction de $I_n, A, \ldots, A^{n-1}$. En effet pour $p = n$, de $P_A(A) = \sum_{k=0}^{n} a_k A^k = 0$ avec $a_n = (-1)^n$ on déduit que $A^n = (-1)^{n+1} \sum_{k=0}^{n-1} a_k A^k$ et pour $p > n$ la division euclidienne de X^p par P_A, $X^p = QP_A + R$ avec $R = 0$ ou $R \neq 0$ et $\deg(R) < n$, donne $A^p = R(A)$.*

4. Méthodes de calcul du polynôme caractéristique

On décrit tout d'abord la méthode de Souriau qui donne un algorithme de calcul du polynôme caractéristique d'une matrice réelle ou complexe. On peut ensuite utiliser, pour les matrices de taille raisonnable, une méthode numérique de résolution d'une équation polynomiale pour en déduire des valeurs approchées des valeurs propres.

Si P_A désigne le polynôme caractéristique de la matrice A, on note :

$$(-1)^n P_A(\lambda) = \det(\lambda I_n - A) = \lambda^n + p_1\lambda^{n-1} + \cdots + p_{n-1}\lambda + p_n.$$

Lemme 2.4 (formules de Newton) : *Soit A une matrice réelle ou complexe de valeurs propres complexes distinctes ou confondues $\lambda_1, \ldots, \lambda_n$. En notant pour tout entier naturel k :*

$$S_k = \mathrm{Tr}\left(A^k\right) = \sum_{i=1}^{n} \lambda_i^k$$

et en utilisant les notations qui précèdent, on a :

$$S_k + p_1 S_{k-1} + \cdots + p_{k-1} S_1 + p_k k = 0 \quad (1 \leq k \leq n).$$

Démonstration : Voir l'exercice 2.14. ∎

***Remarque 2.7** : En utilisant le fait que pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$ on a $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$, on peut déduire des formules de Newton que AB et BA ont même polynôme caractéristique (exercice 2.15).*

On définit les suites de matrices $(A_k)_{1\leq k\leq n}$, $(B_k)_{1\leq k\leq n}$ et la suite de scalaires $(\alpha_k)_{1\leq k\leq n}$ de la manière suivante :

$$\begin{cases} A_1 = A, & \alpha_1 = -\operatorname{Tr}(A_1), \quad B_1 = A_1 + \alpha_1 I_n, \\ A_k = B_{k-1}A, & \alpha_k = -\dfrac{\operatorname{Tr}(A_k)}{k}, \quad B_k = A_k + \alpha_k I_n \ (2 \leq k \leq n). \end{cases}$$

Le résultat qui suit permet alors de calculer très simplement le polynôme caractéristique de A ainsi que l'inverse de A dans le cas où cette matrice est inversible.

Théorème 2.6 : *Avec les notations qui précèdent, les coefficients $(\alpha_k)_{1\leq k\leq n}$ sont les coefficients du polynôme caractéristique de A. Si de plus A est inversible, alors son inverse est $A^{-1} = -\dfrac{1}{\alpha_n}B_{n-1}$.*

Démonstration : On procède par récurrence sur k.

Pour $k = 1$, on a $\alpha_1 = -\operatorname{Tr}(A) = p_1$.

Supposons que $\alpha_j = p_j$ pour $j = 1, \ldots, k-1$, avec $k > 1$. Par définition, on a :

$$A_k = B_{k-1}A = A_{k-1}A + \alpha_{k-1}A$$

et par récurrence il vient :

$$A_k = A^k + \alpha_1 A^{k-1} + \cdots + \alpha_{k-1}A, \tag{2.1}$$

soit

$$\begin{cases} A_k = A^k + p_1 A^{k-1} + \cdots + p_{k-1}A, \\ \operatorname{Tr}(A_k) = \operatorname{Tr}\left(A^k\right) + p_1 \operatorname{Tr}\left(A^{k-1}\right) + \cdots + p_{k-1}\operatorname{Tr}(A). \end{cases}$$

On en déduit alors que :

$$\alpha_k = -\frac{1}{k}\left(S_k + p_1 S_{k-1} + \cdots + p_{k-1}S_1\right).$$

Avec les formules de Newton, cela s'écrit $\alpha_k = p_k$.

En prenant $k = n$ dans (2.1) on a :

$$A_n = A^n + \alpha_1 A^{n-1} + \cdots + \alpha_{n-1}A$$

et $B_n = A_n + \alpha_n I_n = (-1)^n P_A(A) = 0$ avec le théorème de Cayley-Hamilton. On déduit donc que l'inverse de A est $A^{-1} = \dfrac{-1}{\alpha_n}B_{n-1}$. ∎

La méthode de Krylov donne un autre algorithme de calcul du polynôme caractéristique et de l'inverse.

On garde les notations qui précèdent, et on note, pour tout vecteur u de $\mathbb{R}^n$ et pour tout entier naturel k, $u_k = A^k u$. Avec le théorème de Cayley-Hamilton on déduit que, quel que soit le vecteur u, les coefficients p_i sont solutions du système linéaire :

$$p_1 u_{n-1} + p_2 u_{n-2} + \cdots + p_n u_0 = -u_n.$$

En prenant u de manière aléatoire, on a toutes les chances que le système ci-dessus soit non dégénéré, de sorte que les coefficients du polynôme caractéristique s'obtiennent comme solution d'un système linéaire.

Toujours avec le théorème de Cayley-Hamilton, on déduit que l'inverse de A est donné par :

$$A^{-1} = -\frac{1}{p_n}\left(A^{n-1} + p_1 A^{n-2} + \cdots + p_{n-1} I_n\right).$$

La méthode de Leverrier donne un autre algorithme de calcul du polynôme caractéristique et de l'inverse.

En utilisant les formules de Newton, les coefficients p_k apparaissent comme solutions d'un système triangulaire inférieur, la matrice de ce système se calculant facilement avec les $S_k = \operatorname{Tr}\left(A^k\right)$.

5. Le théorème de décomposition des noyaux

Théorème 2.7 : *Soient p un entier supérieur ou égal à 2 et $P = \prod_{k=1}^{p} P_k$ dans $\mathbb{K}[X]$, les polynômes P_k étant deux à deux premiers entre eux. Pour tout endomorphisme u de E on a :*

$$\ker\big(P(u)\big) = \bigoplus_{k=1}^{p} \ker\big(P_k(u)\big).$$

Démonstration : On procède par récurrence sur $p \geq 2$.

Pour $p = 2$, les polynômes P_1 et P_2 sont premiers entre eux et le théorème de Bézout nous dit qu'il existe deux polynômes A et B tels que $AP_1 + BP_2 = 1$. On a alors :

$$A(u) \circ P_1(u) + B(u) \circ P_2(u) = Id$$

et pour tout $x \in \ker\big(P(u)\big)$, on a :

$$x = \big(A(u) \circ P_1(u)\big)(x) + \big(B(u) \circ P_2(u)\big)(x), \tag{2.2}$$

avec :

$$\begin{aligned} x_2 &= \big(A(u) \circ P_1(u)\big)(x) \in \ker\big(P_2(u)\big), \\ x_1 &= \big(B(u) \circ P_2(u)\big)(x) \in \ker\big(P_1(u)\big), \end{aligned}$$

du fait de la commutativité de $\mathbb{K}[u]$. On a donc ainsi montré que :

$$\ker(P(u)) = \ker(P_1(u)) + \ker(P_2(u)),$$

l'inclusion $\ker(P_1(u)) + \ker(P_2(u)) \subset \ker(P(u))$ étant évidente.

De l'égalité (2.2) on déduit immédiatement que :

$$\ker(P_1(u)) \cap \ker(P_2(u)) = \{0\}.$$

On a donc bien la somme directe :

$$\ker(P(u)) = \ker(P_1(u)) \oplus \ker(P_2(u)).$$

On conclut par récurrence sur p en considérant que si P_1 est premier avec $P_2, \ldots, P_p$ alors il est premier avec leur produit. ■

Remarque 2.8 : *Dans le cas où $P(u) = 0$, on a $E = \bigoplus_{k=1}^{p} \ker(P_k(u))$.*

6. Sous-espaces caractéristiques

Soit u dans $\mathcal{L}(E)$ tel que son polynôme caractéristique s'écrive :

$$P_u(X) = (-1)^n \prod_{k=1}^{p} (X - \lambda_k)^{\alpha_k},$$

avec $\alpha_k \in \mathbb{N} - \{0\}$ et les λ_k deux à deux distinctes. Le polynôme minimal de u s'écrit alors :

$$\pi_u(X) = \prod_{k=1}^{p} (X - \lambda_k)^{\beta_k}$$

avec $1 \leq \beta_k \leq \alpha_k$.

Une telle écriture de P_u est assurée pour $\mathbb{K} = \mathbb{C}$, mais pas nécessairement pour $\mathbb{K} = \mathbb{R}$.

Définition 2.7 : *Avec les notations qui précèdent, on appelle sous-espaces caractéristiques de u les sous-espaces vectoriels $N_k = \ker(u - \lambda_k Id)^{\alpha_k}$ pour tout $k \in \{1, 2, \ldots, p\}$.*

Théorème 2.8 : *Avec les notations qui précèdent, on a les résultats suivants :*

(i) $E = \bigoplus_{k=1}^{p} N_k$.

Pour tout $k \in \{1, 2, \ldots, p\}$ on a :

(ii) $N_k = \ker(u - \lambda_k Id)^{\beta_k}$;

(iii) *λ_k est la seule valeur propre de la restriction de u à N_k ;*

(iv) $\dim(N_k) = \alpha_k$;

(v) *la restriction de $u - \lambda_k Id$ à N_k est nilpotente d'indice β_k.*

Démonstration :

(i) De $P_u(u) = 0$ et du théorème de décomposition des noyaux on déduit que $E = \bigoplus_{k=1}^{p} N_k$.

(ii) On pose, pour tout $k \in \{1, 2, \ldots, p\}$, $M_k = \ker(u - \lambda_k Id)^{\beta_k}$ et on a :

$$M_k \subset N_k, \quad E = \bigoplus_{k=1}^{p} N_k = \bigoplus_{k=1}^{p} M_k,$$

donc :

$$1 \leq \dim(M_k) \leq \dim(N_k),$$

$$n = \sum_{k=1}^{p} \dim(N_k) = \sum_{k=1}^{p} \dim(M_k),$$

et nécessairement $M_k = N_k$.

(iii) Soit λ une valeur propre de $u_{|N_k}$. C'est aussi une valeur propre de u. Il existe donc $j \in \{1, 2, \ldots, p\}$ tel que $\lambda = \lambda_j$ et un vecteur $x \in N_k - \{0\}$ tel que $(u - \lambda_j Id)(x) = 0$. On a alors $x \in N_k \cap N_j - \{0\}$ et nécessairement $j = k$ ($N_k \cap N_j = \{0\}$, si $j \neq k$).

(iv) Soit $d_k = \dim(N_k)$, pour $k \in \{1, 2, \ldots, p\}$. De ce qui précède on déduit que le polynôme caractéristique de $u_{|N_k} \in \mathcal{L}(N_k)$ (N_k est stable par u) est $P_k(X) = (\lambda_k - X)^{d_k}$. De $E = \bigoplus_{k=1}^{p} N_k$, les N_k étant stables par u, on déduit que $P_u = \prod_{k=1}^{p} P_k$ et $d_k = \alpha_k$.

(v) Pour tout $k \in \{1, 2, \ldots, p\}$, on note :

$$v_k = (u - \lambda_k Id)_{|N_k} \in \mathcal{L}(N_k)$$

et on a $v_k^{\beta_k} = 0$ $\left(N_k = \ker(u - \lambda_k Id)^{\beta_k}\right)$. Si, pour un $k \in \{1, 2, \ldots, p\}$, $v_k^{\beta_k - 1} = 0$, alors le polynôme $(X - \lambda_k)^{\beta_k - 1} \prod_{\substack{j=1 \\ j \neq k}}^{p} (X - \lambda_j)^{\beta_j}$ annule u puisqu'il annule tous les $u_{|N_j}$ et $E = \bigoplus_{k=1}^{p} N_k$, ce qui contredit le fait que $\prod_{k=1}^{p} (X - \lambda_k)^{\beta_k}$ est le polynôme minimal de u. On a donc $v_k^{\beta_k - 1} \neq 0$ et v_k est nilpotent d'indice β_k. ■

Du théorème précédent on va déduire que tout endomorphisme d de E dont le polynôme caractéristique est scindé sur $\mathbb{K}$ se décompose de manière unique comme somme d'un endomorphisme diagonalisable d (un endomorphisme est diagonalisable s'il existe une base de E dans laquelle sa matrice est diagonale) et d'un endomorphisme nilpotent v avec d et v qui commutent.

Lemme 2.5 : *Avec les notations qui précèdent, pour tout k compris entre 1 et p, le projecteur π_k de E sur N_k parallèlement à $\bigoplus_{\substack{j=1 \\ j\neq k}}^{p} N_j$ est un polynôme en u.*

Démonstration : Les polynômes $P_k(X) = \prod_{\substack{j=1 \\ j\neq k}}^{p} (X-\lambda_j)^{\beta_j}$, $(1 \leq k \leq p)$, étant premiers entre eux dans leur ensemble, on peut donc trouver des polynômes $Q_1, Q_2, \ldots, Q_p$ tels que $\sum_{k=1}^{p} Q_k P_k = 1$ (théorème de Bézout). On a alors :

$$Id = \sum_{k=1}^{p} Q_k(u) \circ P_k(u)$$

et tout x de E s'écrit $x = \sum_{k=1}^{p} \big(Q_k(u) \circ P_k(u)\big)(x)$ avec :

$$\big(Q_k(u) \circ P_k(u)\big)(x) \in Ker\,(u - \lambda_k Id)^{\beta_k} = N_k.$$

C'est-à-dire que $\big(Q_k(u) \circ P_k(u)\big)(x)$ est la projection de x sur N_k parallèlement à $\bigoplus_{\substack{j=1 \\ j\neq k}}^{p} N_j$. On a donc :

$$\pi_k = Q_k(u) \circ P_k(u) \in \mathbb{K}\,[u]\,.$$ ■

Les projecteurs π_k de E sur N_k parallèlement à $\bigoplus_{\substack{j=1 \\ j\neq k}}^{p} N_j$ sont les projecteurs spectraux de u.

Théorème 2.9 (Dunford-Schwarz) : *Soit u un endomorphisme de E dont le polynôme caractéristique est scindé sur $\mathbb{K}$. Il existe un unique couple (d, v) d'endomorphismes de E tel que d soit diagonalisable, v soit nilpotent, d et v commutent et $u = d + v$.*

Démonstration : Sur chaque sous-espace vectoriel N_k, on a vu que l'endomorphisme $v_k = (u - \lambda_k Id)_{|N_k}$ est nilpotent et, en notant $d_k = \lambda_k Id_{|N_k}$, on a $u_{|N_k} = d_k + v_k$ avec d_k diagonalisable, v_k nilpotent et $d_k v_k = v_k d_k$.

On définit alors les endomorphismes d et v par $d = \sum_{k=1}^{p} \lambda_k \pi_k$ et $v = u - d$.

L'endomorphisme d est diagonalisable ($d = d_k$ sur N_k), l'endomorphisme v est nilpotent ($v = v_k$ sur N_k), d et v commutent puisqu'ils sont dans l'algèbre commutative $\mathbb{K}\,[u]$ et $u = d + v$.

Il reste à montrer l'unicité d'un tel couple (d, v) .

Soit (d', v') un autre couple d'endomorphismes vérifiant les mêmes conditions que (d, v). Comme $u = d' + v'$ et d' et v' commutent, ils commutent avec u donc avec d et v qui sont des polynômes en u. On a alors $d - d' = v' - v$, avec $d - d'$ diagonalisable comme somme de deux endomorphismes diagonalisables qui commutent (exercice 3.3) et $v' - v$ nilpotent comme somme de deux endomorphismes nilpotents qui commutent. Et nécessairement $d - d' = v' - v = 0$. D'où l'unicité de la décomposition. ■

Pratiquement, la décomposition de Dunford-Schwarz d'un endomorphisme de E dont le polynôme caractéristique est scindé sur $\mathbb{K}$ passe par le calcul des projecteurs spectraux π_k. Pour ce faire il suffit de disposer d'un polynôme annulateur de u de la forme :

$$P(X) = \prod_{k=1}^{p} (X - \lambda_k)^{m_k},$$

avec $m_k \geq \beta_k$ pour tout k compris entre 1 et p où les λ_k sont les valeurs propres de u et β_k est la multiplicité de λ_k comme racine du polynôme minimal (on peut prendre pour polynôme P le polynôme caractéristique, ou mieux le polynôme minimal, de u).

À partir de la décomposition en éléments simples :

$$\frac{1}{P(X)} = \sum_{k=1}^{p} \sum_{i=1}^{m_k} \frac{\alpha_{ik}}{(X - \lambda_k)^i},$$

en posant, pour tout k compris entre 1 et p :

$$\begin{cases} Q_k(X) = \sum\limits_{i=1}^{m_k} \alpha_{ik} (X - \lambda_k)^{m_k - i}, \\ P_k(X) = \prod\limits_{\substack{j=1 \\ j \neq k}}^{p} (X - \lambda_j)^{m_j}, \end{cases}$$

on obtient :

$$\frac{1}{P(X)} = \sum_{k=1}^{p} \frac{Q_k(X)}{(X - \lambda_k)^{m_k}}$$

et la décomposition de Bézout :

$$1 = \sum_{k=1}^{p} Q_k P_k$$

qui permet d'obtenir les projecteurs spectraux :

$$\pi_k = (Q_k P_k)(u).$$

On a alors $u = d + v$ avec $d = \sum\limits_{k=1}^{p} \lambda_k \pi_k$ et $v = u - d$.

La décomposition de Dunford-Schwarz d'un endomorphisme u de E dont le polynôme caractéristique est scindé sur $\mathbb{K}$ permet le calcul de ses puissances successives. En effet comme d et v commutent, on peut utiliser la formule du binôme de Newton pour écrire :

$$\forall r \geq 1, \quad u^r = (d + v)^r = \sum_{k=0}^{r} C_r^k d^k \circ v^{r-k}.$$

Le calcul des puissances successives de l'endomorphisme d peut se faire dans une base de diagonalisation ou en utilisant les propriétés des projecteurs pour écrire que :

$$\forall r \geq 1, \quad d^r = \left(\sum_{k=1}^{p} \lambda_k \pi_k\right)^r = \sum_{k=1}^{p} \lambda_k^r \pi_k$$

et le calcul des puissances successives de l'endomorphisme nilpotent v s'arrête à v^{q-1} où q est son indice de nilpotence.

On peut aussi calculer v^r avec :

$$\forall r \geq 1, \quad v^r = \left(\sum_{k=1}^{p} (u - \lambda_k Id)\, \pi_k\right)^r = \sum_{k=1}^{p} (u - \lambda_k Id)^r\, \pi_k.$$

Avec l'exercice 2.16 on a un exemple de tels calculs.

7. Exercices

On note toujours $\mathbb{K}$ le corps des réels ou des complexes et on désigne par E un $\mathbb{K}$-espace vectoriel de dimension $n \geq 1$.

Exercice 2.1 : *Soit P une matrice inversible de $\mathcal{M}_n(\mathbb{C})$. On définit les matrices réelles R et J par $R = \mathrm{Re}(P)$ et $J = \mathrm{Im}(P)$. Montrer qu'il existe un réel λ tel que la matrice $R + \lambda J$ soit inversible. En déduire que si A, B dans $\mathcal{M}_n(\mathbb{R})$ sont semblables dans $\mathcal{M}_n(\mathbb{C})$ alors elles sont semblables dans $\mathcal{M}_n(\mathbb{R})$.*

Solution : Si le polynôme $\varphi(X) = \det(R + XJ)$ s'annule pour toute valeur réelle il est alors identiquement nul et $\varphi(i) = \det(R + iJ) = \det(P) = 0$, ce qui contredit P inversible. Il existe donc des réels λ tels que $\varphi(\lambda) = \det(R + \lambda J) \neq 0$.

Si A, $B \in \mathcal{M}_n(\mathbb{R})$ sont semblables dans $\mathcal{M}_n(\mathbb{C})$ il existe alors une matrice P inversible dans $\mathcal{M}_n(\mathbb{C})$ telle que $A = P^{-1}BP$. On a alors, en notant R la partie réelle de P et J sa partie imaginaire :

$$(R + iJ)\, A = B\, (R + iJ)$$

et, en identifiant parties réelles et parties imaginaires, $RA = BR$, $JA = BJ$. Pour tout réel λ tel que $R + \lambda J$ soit inversible on a alors $(R + \lambda J)\, A = B\, (R + \lambda J)$, ce qui prouve que les matrices A et B sont semblables dans $\mathcal{M}_n(\mathbb{R})$.

Exercice 2.2 : *Montrer que pour tout endomorphisme u de E il existe un sous-espace vectoriel de E de dimension 1 ou 2 stable par u.*

Solution : Le polynôme minimal π_u se décompose dans l'anneau factoriel $\mathbb{K}[X]$ en produit de facteurs irréductibles, ces facteurs étant de degré 1 ou 2 ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). Si π_u admet un facteur irréductible de degré 1, ce dernier est de la forme $X - \lambda$ avec λ dans $\mathbb{K}$ et λ est une valeur propre de u. Il est alors facile de vérifier que pour tout vecteur propre x dans $E - \{0\}$ associé à cette valeur propre la droite vectorielle $D = \mathbb{K}x$ est stable par u. Si tous ces facteurs irréductibles sont de degré 2, ils sont de la forme $X^2 + bX + c$ et le polynôme π_u s'écrit $\pi_u(X) = \left(X^2 + bX + c\right) Q(X)$. De l'égalité :

$$0 = \pi_u(u) = \left(u^2 + bu + cId\right) \circ Q(u)$$

et du caractère minimal de π_u on déduit que l'endomorphisme $u^2 + bu + cId$ n'est pas injectif, c'est-à-dire que son noyau n'est pas réduit à $\{0\}$. Pour tout vecteur x non nul dans ce noyau on vérifie alors que $F = \text{Vect}\{x, u(x)\}$ est un sous-espace vectoriel de dimension 2 stable par u. En effet on sait déjà que F est de dimension 1 ou 2. Si F est de dimension 1 alors x est vecteur propre de u et la valeur propre correspondante est racine de π_u, ce qui contredit le fait que π_u n'a pas de facteur irréductible de degré 1. On a donc F de dimension 2. Avec $u^2(x) + bu(x) + cx = 0$ on déduit que $u^2(x)$ est dans F, ce qui entraîne que F est stable par u.

Exercice 2.3 : *Soient E un $\mathbb{C}$-espace vectoriel de dimension finie, u un endomorphisme de E et P un polynôme de degré strictement positif à coefficients dans $\mathbb{C}$. Déterminer les valeurs propres de $P(u)$.*

Solution : Si $\lambda \in \mathbb{C}$ est valeur propre de u alors pour tout entier naturel k, il est facile de vérifier que λ^k est valeur propre de u^k. On déduit alors que pour tout polynôme P à coefficients dans $\mathbb{C}$, $P(\lambda)$ est valeur propre de $P(u)$.

Si $\mu \in \mathbb{C}$ est valeur propre de $P(u)$ alors, en notant $Q(X) = P(X) - \mu$, l'endomorphisme $Q(u)$ est non injectif. En écrivant, dans $\mathbb{C}[X]$, que $Q(X) = \alpha \prod_{i=1}^{p} (X - \lambda_i)^{m_i}$, on déduit qu'il existe un indice i tel que l'endomorphisme $u - \lambda_i Id$ soit non injectif, c'est-à-dire que λ_i est une valeur propre de u et, avec $Q(\lambda_i) = 0$, on déduit que $\mu = P(\lambda_i)$. En définitive, on a $\text{sp}\big(P(u)\big) = \{P(\lambda) \mid \lambda \in \text{sp}(u)\}$.

Exercice 2.4 : *Soient A et B dans $\mathcal{M}_n(\mathbb{K})$. En remarquant que :*

$$\begin{pmatrix} AB - \lambda I_n & 0 \\ B & -\lambda I_n \end{pmatrix} = \begin{pmatrix} A & \lambda I_n \\ I_n & 0 \end{pmatrix} \begin{pmatrix} B & -\lambda I_n \\ -I_n & A \end{pmatrix},$$

montrer que AB et BA ont même polynôme caractéristique.

Solution : On a :

$$\begin{aligned}(-\lambda)^n \det(AB - \lambda I_n) &= \det\begin{pmatrix} AB - \lambda I_n & 0 \\ B & -\lambda I_n \end{pmatrix} \\ &= \det\begin{pmatrix} A & \lambda I_n \\ I_n & 0 \end{pmatrix}\begin{pmatrix} B & -\lambda I_n \\ -I_n & A \end{pmatrix} \\ &= \det\begin{pmatrix} B & -\lambda I_n \\ -I_n & A \end{pmatrix}\begin{pmatrix} A & \lambda I_n \\ I_n & 0 \end{pmatrix} \\ &= \det\begin{pmatrix} BA - \lambda I_n & B \\ 0 & -\lambda I_n \end{pmatrix} \\ &= (-\lambda)^n \det(BA - \lambda I_n).\end{aligned}$$

Il en résulte que AB et BA ont même polynôme caractéristique.

Exercice 2.5 : *Montrer qu'une matrice $A \in \mathcal{M}_n(\mathbb{K})$ est de rang r si et seulement si elle est équivalente à $A_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$, où I_r désigne la matrice identité d'ordre r. En déduire que pour toutes matrices A et B dans $\mathcal{M}_n(\mathbb{K})$, AB et BA ont même polynôme caractéristique (considérer d'abord le cas où $A = A_r$).*

Solution : Soient $u \in \mathcal{L}(\mathbb{K}^n)$ de matrice A, de rang r, dans la base canonique de $\mathbb{K}^n$, H un supplémentaire de $\ker(u)$ dans $\mathbb{K}^n$, donc de dimension r, $\mathcal{B}_1 = \{e_1, \ldots, e_r\}$ une base de H et $\mathcal{B}_2$ une base de $\ker(u)$. Le système $u(\mathcal{B}_1) = \{u(e_1), \ldots, u(e_r)\}$ est alors libre dans $\mathbb{K}^n$ (si $\sum_{k=1}^{r} \lambda_k u(e_k) = 0$, alors $\sum_{k=1}^{r} \lambda_k e_k \in H \cap \ker(u) = \{0\}$ et tous les λ_k sont nuls du fait que $\mathcal{B}_1$ est libre) et il se complète en une base de $\mathbb{K}^n$:

$$\mathcal{B} = \{u(e_1), \ldots, u(e_r), f_{r+1}, \ldots, f_n\}.$$

La matrice de u dans les bases $\mathcal{B}_1 \cup \mathcal{B}_2$ et $\mathcal{B}$ a alors la forme indiquée. La réciproque est évidente.

Pour $A = A_r$, en écrivant $B = \begin{pmatrix} B_1 & B_2 \\ B_3 & B_4 \end{pmatrix}$ avec B_1 dans $M_r(\mathbb{K})$ et B_4 dans $\mathcal{M}_{n-r}(\mathbb{K})$, on a :

$$AB = \begin{pmatrix} B_1 & B_2 \\ 0 & 0 \end{pmatrix}, \quad BA = \begin{pmatrix} B_1 & 0 \\ B_3 & 0 \end{pmatrix},$$

et

$$\det(AB - \lambda I_n) = (-\lambda)^{n-r} \det(B_1 - \lambda I_n) = \det(BA - \lambda I_n).$$

Dans le cas général, on a $A = PA_rQ$ (le cas $A = 0$ est trivial) et, en notant P_M le polynôme caractéristique de la matrice M, on a :

$$P_{AB} = P_{P(A_rQBP)P^{-1}} = P_{A_r(QBP)} = P_{(QBP)A_r} = P_{Q(BPA_rQ)Q^{-1}} = P_{BA}.$$

Exercice 2.6 : *Montrer que pour toute matrice A dans $\mathcal{M}_n(\mathbb{K})$ il existe une infinité de scalaires λ tels que $A - \lambda I_n$ soit inversible. En utilisant ce résultat, montrer que pour toutes matrices A et B dans $\mathcal{M}_n(\mathbb{K})$, AB et BA ont même polynôme caractéristique (considérer d'abord le cas où la matrice A est inversible).*

Solution : L'application polynomiale $\lambda \mapsto \det(A - \lambda I_n)$ ayant au plus n racines dans $\mathbb{K}$, il en résulte que ce déterminant est non nul pour une infinité de valeurs de λ et que pour ces valeurs la matrice $A - \lambda I_n$ est inversible.

Si A est inversible, alors AB et $BA = A^{-1}(AB)A$ sont semblables et elles ont donc même polynôme caractéristique.

Pour A dans $\mathcal{M}_n(\mathbb{K})$ il existe une infinité de scalaires λ tels que $A - \lambda I_n$ soit inversible et pour ces valeurs de λ on a, en notant P_M le polynôme caractéristique de la matrice M, $P_{(A-\lambda I_n)B}(X) = P_{B(A-\lambda I_n)}(X)$. Pour tout x fixé dans $\mathbb{K}$ on a donc deux polynômes en λ qui coïncident pour une infinité de valeurs. Ils sont donc égaux et $\lambda = 0$ donne $P_{AB}(X) = P_{BA}(X)$.

Exercice 2.7 : *Soient E un $\mathbb{C}$-espace vectoriel de dimension n et u, v dans $\mathcal{L}(E)$.*

1. Montrer que si u et v commutent alors ils ont un vecteur propre en commun.

2. Montrer que si $u \circ v - v \circ u = u$ alors u et v ont un vecteur propre en commun.

3. Montrer que si $u \circ v - v \circ u = \alpha u + \beta v$ alors u et v ont un vecteur propre en commun.

Solution :

1. Sur $\mathbb{C}$ l'endomorphisme u a au moins une valeur propre et, avec la condition $u \circ v = v \circ u$, on vérifie facilement que l'espace propre $E(u, \lambda)$ est stable par v. La restriction de v à $E(u, \lambda)$ définit alors un endomorphisme de $E(u, \lambda)$ et cet endomorphisme admet des valeurs propres dans $\mathbb{C}$. Tout vecteur propre de la restriction de v à $E(u, \lambda)$ associé à l'une de ces valeurs propres est alors un vecteur propre commun à u et v.

2. Si $u \circ v - v \circ u = u$ on vérifie alors par récurrence que pour tout entier naturel k strictement positif on a $u^k \circ v - v \circ u^k = ku^k$. En effet, le résultat est vrai pour $k = 1$ et en le supposant vrai pour $k \geq 1$ on a :

$$\begin{aligned} u^{k+1} \circ v - v \circ u^{k+1} &= u \circ \left(u^k \circ v - v \circ u^k\right) + (u \circ v - v \circ u) \circ u^k \\ &= u \circ \left(ku^k\right) + u \circ u^k = (k+1)\, u^k. \end{aligned}$$

Ce résultat peut se traduire en disant que l'application linéaire définie sur $\mathcal{L}(E)$ par :

$$\varphi : w \mapsto w \circ v - v \circ w$$

est telle que :

$$\forall k \geq 1, \quad \varphi\left(u^k\right) = ku^k.$$

Si u^k est non nul pour tout entier $k \geq 1$ alors chacun de ces vecteurs de $\mathcal{L}(E)$ est un vecteur propre de φ et cet endomorphisme de $\mathcal{L}(E)$ a une infinité de valeurs propres, ce qui est impossible. Il existe donc un entier $k \geq 1$ tel que $u^k = 0$, ce qui entraîne que u n'est pas injectif et 0 est valeur propre de u. Avec la condition $u \circ v - v \circ u = u$ on vérifie alors facilement que le noyau de u est stable par v. Tout vecteur propre de la restriction de v au noyau de u est alors une vecteur propre commun à u et v.

On peut aussi montrer que la condition $u \circ v - v \circ u = u$ entraîne que u n'est pas inversible. En effet, si u est inversible on peut alors écrire que :

$$v = u^{-1} \circ (v + Id) \circ u$$

et on en déduit que v et $v + Id$ ont même polynôme caractéristique. En notant P_v le polynôme caractéristique de v on a alors :

$$\forall \lambda \in \mathbb{C}, \quad P_v(\lambda) = \det(v + Id - \lambda Id) = P_v(\lambda - 1).$$

On déduit alors par récurrence que $P_v(k) = P_v(0)$ pour tout entier naturel k, le polynôme P_v étant non constant, ce qui est impossible. On a donc $\ker(u) \neq \{0\}$ et on conclut comme dans la démonstration qui précède.

3. D'après ce qui précède on peut supposer α et β non nuls et la condition $u \circ v - v \circ u = \alpha u + \beta v$ peut s'écrire :

$$(\alpha u + \beta v) \circ \left(\frac{1}{\alpha} v\right) - \left(\frac{1}{\alpha} v\right) \circ (\alpha u + \beta v) = \alpha u + \beta v.$$

Avec ce qui précède on déduit alors qu'il existe un vecteur non nul x dans le noyau de $\alpha u + \beta v$ qui soit un vecteur propre de v, c'est-à-dire que :

$$(\alpha u + \beta v)(x) = 0, \quad v(x) = \lambda x.$$

On a alors $u(x) = -\dfrac{\beta}{\alpha} \lambda x$ et x est un vecteur propre commun à u et v.

Exercice 2.8 : *Soient α, β dans $\mathbb{K}$ et $A(\alpha, \beta) = \left((a_{ij})\right)_{1 \leq i,j \leq n}$ la matrice d'ordre n supérieur ou égal à 3 définie par :*

$$\forall i \in \{1, 2, \ldots, n\}, \begin{cases} a_{ii} = \beta, \\ a_{ij} = \alpha & \text{si } j \in \{1, 2, \ldots, n\} - \{i\}. \end{cases}$$

1. Calculer $\Delta(\alpha, \beta) = \det\big(A(\alpha, \beta)\big)$.

2. Calculer le polynôme caractéristique et les valeurs propres avec leur multiplicité de la matrice $A(\alpha, \beta)$.

3. Calculer le polynôme minimal de la matrice $A(\alpha, \beta)$.

4. Dans le cas où la matrice $A(\alpha, \beta)$ est inversible, calculer son inverse.

5. Calculer $A(\alpha, \beta)^k$ pour tout entier naturel k.

Solution :

1. La matrice $A(\alpha,\beta)$ est de la forme :

$$A(\alpha,\beta) = \begin{pmatrix} \beta & \alpha & \alpha & \dots & \alpha \\ \alpha & \beta & \alpha & \dots & \alpha \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \alpha & \dots & \alpha & \beta & \alpha \\ \alpha & \dots & \alpha & \alpha & \beta \end{pmatrix}.$$

En ajoutant les lignes 2 à n à la première ligne on a :

$$\Delta(\alpha,\beta) = \det\big(A(\alpha,\beta)\big) = \beta + (n-1)\alpha \begin{vmatrix} 1 & 1 & 1 & \dots & 1 \\ \alpha & \beta & \alpha & \dots & \alpha \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \alpha & \dots & \alpha & \beta & \alpha \\ \alpha & \dots & \alpha & \alpha & \beta \end{vmatrix}.$$

Puis en retranchant la première colonne aux colonnes 2 à n on obtient :

$$\Delta(\alpha,\beta) = \big(\beta + (n-1)\alpha\big)(\beta-\alpha)^{n-1}.$$

2. Le polynôme caractéristique de $A(\alpha,\beta)$ est donné par :

$$\begin{aligned} P_{(\alpha,\beta)}(X) &= \Delta(\alpha,\beta - X) \\ &= (-1)^n\Big(X - \big(\beta + (n-1)\alpha\big)\Big)\big(X - (\beta-\alpha)\big)^{n-1}. \end{aligned}$$

Les valeurs propres de $A(\alpha,\beta)$ sont donc $\lambda_1 = \beta + (n-1)\alpha$ et $\lambda_2 = \beta - \alpha$. Pour $\alpha = 0$, la matrice $A(0,\beta)$ est celle d'une homothétie et $\lambda_1 = \lambda_2 = \beta$ est valeur propre d'ordre n de cette matrice. Pour $\alpha \neq 0$, λ_1 est valeur propre simple et λ_2 est valeur propre d'ordre $n-1$.

On peut aussi remarquer que $A(\alpha,\beta) + (\alpha-\beta)I_n = \alpha A(1,1)$ avec $A(1,1)$ de rang 1. On déduit donc que 0 est valeur propre d'ordre $n-1$ de $A(1,1)$ et que $\mathrm{Tr}\big(A(1,1)\big) = n-1$ est valeur propre simple. On retrouve alors facilement les valeurs propres de $A(\alpha,\beta)$.

3. Si $\alpha = 0$ alors $A(\alpha,\beta) = \beta I_n$ et son polynôme minimal est :

$$\pi_{(\alpha,\beta)}(X) = X - \beta.$$

Pour α non nul, on a :

$$\begin{gathered} A(\alpha,\beta) - (\beta-\alpha)I_n = \alpha A(1,1), \\ A(\alpha,\beta) - \big(\beta + (n-1)\alpha\big)I_n = \alpha A(1, 1-n), \end{gathered}$$

avec $A(1,1)\,A(1,1-n) = 0$. La matrice $A(\alpha,\beta)$ n'étant pas celle d'une homothétie, on déduit que son polynôme minimal est :

$$\pi_{(\alpha,\beta)}(X) = \big(X - (\beta-\alpha)\big)\Big(X - \big(\beta + (n-1)\alpha\big)\Big).$$

4. Pour $\alpha = 0$, on a $A(\alpha,\beta) = \beta I_n$ et cette matrice est inversible si et seulement si $\beta \neq 0$ avec $A(\alpha,\beta)^{-1} = \beta^{-1} I_n$.

On suppose que α est non nul. Dans ce cas, la matrice $A(\alpha,\beta)$ est inversible si et seulement si $\beta \neq \alpha$ et $\beta \neq -(n-1)\alpha$ et son polynôme minimal est $\pi_{(\alpha,\beta)}(X) = (X-\lambda_1)(X-\lambda_2)$. L'égalité $\pi_{(\alpha,\beta)}\big(A(\alpha,\beta)\big) = 0$ donne alors :

$$\begin{aligned} A(\alpha,\beta)^{-1} &= \frac{1}{\lambda_1\lambda_2}\left((\lambda_1+\lambda_2) I_n - A(\alpha,\beta)\right) \\ &= \frac{1}{\lambda_1\lambda_2} A(-\alpha, \lambda_1+\lambda_2-\beta) = \frac{1}{\lambda_1\lambda_2} A\big(-\alpha, \beta+(n-1)\alpha\big). \end{aligned}$$

5. Pour $\alpha = 0$, on a $A(\alpha,\beta) = \beta I_n$ et pour tout entier naturel k, on a $A(\alpha,\beta)^k = \beta^k I_n$.

On suppose que α est non nul. Pour $k=0$ et $k=1$, on a :

$$A(\alpha,\beta)^k = a_k A(\alpha,\beta) + b_k I_n$$

avec :

$$(a_0,b_0) = (0,1), \quad (a_1,b_1) = (1,0).$$

Pour $k \geq 2$, la division euclidienne de X^k par $\pi_{(\alpha,\beta)}(X)$ s'écrit :

$$X^k = Q(X)\pi_{(\alpha,\beta)}(X) + a_k X + b_k,$$

avec :

$$\begin{cases} a_k = \dfrac{\lambda_1^k - \lambda_2^k}{\lambda_1-\lambda_2} = \dfrac{\lambda_1^k - \lambda_2^k}{n\alpha}, \\ b_k = \dfrac{\lambda_1\lambda_2^k - \lambda_2\lambda_1^k}{\lambda_1-\lambda_2} = \dfrac{\lambda_1\lambda_2}{n\alpha}\left(\lambda_2^{k-1} - \lambda_1^{k-1}\right). \end{cases}$$

Et on a :

$$\forall k \in \mathbb{N}, \quad A(\alpha,\beta)^k = a_k A(\alpha,\beta) + b_k I_n.$$

On peut aussi écrire :

$$\forall k \in \mathbb{N}, \quad A(\alpha,\beta)^k = \frac{1}{n\alpha}\Big(\lambda_1^k\big(A(\alpha,\beta) - \lambda_2 I_n\big) + \lambda_2^k\big(\lambda_1 I_n - A(\alpha,\beta)\big)\Big)$$

$\Big(\frac{1}{nb}\left(A(\alpha,\beta) - \lambda_2 I_n\right)$ et $\frac{1}{nb}\left(\lambda_1 I_n - A(\alpha,\beta)\right)$ sont les projecteurs spectraux$\Big)$, ou encore :

$$\forall k \in \mathbb{N}, \quad A(\alpha,\beta)^k = \frac{1}{n}\big(\lambda_2^k A(1,1) - \lambda_1^k A(1,1-n)\big).$$

Exercice 2.9 :

1. Pour toute matrice complexe $A = \left((a_{ij})\right)_{1\leq i,j\leq n}$ *d'ordre* n *supérieur ou égal à 2 et pour tout nombre complexe* t*, on note* $A_t = \left((a_{ij}+t)\right)_{1\leq i,j\leq n}$*. Montrer que :*

$$\forall t \in \mathbb{C}, \quad \det(A_t) = \det(A) + tS(A),$$

la constante réelle $S(A)$ *ne dépendant que des coefficients de* A*.*

2. Soient a, b, c *des nombres complexes et* $M(a,b,c) = \left((m_{ij})\right)_{1\leq i,j\leq n}$ *la matrice d'ordre* n *supérieur ou égal à 2 définie par :*

$$\forall i \in \{1,2,\ldots,n\}, \begin{cases} m_{ii} = b, \\ m_{ij} = c & si\ j \in \{i+1,\ldots,n\}, \\ m_{ij} = a & si\ j \in \{1,\ldots,i-1\}. \end{cases}$$

(a) En prenant, pour $a \neq c$*,* $A = M(a,b,c)$*,* $t = -a$ *et* $t = -c$ *dans la question précédente, calculer le déterminant de* $M(a,b,c)$*.*
(b) Calculer le polynôme caractéristique et les valeurs propres de $M(a,b,c)$*.*

Solution :

1. En notant C_j la colonne numéro j de A, T le vecteur de composantes toutes égales à t, et en utilisant la n-linéarité du déterminant on déduit que :

$$\det(A_t) = \det(C_1 + T, \ldots, C_n + T) = \det(A) + tS(A),$$

où on a noté :

$$S(A) = \sum_{j=1}^{n} \det(C_1, \ldots, C_{j-1}, E, C_j, \ldots, C_n),$$

E désignant le vecteur de composantes toutes égales à 1.

2. La matrice $M(a,b,c)$ s'écrit :

$$M(a,b,c) = \begin{pmatrix} b & c & c & \ldots & c \\ a & b & c & \ldots & c \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ a & \ldots & a & b & c \\ a & \ldots & a & a & b \end{pmatrix}.$$

(a) Pour $a = c$, la matrice $M(a,b,a)$ est symétrique et ce cas a été étudié avec l'exercice 2.8. Dans ce cas, on a :

$$\det\big(M(a,b,a)\big) = \big(b + (n-1)a\big)(b-a)^{n-1}.$$

Pour $a \neq c$, en prenant $A = M(a,b,c)$ dans la question précédente, on a :

$$\det(A_{-a}) = (b-a)^n = \det(A) - aS(A),$$
$$\det(A_{-c}) = (b-c)^n = \det(A) - cS(A).$$

On en déduit facilement que :

$$\det\big(M(a,b,c)\big) = \frac{c(b-a)^n - a(b-c)^n}{c-a}.$$

(b) Pour $a = c$, le polynôme caractéristique et les valeurs propres de la matrice $M(a, b, a)$ ont été calculés à l'exercice 2.8. On suppose donc que $a \neq c$. Dans ce cas, le polynôme caractéristique de $M(a, b, c)$ est donné par :

$$P_{a,b,c}(\lambda) = \frac{c(b - \lambda - a)^n - a(b - \lambda - c)^n}{c - a}.$$

Pour $a = 0$ ou $c = 0$, la matrice $M(a, b, c)$ est triangulaire avec b pour valeur propre d'ordre n. Pour $a \neq 0$ et $c \neq 0$, on a :

$$P_{a,b,c}(b - a) = a(a - c)^{n-1} \neq 0, \quad P_{a,b,c}(b - c) = c(c - a)^{n-1} \neq 0.$$

On peut donc écrire que les valeurs propres de $M(a, b, c)$ sont définies par :

$$\left(\frac{b - \lambda - a}{b - \lambda - c}\right)^n = \frac{a}{c}.$$

En notant $\omega_1, \ldots, \omega_n$ les racines n-ième dans $\mathbb{C}$ de $\dfrac{a}{c}$ (qui est non nul), on déduit alors que les valeurs propres de $M(a, b, c)$ sont :

$$\lambda_k = b + \frac{c\omega_k - a}{1 - \omega_k} \quad (1 \leq k \leq n)$$

(comme $a \neq c$, on a $\omega_k \neq 1$).

Exercice 2.10 : *Soient a, b dans $\mathbb{R}$ et $A(a, b)$ la matrice réelle d'ordre n supérieur ou égal à 2 définie par :*

$$A(a, b) = \begin{pmatrix} a & b & 0 & \ldots & 0 \\ b & a & b & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b & a & b \\ 0 & \ldots & 0 & b & a \end{pmatrix}.$$

Calculer les valeurs propres et les vecteurs propres associés de $A(a, b)$.

Solution : On peut écrire $A(a, b) = aI_n + bA(0, 1)$. Il suffit donc de considérer le cas où $(a, b) = (0, 1)$.

La matrice $A(0, 1)$ est symétrique réelle, donc toutes ses valeurs propres sont réelles. Avec le théorème de Gerschgörin-Hadamard, on déduit que pour toute valeur propre $\lambda \in \mathbb{R}$ on a $|\lambda| \leq 2$. Une telle valeur propre peut donc s'écrire $\lambda = 2\cos(\alpha)$ avec $\alpha \in [0, \pi]$. Si x est un vecteur propre non nul associé ses composantes sont solutions de la récurrence :

$$x_{k-1} - \lambda x_k + x_{k+1} = 0 \quad (1 \leq k \leq n),$$

avec les conditions aux limites $x_0 = 0$ et $x_{n+1} = 0$.

Le polynôme caractéristique de cette récurrence est :

$$P(r) = r^2 - 2\cos(\alpha) r + 1,$$

soit $P(r) = \left(r - e^{i\alpha}\right)\left(r - e^{-i\alpha}\right)$. Les racines sont donc $r_1 = e^{i\alpha}$ et $r_2 = e^{-i\alpha}$. Ce qui donne :

$$x_k = c_1 e^{ik\alpha} + c_2 e^{-ik\alpha} \quad (0 \leq k \leq n+1).$$

Avec $x_0 = 0$, on déduit que $c_2 = -c_1$. Avec $x_{n+1} = 0$ et $(c_1, c_2) \neq (0,0)$, on déduit que $\sin\big((n+1)\alpha\big) = 0$ et $\alpha = \dfrac{j\pi}{n+1}$ avec $1 \leq j \leq n$.

Les valeurs propres de $A(a,b)$ sont donc :

$$\lambda_k(a,b) = a + 2b\cos\left(\frac{k\pi}{n+1}\right) \quad (1 \leq k \leq n).$$

L'espace propre associé à $\lambda_k(a,b)$ est la droite engendrée par le vecteur $v^{(k)}$ de composantes :

$$v_j^{(k)} = \sin\left(j\frac{k\pi}{n+1}\right) \quad (1 \leq k, j \leq n).$$

Exercice 2.11 : *Soit u un endomorphisme de E nilpotent d'ordre 2. Montrer que si $\mathcal{B}_1 = \{u(e_1), \ldots, u(e_r)\}$ est une base de* $\mathrm{Im}(u)$, *alors elle se complète en une base $\mathcal{B}_2 = \left\{u(e_1), \ldots, u(e_r), e'_1, \ldots, e'_p\right\}$ de* $\ker(u)$ *et le système :*

$$\mathcal{B} = \left\{u(e_1), \ldots, u(e_r), e'_1, \ldots, e'_p, e_1, \ldots, e_r\right\}$$

est une base de E. Écrire la matrice de u dans cette base.

Solution : De $u^2 = 0$ on déduit que $\mathrm{Im}(u)$ est contenue dans $\ker(u)$. On peut donc compléter $\mathcal{B}_1$ en une base $\mathcal{B}_2$ de $\ker(u)$. En utilisant le théorème du rang on a :

$$\mathrm{card}(\mathcal{B}_2) = n - r = r + p,$$

et on déduit que $p = n - 2r$ et $\mathrm{card}(\mathcal{B}) = n$. Il suffit donc de montrer que le système $\mathcal{B}$ est libre. Si

$$\sum_{k=1}^{r} \alpha_k e_k + \sum_{k=1}^{r} \beta_k u(e_k) + \sum_{k=1}^{p} \gamma_k e'_k = 0,$$

en appliquant u on a $\sum\limits_{k=1}^{r} \alpha_k u(e_k) = 0$ et tous les α_k sont nuls. Le système $\mathcal{B}_2$ étant libre on en déduit ensuite que tous les β_k et les γ_k sont nuls. En conclusion, le système $\mathcal{B}$ est libre et c'est une base de E.

La matrice de u dans cette base est alors :

$$A = \begin{pmatrix} 0 & 0 & I_r \\ 0 & 0 & 0 \end{pmatrix}.$$

Exercice 2.12 : *Soient $a_1, \ldots, a_n$ des scalaires. Calculer le polynôme caractéristique et le polynôme minimal de la matrice :*

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & 0 & 1 \\ a_1 & a_2 & \ldots & a_{n-1} & a_n \end{pmatrix}.$$

Solution : On désigne par $u \in \mathcal{L}(\mathbb{K}^n)$ l'endomorphisme que définit la matrice A dans la base canonique $\{e_1, \ldots, e_n\}$ de $\mathbb{K}^n$ et on a :

$$\begin{cases} u(e_1) = a_1 e_n, \\ u(e_k) = e_{k-1} + a_k e_n \quad (2 \leq k \leq n). \end{cases}$$

Par récurrence sur $k \in \{1, \ldots, n-1\}$ on déduit alors qu'il existe des coefficients $\alpha_{k,j}$ tels que :

$$u^k(e_n) = e_{n-k} + \sum_{j=n-k+1}^{n} \alpha_{k,j} e_j.$$

En effet, le résultat est vrai pour $k = 1$ et, en le supposant acquis pour $k \in \{1, \ldots, n-2\}$, on a :

$$\begin{aligned} u^{k+1}(e_n) &= e_{n-k-1} + a_{n-k} e_n + \sum_{j=n-k+1}^{n} \alpha_{k,j} (e_{j-1} + a_j e_n) \\ &= e_{n-(k+1)} + \sum_{j=n-k}^{n} \alpha_{k+1,j} e_j. \end{aligned}$$

On déduit alors que le système $\{e_n, u(e_n), \ldots, u^{n-1}(e_n)\}$ est libre dans $\mathbb{K}^n$ (la matrice du système $\{u^{n-1}(e_n), \ldots, u(e_n), e_n\}$ dans la base $\{e_1, \ldots, e_n\}$ est triangulaire inférieure de diagonale unité) et le polynôme minimal de u est nécessairement de degré supérieur ou égal à n. Ce polynôme étant de degré au plus égal à n (théorème de Cayley-Hamilton), il en résulte que $\deg(\pi_A) = n$ et $\pi_A = (-1)^n P_A$ où P_A désigne le polynôme caractéristique de A.

En notant $P_A = P_{(a_1,\ldots,a_n)}$ et en développant ce déterminant par rapport à la première colonne, on a :

$$P_{(a_1,\ldots,a_n)}(X) = -X\, P_{(a_2,\ldots,a_n)}(X) + (-1)^{n+1} a_1,$$

et, par récurrence, $P_A(X) = (-1)^n \left(X^n - \sum\limits_{k=1}^{n} a_k X^{k-1} \right)$.

Exercice 2.13 : *Montrer que l'inverse d'une matrice triangulaire supérieure [resp. inférieure] et inversible est une matrice triangulaire supérieure [resp. inférieure].*

Solution : Résulte immédiatement du fait que A^{-1} est un polynôme en A.

Exercice 2.14 : *Soit $P(X) = X^n + p_1 X^{n-1} + \cdots + p_{n-1}X + p_n$ un polynôme unitaire de degré $n \geq 2$ à coefficients dans $\mathbb{C}$. On désigne par $\lambda_1, \ldots, \lambda_n$ les racines complexes de P et par $(\sigma_k)_{1\leq k\leq n}$ les fonctions symétriques élémentaires des racines définies par :*

$$\sigma_k(\lambda_1, \ldots, \lambda_n) = \sum_{1\leq i_1<\ldots<i_k\leq n} \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_k} \quad (1 \leq k \leq n).$$

On définit également les fonctions $(S_k)_{1\leq k\leq n}$ par :

$$S_k(\lambda_1, \ldots, \lambda_n) = \sum_{i=1}^{n} \lambda_i^k \quad (1 \leq k \leq n).$$

On notera σ_k pour $\sigma_k(\lambda_1, \ldots, \lambda_n)$ et S_k pour $S_k(\lambda_1, \ldots, \lambda_n)$.

1. Montrer que :

$$\forall j \in \{1, \ldots, n\}, \quad p_j = (-1)^j \sigma_j.$$

2. Pour tout entier i compris entre 1 et n on désigne par P_i le polynôme de degré $n-1$ défini par :

$$P(X) = (X - \lambda_i) P_i(X).$$

(a) En écrivant que $P(X) = P(X) - P(\lambda_i)$, montrer que pour tout entier k compris entre 1 et $n-1$ le coefficient de X^{n-k-1} dans P_i est donné par :

$$\alpha_{i,k} = \lambda_i^k + \sum_{j=1}^{k} (-1)^j \sigma_j \lambda_i^{k-j}.$$

(b) En écrivant le polynôme dérivé de P sous la forme $P' = \sum\limits_{i=1}^{n} P_i$, montrer que pour tout entier k compris entre 1 et $n-1$ on a :

$$(-1)^k (n-k) \sigma_k = \sum_{i=1}^{n} \alpha_{i,k}.$$

(c) Déduire de ce qui précède que pour tout entier k compris entre 1 et n on a :

$$S_k - \sigma_1 S_{k-1} + \cdots + (-1)^{k-1}\sigma_{k-1}S_1 + (-1)^k \sigma_k k = 0,$$

puis que :

$$S_k + p_1 S_{k-1} + \cdots + p_{k-1}S_1 + p_k k = 0$$

(formules de Newton).

Solution :

1. On procède par récurrence sur $n \geq 1$. Pour ce faire on note $\lambda = (\lambda_1, \ldots, \lambda_n)$, $\lambda' = (\lambda_1, \ldots, \lambda_{n-1})$ et on exprime, pour tout entier k compris entre 0 et n, $\sigma_{n,k}(\lambda)$ en fonction de λ_n et des $\sigma_{n-1,j}(\lambda')$ $(0 \leq j \leq n-1)$. On vérifie facilement, en posant $\sigma_{n,0}(\lambda) = 1$, que :

$$\begin{aligned}
&\sigma_{n,0}(\lambda) = \sigma_{n-1,0}(\lambda') = 1, \\
&\sigma_{n,k}(\lambda) = \sigma_{n-1,k}(\lambda') + \lambda_n \sigma_{n-1,k-1}(\lambda') \quad (1 \leq k \leq n-1), \\
&\sigma_{n,n}(\lambda) = \lambda_n \sigma_{n-1,n-1}(\lambda').
\end{aligned} \tag{2.3}$$

Pour $n = 1$, on a $P(X) = X - \lambda_1 = X - \sigma_{1,1}(\lambda)$. Supposons le résultat acquis pour $n - 1 \geq 1$. On a alors :

$$\begin{aligned}
P(X) &= \prod_{i=1}^{n}(X - \lambda_i) = \prod_{i=1}^{n-1}(X - \lambda_i)(X - \lambda_n) \\
&= X^n + \sum_{k=1}^{n-1}(-1)^k\left(\sigma_{n-1,k}(\lambda') + \sigma_{n-1,k-1}(\lambda')\lambda_n\right)X^{n-k} \\
&\quad + (-1)^n \sigma_{n-1,n-1}(\lambda')\lambda_n
\end{aligned}$$

et on conclut alors avec les identités (2.3).

2. (a) On a :

$$\begin{aligned}
(X - \lambda_i) P_i(X) = P(X) &= P(X) - P(\lambda_i) \\
&= X^n - \lambda_i^n + \sum_{j=1}^{n-1}(-1)^j \sigma_j \left(X^{n-j} - \lambda_i^{n-j}\right).
\end{aligned}$$

Donc :

$$P_i(X) = \frac{X^n - \lambda_i^n}{X - \lambda_i} + \sum_{j=1}^{n-1}(-1)^j \sigma_j \frac{X^{n-j} - \lambda_i^{n-j}}{X - \lambda_i},$$

où pour tout entier p strictement positif :

$$\frac{X^p - \lambda_i^p}{X - \lambda_i} = X^{p-1} + X^{p-2}\lambda_i + \cdots + X\lambda_i^{p-2} + \lambda_i^{p-1}$$

est un polynôme de degré $p - 1$. On déduit donc que pour tout entier k compris entre 1 et $n - 1$ le coefficient de X^{n-k-1} dans P_i est donné par :

$$\alpha_{i,k} = \lambda_i^k + \sum_{j=1}^{k}(-1)^j \sigma_j \lambda_i^{k-j}.$$

(b) On a :

$$P'(X) = nX^{n-1} + \sum_{j=1}^{n-1} (-1)^j (n-j)\, \sigma_j X^{n-j-1} = \sum_{i=1}^{n} P_i$$

et en identifiant les coefficients de X^{n-k-1}, pour tout entier k compris entre 1 et $n-1$, on déduit que :

$$(-1)^k (n-k)\, \sigma_k = \sum_{i=1}^{n} \alpha_{i,k}.$$

(c) L'identité précédente peut s'écrire, pour tout entier k compris entre 1 et $n-1$:

$$(-1)^k (n-k)\, \sigma_k = S_k + \sum_{j=1}^{k-1} (-1)^j \sigma_j S_{k-j} + (-1)^k n \sigma_k,$$

ou encore :

$$S_k - \sigma_1 S_{k-1} + \cdots + (-1)^{k-1} \sigma_{k-1} S_1 + (-1)^k \sigma_k k = 0.$$

En utilisant les égalités $p_j = (-1)^j \sigma_j$, on déduit que, pour tout entier k compris entre 1 et $n-1$, on a :

$$S_k + p_1 S_{k-1} + \cdots + p_{k-1} S_1 + p_k k = 0.$$

Pour $k = n$ l'identité s'obtient en écrivant que pour tout entier i compris entre 1 et n on a :

$$0 = P(\lambda_i) = \lambda_i^n + p_1 \lambda_i^{n-1} + \cdots + p_{n-1} \lambda_i + p_n$$

et en faisant la somme de toutes ces égalités.

Exercice 2.15 : *En utilisant les formules de Newton (exercice 2.14), montrer que pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$ les matrices AB et BA ont même polynôme caractéristique.*

Solution : On note :

$$P_{AB}(\lambda) = (-1)^n \left(\lambda^n + p_1 \lambda^{n-1} + \cdots + p_{n-1} \lambda + p_n\right),$$
$$P_{BA}(\lambda) = (-1)^n \left(\lambda^n + q_1 \lambda^{n-1} + \cdots + q_{n-1} \lambda + q_n\right),$$

les polynômes caractéristiques des matrices AB et BA.

On a :

$$p_1 = -\operatorname{Tr}(AB) = -\operatorname{Tr}(BA) = q_1.$$

Supposons que $p_j = q_j$ pour tout entier j compris entre 1 et $k-1$ avec $2 \leq k \leq n$. On a alors, en utilisant les formules de Newton relatives à la matrice AB :

$$\begin{aligned} kp_k &= -\left(S_k + p_1 S_{k-1} + \cdots + p_{k-1} S_1\right) \\ &= -\left(S_k + q_1 S_{k-1} + \cdots + q_{k-1} S_1\right) \end{aligned}$$

avec, pour tout entier j compris entre 1 et $k-1$:

$$S_j = \mathrm{Tr}\left(A\left(B\left(AB\right)^{j-1}\right)\right) = \mathrm{Tr}\left(\left(B\left(AB\right)^{j-1}\right)A\right) = \mathrm{Tr}\left(\left(BA\right)^j\right).$$

On déduit alors que $kp_k = kq_k$.

On a donc ainsi montré par récurrence que AB et BA ont même polynôme caractéristique.

Cette démonstration est valable pour tout corps commutatif $\mathbb{K}$.

Exercice 2.16 : *Écrire la décomposition de Dunford-Schwarz de la matrice :*

$$A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in M_4\left(\mathbb{C}\right).$$

En déduire un calcul de A^r pour tout entier r strictement positif.

Solution : Le polynôme caractéristique de A est $P_A(X) = X\left(X-1\right)^3$ et son polynôme minimal est $\pi_A\left(X\right) = X\left(X-1\right)^2$. On a la décomposition en éléments simples :

$$\frac{1}{\pi_A(X)} = \frac{1}{X} + \left(\frac{1}{\left(X-1\right)^2} - \frac{1}{X-1}\right)$$

qui donne la décomposition de Bézout :

$$1 = \left(X-1\right)^2 + \left(2X - X^2\right)$$

et les projecteurs spectraux :

$$\pi_1 = \left(A - I_4\right)^2, \quad \pi_2 = 2A - A^2$$

(on a $\pi_1 + \pi_2 = I_4$). On obtient alors la décomposition $A = D + V$ avec :

$$D = \pi_2 = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad V = A - D = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Pour $r > 0$, on a :

$$A^r = D^r + rD^{r-1}V$$

$(D^2 = 0)$ avec $D^r = \pi_2^r = \pi_2 = D$. Soit :

$$\forall r \geq 2, \quad A^r = D\left(I_4 + rV\right) = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & r & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

3 Réduction des endomorphismes et des matrices

On garde les notations du chapitre 2.

1. Trigonalisation

Définition 3.1 : *On dit qu'un endomorphisme $u \in \mathcal{L}(E)$ est trigonalisable s'il existe une base de E dans laquelle la matrice de u est triangulaire.*

Si $u \in \mathcal{L}(E)$ a pour matrice A dans une base $\mathcal{B} = \{e_1, e_2, \dots, e_n\}$ et si de plus u est trigonalisable il existe alors une base $\mathcal{B}' = \{f_1, f_2, \dots, f_n\}$ dans laquelle la matrice T de u est triangulaire. On sait alors que les matrices A et T sont semblables, c'est-à-dire qu'il existe une matrice d'ordre n à coefficients dans $\mathbb{K}$ inversible P telle que $T = P^{-1}AP$. Ce qui nous amène à poser la définition suivante.

Définition 3.2 : *On dit qu'une matrice A d'ordre n à coefficients dans $\mathbb{K}$ est trigonalisable si elle est semblable à une matrice triangulaire.*

Deux matrices semblables ayant même déterminant on en déduit qu'elles ont même polynôme caractéristique. En conséquence, si la matrice A est semblable à une matrice triangulaire T alors les termes diagonaux de T sont les valeurs propres dé A.

On en déduit également qu'une matrice à coefficients réels n'ayant pas toutes ses valeurs propres réelles n'est pas trigonalisable dans $\mathcal{M}_n(\mathbb{R})$. Par exemple la matrice :

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

de valeurs propres complexes i et $-i$ n'est pas trigonalisable dans $M_2(\mathbb{R})$.

Une conséquence importante du théorème de d'Alembert-Gauss est que pour $\mathbb{K} = \mathbb{C}$ tout endomorphisme et toute matrice sont trigonalisables.

Lemme 3.1 : *Soient E un $\mathbb{C}$-espace vectoriel de dimension $n \geq 2$ et u un endomorphisme de E. Il existe un hyperplan de E stable par u.*

Démonstration : Le corps $\mathbb{C}$ étant algébriquement clos, on en déduit que u admet au moins une valeur propre $\lambda \in \mathbb{C}$. L'endomorphisme $u - \lambda Id$ est alors non injectif et son image $\mathrm{Im}\,(u - \lambda Id)$ est un sous-espace vectoriel de dimension inférieure ou égale à $n - 1$. Il existe donc un hyperplan H de E qui contient $\mathrm{Im}\,(u - \lambda Id)$. On a alors :

$$\forall x \in E, \quad u\,(x) - \lambda x \in H$$

et

$$\forall x \in H, \quad u\,(x) = \big(u\,(x) - \lambda x\big) + \lambda x \in H,$$

c'est-à-dire que H est stable par u. ∎

Théorème 3.1 : *Si E est un $\mathbb{C}$-espace vectoriel de dimension $n \geq 1$ alors tout endomorphisme de E est trigonalisable.*

Démonstration : On raisonne par récurrence sur la dimension $n \geq 1$ de E. Pour $n = 1$ le résultat est évident. Supposons-le acquis pour les espaces vectoriels complexes de dimension $n-1$. Si u est un endomorphisme sur l'espace vectoriel complexe E de dimension $n \geq 2$ alors il existe un hyperplan H de E stable par u et la restriction v de u à H est trigonalisable, c'est-à-dire qu'il existe une base $\mathcal{B}_1 = \{e_1, e_2, \ldots, e_{n-1}\}$ de H dans laquelle la matrice de v est triangulaire supérieure. Pour tout vecteur $e_n \in E - H$ le système $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ est une base de E et la matrice de u dans cette base est triangulaire supérieure. ∎

Corollaire 3.1 : *Soient E un $\mathbb{C}$-espace vectoriel de dimension $n \geq 1$ et u un endomorphisme de E. La trace de u est égale à la somme des valeurs propres de u et le déterminant de u est égal au produit des valeurs propres de u.*

Corollaire 3.2 : *Toute matrice à coefficients complexes est trigonalisable.*

On a déjà vu que le résultat du théorème 3.1 est faux si le corps de base est $\mathbb{R}$. Ce qui est important dans la démonstration précédente c'est que u admet des valeurs propres ainsi que la restriction de u à tout sous-espace vectoriel de E stable par u, ce qui est une conséquence du fait que tout polynôme de $\mathbb{C}\,[X]$ est scindé sur $\mathbb{C}$. Pour $\mathbb{K} = \mathbb{R}$ on a le résultat suivant, valable en fait pour tout corps commutatif.

Théorème 3.2 : *Soient E un $\mathbb{R}$-espace vectoriel de dimension $n \geq 1$ et u un endomorphisme de E. L'endomorphisme u est trigonalisable sur $\mathbb{R}$ si et seulement si son polynôme caractéristique est scindé sur $\mathbb{R}$.*

Démonstration : On note P_u le polynôme caractéristique de u. Il est clair que si u est trigonalisable alors son polynôme caractéristique est scindé sur $\mathbb{R}$ (raisonner avec la matrice de u dans une base de trigonalisation).

Réciproquement, si P_u est scindé sur $\mathbb{R}$ alors u admet au moins une valeur propre $\lambda \in \mathbb{R}$ et comme dans le cas complexe on montre qu'il existe un hyperplan H de E qui est stable par u. En utilisant le lemme 2.1 on déduit que le polynôme caractéristique de la restriction de u à H est également scindé sur $\mathbb{R}$. La démonstration se termine alors par récurrence comme dans le cas complexe. ∎

Corollaire 3.3 : *Soient E un $\mathbb{R}$-espace vectoriel de dimension $n \geq 1$ et u un endomorphisme de E. Si u est trigonalisable et si F est un sous-espace vectoriel de E stable par u alors la restriction de u à F est aussi trigonalisable.*

Corollaire 3.4 : *Soient E un $\mathbb{R}$-espace vectoriel de dimension $n \geq 1$ et u un endomorphisme de E. Si u est trigonalisable alors la trace de u est égale à la somme des valeurs propres de u et le déterminant de u est égal au produit des valeurs propres de u.*

Corollaire 3.5 : *Toute matrice à coefficients réels dont le polynôme caractéristique est scindé sur $\mathbb{R}$ est trigonalisable.*

2. Diagonalisation

Définition 3.3 : *On dit qu'un endomorphisme $u \in \mathcal{L}(E)$ est diagonalisable s'il existe une base de E dans laquelle la matrice de u est diagonale.*

Si $u \in \mathcal{L}(E)$ a pour matrice A dans une base $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ et si de plus u est diagonalisable il existe alors une base $\mathcal{B}' = \{f_1, f_2, \ldots, f_n\}$ dans laquelle la matrice D de u est diagonale. On sait alors que les matrices A et D sont semblables, c'est-à-dire qu'il existe une matrice d'ordre n à coefficients dans $\mathbb{K}$ inversible P telle que $D = P^{-1}AP$. Ce qui nous amène à poser la définition suivante.

Définition 3.4 : *On dit qu'une matrice A d'ordre n à coefficients dans $\mathbb{K}$ est diagonalisable si elle est semblable à une matrice diagonale.*

Une condition suffisante de diagonalisation est donnée par le résultat suivant.

Théorème 3.3 : *Si u est un endomorphisme de E ayant n valeurs propres distinctes dans $\mathbb{K}$ alors u est diagonalisable.*

Démonstration : Si u a n valeurs propres distinctes alors ces valeurs propres sont toutes simples et chaque sous-espace propre de u est de dimension 1. Il est facile de vérifier que ces sous-espaces propres sont en somme directe. Il en résulte alors que u est diagonalisable. ■

Des conditions nécessaires et suffisantes pour qu'un endomorphisme soit diagonalisable sont données par le résultat suivant.

Théorème 3.4 : *Soit u un endomorphisme de E. Les conditions suivantes sont équivalentes :*

(i) *l'endomorphisme u est diagonalisable ;*

(ii) *si $\lambda_1, \ldots, \lambda_p$ sont les valeurs propres de u deux à deux distinctes dans $\mathbb{K}$, alors $E = \bigoplus\limits_{k=1}^{p} \ker(u - \lambda_k Id)$;*

(iii) *si $\lambda_1, \ldots, \lambda_p$ sont les valeurs propres de u deux à deux distinctes dans $\mathbb{K}$, alors $\sum\limits_{k=1}^{p} \dim\big(\ker(u - \lambda_k Id)\big) = n$;*

(iv) *le polynôme caractéristique de u est scindé sur $\mathbb{K}$ de racines deux à deux distinctes $\lambda_1, \ldots, \lambda_p$ dans $\mathbb{K}$, chaque λ_k $(1 \leq k \leq p)$ étant de multiplicité $\alpha_k = \dim\big(\ker(u - \lambda_k Id)\big)$;*

(v) *il existe un polynôme annulateur de u scindé à racines simples dans $\mathbb{K}$;*

(vi) *le polynôme minimal π_u est scindé à racines simples dans $\mathbb{K}$.*

Démonstration :

(i) $\Leftrightarrow$ (ii) Si u est diagonalisable il existe alors une base de E :

$$\mathcal{B} = \left\{e_{11}, \ldots, e_{1,\alpha_1}, \ldots, e_{p1}, \ldots, e_{p,\alpha_p}\right\}$$

formée de vecteurs propres avec :

$$u\left(e_{k,j}\right) = \lambda_k e_{k,j} \quad \left(1 \leq k \leq p, 1 \leq j \leq \alpha_k\right),$$

où les $\lambda_1, \ldots, \lambda_p$ sont les valeurs propres distinctes deux à deux de u. On déduit alors facilement que $E = \bigoplus_{k=1}^{p} \ker\left(u - \lambda_k Id\right)$ avec $\dim\left(\ker\left(u - \lambda_k Id\right)\right) = \alpha_k$. La réciproque est évidente.

(ii) $\Leftrightarrow$ (iii) Cette équivalence est évidente (les espaces propres sont en somme directe).

(iii) $\Rightarrow$ (iv) Si $\sum_{k=1}^{p} \dim\left(\ker\left(u - \lambda_k Id\right)\right) = n$ alors $E = \bigoplus_{k=1}^{p} \ker\left(u - \lambda_k Id\right)$ et le polynôme caractéristique de u s'écrit $P_u\left(X\right) = \prod_{k=1}^{p} \left(X - \lambda_k\right)^{\alpha_k}$.

(iv) $\Rightarrow$ (v) Du fait que $\dim\left(\ker\left(u - \lambda_k Id\right)\right)$ est égale à la multiplicité de λ_k on déduit que $\sum_{k=1}^{p} \dim\left(\ker\left(u - \lambda_k Id\right)\right) = n$ et donc que $E = \bigoplus_{k=1}^{p} \ker\left(u - \lambda_k Id\right)$. On a alors $P\left(u\right) = 0$ avec $P\left(X\right) = \prod_{k=1}^{p} \left(X - \lambda_k\right)$ scindé sur $\mathbb{K}$ à racines simples.

(v) $\Rightarrow$ (vi) Le polynôme minimal étant un diviseur de tout polynôme annulateur, il est également scindé sur $\mathbb{K}$ à racines simples si (v) est vérifié.

(vi) $\Rightarrow$ (i) Si le polynôme minimal de u s'écrit $\pi_u\left(X\right) = \prod_{k=1}^{p} \left(X - \lambda_k\right)$, les λ_k étant deux à deux distincts, alors le théorème de décomposition des noyaux donne :

$$E = \bigoplus_{k=1}^{p} \ker\left(u - \lambda_k Id\right),$$

c'est-à-dire que u est diagonalisable. ∎

3. Espaces vectoriels euclidiens

Pour ce paragraphe, E est un espace vectoriel sur $\mathbb{R}$ de dimension $n \geq 1$.

Définition 3.5 : *On appelle produit scalaire euclidien sur E toute application*

$$\begin{array}{rccc} \varphi: & E \times E & \to & \mathbb{R} \\ & (x, y) & \mapsto & \varphi\left(x, y\right) \end{array}$$

qui vérifie les propriétés suivantes :

- *pour tout x dans E l'application $y \mapsto \varphi(x,y)$ est linéaire et pour tout y dans E l'application $x \mapsto \varphi(x,y)$ est linéaire (φ est une forme bilinéaire sur E) ;*
- *$\varphi(x,y) = \varphi(y,x)$ pour tous x,y dans E (φ est symétrique) ;*
- *$\varphi(x,x) \geq 0$ pour tout x dans E (φ est positive) ;*
- *pour x dans E, $\varphi(x,x) = 0$ équivaut à $x = 0$ (φ est définie).*

On note en général :

$$(x,y) \longmapsto \langle x \mid y \rangle$$

un tel produit scalaire.

Définition 3.6 : *Un espace euclidien est un $\mathbb{R}$-espace vectoriel de dimension finie muni d'un produit scalaire euclidien.*

Exemple 3.1 : *L'espace vectoriel $\mathbb{R}^n$ est muni de sa base canonique $\{e_1, \ldots, e_n\}$. On se donne un vecteur ω dans $\mathbb{R}^n$ et pour $x = \sum\limits_{k=1}^{n} x_k e_k$, $y = \sum\limits_{k=1}^{n} y_k e_k$ dans $\mathbb{R}^n$ on note :*

$$\langle x \mid y \rangle = \sum_{k=1}^{n} \omega_k x_k y_k.$$

L'application ainsi définie est un produit scalaire sur $\mathbb{R}^n$ si et seulement si toutes les composantes de ω sont strictement positives.

Si tous les coefficients ω_k sont égaux à 1, le produit scalaire obtenu est appelé produit scalaire canonique de $\mathbb{R}^n$.

Dans la suite de ce paragraphe E désigne un espace euclidien et on note pour tout x dans E, $\|x\| = \sqrt{\langle x|x \rangle}$.

Théorème 3.5 (inégalité de Cauchy-Schwarz) : *Pour tous x, y dans E on a :*

$$|\langle x \mid y \rangle| \leq \|x\| \, \|y\|,$$

l'égalité étant réalisée si et seulement si x et y sont liés.

Démonstration : Pour x, y fixés dans E on désigne par P la fonction polynomiale définie par :

$$P(t) = \|x + ty\|^2 = \|y\|^2 t^2 + 2 \langle x \mid y \rangle t + \|x\|^2.$$

On a alors $P(t) \geq 0$ pour tout réel t et nécessairement le discriminant de ce polynôme de degré 2 est toujours négatif ou nul, soit :

$$\langle x \mid y \rangle^2 - \|x\|^2 \|y\|^2 \leq 0,$$

ce qui équivaut à $|\langle x \mid y \rangle| \leq \|x\| \, \|y\|$.

Si x et y sont liés il est clair que l'égalité est réalisée. Réciproquement, si l'égalité est réalisée alors le polynôme P admet une racine réelle λ et $x + \lambda y = 0$, c'est-à-dire que x et y sont liés. ∎

Théorème 3.6 (inégalité de Minkowski) : *Pour tous x, y dans E on a :*

$$\|x+y\| \leq \|x\| + \|y\|,$$

l'égalité étant réalisée si et seulement si $x = \lambda y$ avec $\lambda \geq 0$ (on dit que x et y sont positivement liés).

Démonstration : On a :

$$\|x+y\|^2 = \|x\|^2 + 2\langle x \mid y \rangle + \|y\|^2$$

et avec l'inégalité de Cauchy-Schwarz :

$$\|x+y\|^2 \leq \|x\|^2 + 2\|x\|\|y\| + \|y\|^2 = (\|x\| + \|y\|)^2$$

ce qui équivaut à $\|x+y\| \leq \|x\| + \|y\|$.

L'égalité est réalisée si et seulement si $\langle x \mid y \rangle = \|x\|\|y\|$ ce qui entraîne $|\langle x \mid y \rangle| = \|x\|\|y\|$ et il existe un réel λ tel que $x = \lambda y$. Si $x = 0$ l'égalité est réalisée et si $x \neq 0$ on a $\lambda \|y\|^2 = |\lambda| \|y\|^2$ avec $\|y\| \neq 0$, donc $\lambda = |\lambda| > 0$. ∎

Corollaire 3.6 : *L'application $x \mapsto \|x\| = \sqrt{\langle x|x \rangle}$ définit une norme sur E.*

Les deux égalités qui suivent sont utiles en pratique.

Proposition 3.1 : *Pour tous x, y dans E on a :*

$$\begin{aligned} \langle x \mid y \rangle &= \frac{1}{2}\left(\|x+y\|^2 - \|x\|^2 - \|y\|^2\right) \\ &= \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right), \end{aligned}$$

$$\|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right).$$

La deuxième identité est l'égalité du parallélogramme. Elle est caractéristique des normes réelles déduites d'un produit scalaire (exercice 3.4).

Remarque 3.1 *: Pour x et y non nuls, on a :*

$$-1 \leq \frac{\langle x \mid y \rangle}{\|x\|\|y\|} \leq 1.$$

Il existe donc un unique réel θ dans $[0, \pi]$ tel que :

$$\langle x \mid y \rangle = \cos(\theta)\|x\|\|y\|.$$

On dit que θ est la mesure dans $[0, \pi]$ de l'angle que font les vecteurs x et y dans $E - \{0\}$.

Pour $\theta = \frac{\pi}{2}$, on a $\langle x \mid y \rangle = 0$ et on dit que les vecteurs x et y sont orthogonaux.

Définition 3.7 : *On dit que deux vecteurs x et y dans E sont orthogonaux si $\langle x \mid y \rangle = 0$.*

Le résultat suivant se démontre facilement.

Théorème 3.7 (Pythagore) : *Les vecteurs x et y sont orthogonaux dans E si et seulement si :*

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

Définition 3.8 : *On appelle famille orthogonale dans E toute famille $\{e_i \mid i \in I\}$ de vecteurs de E telle que $\langle e_i \mid e_j \rangle = 0$ pour tous i, j dans I tels que $i \neq j$. Si de plus $\|e_i\| = 1$ pour tout $i \in I$ on dit alors que cette famille est orthonormée ou orthonormale.*

Remarque 3.2 : *Une famille orthogonale de vecteurs non nuls de E est libre. En effet si $\{e_i \mid i \in I\}$ est une telle famille et si $\sum\limits_{j\in J} \lambda_j e_j = 0$, où J est une partie finie de I, on a alors pour tout $k \in J$:*

$$0 = \left\langle \sum_{j\in J} \lambda_j e_j \mid e_k \right\rangle = \lambda_k \|e_k\|^2,$$

avec $\|e_k\| \neq 0$ et nécessairement $\lambda_k = 0$.

Théorème 3.8 (procédé d'orthonormalisation de Gram-Schmidt) : *Pour toute famille libre $\{x_1, x_2, \ldots, x_p\}$ dans E, il existe une unique famille orthonormée $\{e_1, e_2, \ldots, e_p\} \subset E$ telle que :*

$$\forall k \in \{1,2,\ldots,p\}, \begin{cases} \mathrm{Vect}\{e_1,\ldots,e_k\} = \mathrm{Vect}\{x_1,\ldots,x_k\}, \\ \langle x_k \mid e_k \rangle > 0. \end{cases}$$

Démonstration : On procède par récurrence sur $p \geq 1$. Pour $p = 1$, on a nécessairement $e_1 = \lambda_1 x_1$ avec $\lambda_1 \in \mathbb{R}^*$ et $1 = \|e_1\|^2 = \lambda_1^2 \|x_1\|^2$, donc $\lambda_1^2 = \dfrac{1}{\|x_1\|^2}$, ce qui donne deux solutions pour λ_1. La condition supplémentaire $\langle x_1 \mid e_1 \rangle > 0$ entraîne $\lambda_1 > 0$ et on a l'unique solution $e_1 = \dfrac{1}{\|x_1\|} x_1$.

Supposons $p \geq 2$ et construite la famille orthonormée $\{e_1, e_2, \ldots, e_{p-1}\}$ vérifiant les conditions :

$$\forall k \in \{1,2,\ldots,p-1\}, \begin{cases} \mathrm{Vect}\{e_1,\ldots,e_k\} = \mathrm{Vect}\{x_1,\ldots,x_k\}, \\ \langle x_k \mid e_k \rangle > 0. \end{cases}$$

Si $\left\{e'_1, e'_2, \ldots, e'_{p-1}, e_p\right\}$ est une solution à notre problème on a alors nécessairement $e'_k = e_k$ pour tout k compris entre 1 et $p-1$ (unicité pour le cas $p-1$). Les conditions $\mathrm{Vect}\{e_1,\ldots,e_k\} = \mathrm{Vect}\{x_1,\ldots,x_k\}$ pour tout k compris entre 1 et p entraînent :

$$e_p = \sum_{j=1}^{p-1} \lambda_j e_j + \lambda_p x_p.$$

Avec les conditions d'orthogonalité :

$$\forall j \in \{1, \ldots, p-1\}, \quad \langle e_p \mid e_j \rangle = 0,$$

on déduit que :

$$\lambda_j + \lambda_p \langle x_p \mid e_j \rangle = 0 \quad (1 \leq j \leq p-1).$$

et

$$e_p = \lambda_p \left(x_p - \sum_{j=1}^{p-1} \langle x_p \mid e_j \rangle e_j \right) = \lambda_p y_p.$$

Du fait que $x_p \notin \mathrm{Vect}\{x_1, \ldots, x_{p-1}\} = \mathrm{Vect}\{e_1, \ldots, e_{p-1}\}$ on déduit que $y_p \neq 0$ et la condition $\|e_p\| = 1$ donne :

$$|\lambda_p| = \frac{1}{\|y_p\|}.$$

La condition supplémentaire :

$$0 < \langle x_p \mid e_p \rangle = \left\langle \frac{1}{\lambda_p} \left(e_p - \sum_{j=1}^{p-1} \lambda_j e_j \right) \mid e_p \right\rangle = \frac{1}{\lambda_p}$$

entraîne $\lambda_p > 0$. Ce qui donne en définitive une unique solution pour e_p. ∎

Si on désigne par $\mathcal{E} = \{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$ qui est muni de sa structure euclidienne canonique avec le produit scalaire :

$$(x, y) \mapsto \langle x \mid y \rangle = \sum_{k=1}^{n} x_k y_k,$$

alors tout système orthonormé $\mathcal{F} = \{f_1, \ldots, f_n\}$ est une base de $\mathbb{R}^n$ (ce système est libre formé de n vecteurs) et la matrice de passage Ω de la base canonique $\mathcal{E}$ à la base $\mathcal{F}$ est telle que ${}^t\Omega\Omega = I_n$. En effet les colonnes de Ω sont formées des composantes des vecteurs f_j dans la base $\mathcal{E}$ et le coefficient d'indice (i, j) de ${}^t\Omega\Omega$ est $\langle f_i \mid f_j \rangle = \delta_{ij}$.

Définition 3.9 : *On dit qu'une matrice carrée A à coefficients réels d'ordre n est orthogonale si ${}^tAA = I_n$. On note $\mathcal{O}_n(\mathbb{R})$ l'ensemble des matrices orthogonales.*

On peut également dire qu'une matrice orthogonale est la matrice de passage de la base canonique de $\mathbb{R}^n$ à une base orthonormée où $\mathbb{R}^n$ est muni de sa structure euclidienne canonique.

Remarque 3.3 : *$\mathcal{O}_n(\mathbb{R})$ est un sous-groupe multiplicatif de $GL_n(\mathbb{R})$ et l'application $A \mapsto \det(A)$ réalise un morphisme de groupe de $\mathcal{O}_n(\mathbb{R})$ sur $\{-1, 1\}$. Le noyau de ce morphisme est un sous-groupe distingué de $\mathcal{O}_n(\mathbb{R})$ appelé groupe des matrices de rotation. Il est noté $\mathcal{O}_n^+(\mathbb{R})$.*

On note $\mathcal{O}_n^-(\mathbb{R})$ le sous-ensemble de $\mathcal{O}_n(\mathbb{R})$ formé des matrices telles que $\det(A) = -1$.

Corollaire 3.7 : *Toute matrice $A \in GL_n(\mathbb{R})$ s'écrit de manière unique $A = \Omega T$, où Ω est une matrice orthogonale et T une matrice triangulaire supérieure à coefficients diagonaux strictement positifs.*

Démonstration : L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique avec le produit scalaire :

$$(x, y) \mapsto \langle x \mid y \rangle = \sum_{k=1}^{n} x_k y_k.$$

On désigne par $\mathcal{E} = \{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$ et par $C_1, \ldots, C_n$ les colonnes de la matrice A (ce sont des vecteurs de $\mathbb{R}^n$). Si $A \in GL_n(\mathbb{R})$ alors le système $\mathcal{C} = \{C_1, \ldots, C_n\}$ est une base de $\mathbb{R}^n$ et on peut lui appliquer le procédé de Gram-Schmidt. Il existe donc une base orthonormée $\mathcal{F} = \{f_1, \ldots, f_n\}$ de $\mathbb{R}^n$ telle que $\operatorname{Vect}\{f_1, \ldots, f_k\} = \operatorname{Vect}\{C_1, \ldots, C_k\}$ pour tout $k \in \{1, \ldots, n\}$, ce qui entraîne que la matrice de passage T_1 de la base $\mathcal{C}$ à la base $\mathcal{F}$ est triangulaire supérieure et la matrice de passage Ω de la base canonique $\mathcal{E}$ à la base $\mathcal{F}$ est orthogonale.

En considérant que A est la matrice de passage de la base canonique $\mathcal{E}$ à la base $\mathcal{C}$, on a (relation de Chasles pour les matrices de passage) :

$$A = P_{\mathcal{E},\mathcal{C}} = P_{\mathcal{E},\mathcal{F}} P_{\mathcal{F},\mathcal{C}} = \Omega T_1^{-1} = \Omega T,$$

avec Ω orthogonale et T triangulaire supérieure.

En considérant que les termes diagonaux de T sont les :

$$t_{jj} = \langle C_j \mid f_j \rangle \quad (1 \leq j \leq n),$$

on déduit que T est à termes diagonaux strictement positifs si on a de plus les conditions $\langle C_j \mid f_j \rangle > 0$ pour tout $j \in \{1, \ldots, n\}$.

S'il existe deux décompositions $A = \Omega T$ et $A = \Omega' T'$ avec Ω, Ω' orthogonales et T, T' triangulaires supérieures alors la matrice $\Delta = T' T^{-1} = {}^t\Omega' \Omega$ est triangulaire supérieure orthogonale et $\Delta^{-1} = {}^t\Delta$, à la fois triangulaire supérieure et inférieure, est donc diagonale et orthogonale. Les termes diagonaux de Δ sont donc égaux à ± 1. Si on suppose de plus que T et T' sont à termes diagonaux strictement positifs il en est de même de Δ et nécessairement $\Delta = I_n$, ce qui donne $T = T'$ et $\Omega = \Omega'$. D'où l'unicité de la décomposition. ■

Remarque 3.4 *: Cette décomposition est parfois notée $A = QR$ avec Q orthogonale et R triangulaire supérieure de termes diagonaux strictement positifs et appelée décomposition QR de la matrice A. Cette décomposition conduit à un algorithme de calcul des valeurs propres d'une matrice symétrique réelle moyennant certaines conditions.*

Théorème 3.9 : *Dans un espace vectoriel E euclidien de dimension n tout système orthonormé peut se prolonger en une base orthonormée.*

Démonstration : Soit $\{e_1, \ldots, e_p\}$ un système orthonormé dans E. Si $p = n$ c'est une base de E. On suppose donc que $1 \leq p < n$. Ce système étant libre dans E, il se prolonge donc en une base $\{e_1, \ldots, e_p, x_{p+1}, \ldots, x_n\}$ de E. Le procédé d'orthonormalisation de Gram-Schmidt fournit, vu l'unicité, le système orthonormé $\{e_1, \ldots, e_p, e_{p+1}, \ldots, e_n\}$ qui est une base de E (c'est un système libre à n éléments). ■

Remarque 3.5 *: Si $\{e_1, \ldots, e_n\}$ est une base orthonormée de E alors tout vecteur x de E s'écrit $x = \sum\limits_{k=1}^{n} \lambda_k e_k$ avec $\lambda_k = \langle x \mid e_k \rangle$ pour tout k compris entre 1 et n et $\|x\|^2 = \sum\limits_{k=1}^{n} |\lambda_k|^2$.*

Corollaire 3.8 (Schur) : *Soit A une matrice réelle d'ordre $n > 0$ dont toutes les valeurs propres sont réelles. Il existe alors une matrice orthogonale Ω telle que ${}^t\Omega A \Omega$ soit triangulaire supérieure. C'est-à-dire que la matrice A se trigonalise dans une base orthonormée.*

Démonstration : On procède par récurrence sur $n \geq 1$. Pour $n = 1$ le résultat est évident. Supposons-le acquis pour $n - 1 \geq 1$ et soit A une matrice réelle d'ordre n dont toutes les valeurs propres sont réelles. On note u l'application linéaire ayant A pour matrice dans la base canonique de $\mathbb{R}^n$. Si λ_1 est une valeur propre de u et e_1 un vecteur propre associé unitaire, on complète $\{e_1\}$ en une base orthonormée $\mathcal{B}_1 = \{e_1, \ldots, e_n\}$ de $\mathbb{R}^n$. La matrice de passage Ω_1 de la base canonique à cette base est orthogonale et la matrice de u dans $\mathcal{B}_1$ s'écrit :

$$A_1 = {}^t\Omega_1 A \Omega_1 = \begin{pmatrix} \lambda_1 & a_1 \\ 0 & B_1 \end{pmatrix},$$

avec $a_1 \in \mathcal{M}_{1,n-1}(\mathbb{K})$ (ensemble des matrices à une ligne et $n-1$ colonnes) et $B_1 \in \mathcal{M}_{n-1}(\mathbb{K})$.

Le polynôme caractéristique de A (égal à celui de A_1) est alors donné par :

$$P_A(\lambda) = (\lambda - \lambda_1) \det (B_1 - \lambda I_{n-1}).$$

On en déduit alors que la matrice B_1 a toutes ses valeurs propres réelles. On peut donc utiliser l'hypothèse de récurrence pour écrire qu'il existe une matrice orthogonale d'ordre $n-1$ Ω_2' telle que ${}^t\Omega_2' B_1 \Omega_2' = T_2$ soit triangulaire supérieure.

La matrice $\Omega_2 = \begin{pmatrix} 1 & 0 \\ 0 & \Omega_2' \end{pmatrix}$ est alors orthogonale d'ordre n et :

$${}^t\Omega_2 \, {}^t\Omega_1 A \Omega_1 \Omega_2 = \begin{pmatrix} \lambda_1 & a_1 \\ 0 & {}^t\Omega_2' B_1 \Omega_2' \end{pmatrix} = \begin{pmatrix} \lambda_1 & a_1 \\ 0 & T_2 \end{pmatrix}$$

est triangulaire supérieure. ■

4. Réduction des matrices orthogonales

Pour ce paragraphe, on désigne par E l'espace vectoriel $\mathbb{R}^n$, par $\{e_1, e_2, \ldots, e_n\}$ la base canonique de E et le produit scalaire euclidien canonique de E est noté :

$$(x, y) = \left(\sum_{i=1}^{n} x_i e_i, \sum_{i=1}^{n} y_i e_i \right) \longmapsto \langle x \mid y \rangle = \sum_{i=1}^{n} x_i y_i.$$

La norme associée est la norme euclidienne notée $x \longmapsto \|x\|_2$.

Un endomorphisme $u \in \mathcal{L}(E)$ est dit orthogonal si :

$$\forall (x, y) \in E^2, \quad \langle u(x) \mid u(y) \rangle = \langle x \mid y \rangle.$$

La matrice d'un endomorphisme orthogonal dans une base orthonormée de E est une matrice orthogonale.

On note $\mathcal{O}(E)$ l'ensemble des endomorphismes orthogonaux de E.

Lemme 3.2 : *Soit $u \in \mathcal{L}(\mathbb{R}^n)$. Il existe un sous-espace vectoriel P de $\mathbb{R}^n$ de dimension égale à 1 ou 2 et stable par u.*

Démonstration : Voir l'exercice 2.2. ■

Lemme 3.3 : *Soient E un espace vectoriel euclidien et $u \in \mathcal{O}(E)$ un endomorphisme orthogonal. Il existe des sous-espaces vectoriels de E, $P_1, \ldots, P_r$, de dimension égale à 1 ou 2, deux à deux orthogonaux et stables par u, tels que $E = \overset{r}{\underset{j=1}{\oplus}} P_j$.*

Démonstration : On procède par récurrence sur la dimension de E, $n > 1$.

Pour $n = 2$, le résultat est évident. Supposons-le acquis pour tout endomorphisme orthogonal sur un espace vectoriel euclidien de dimension p comprise entre 1 et $n - 1$, avec $n > 2$.

Si P_1 est un sous-espace vectoriel de E non réduit à $\{0\}$ de dimension au plus égale à 2 stable par u, endomorphisme orthogonal, alors $P_1^{\perp}$ est stable par u. En effet $u(P_1) \subset P_1$ et $u \in GL(E)$ entraînent $u(P_1) = P_1$ (un isomorphisme conserve la dimension), donc tout $y \in P_1$ s'écrit $y = u(x)$ avec $x \in P_1$ et pour tout $z \in P_1^{\perp}$, on a :

$$\langle u(z) \mid y \rangle = \langle u(z) \mid u(x) \rangle = \langle z \mid x \rangle = 0,$$

c'est-à-dire que $u(z) \in P_1^{\perp}$.

Comme $0 < \dim(P_1^{\perp}) < \dim(E)$, on peut trouver des sous-espaces vectoriels de E, $P_2, \ldots, P_r$, de dimension au plus égale à 2, deux à deux orthogonaux et stables par la restriction de u à $P_1^{\perp}$, donc par u, tels que $P_1^{\perp} = \overset{r}{\underset{j=2}{\oplus}} P_j$. On a alors

$E = P_1 \oplus P_1^{\perp} = \overset{r}{\underset{j=1}{\oplus}} P_j$. ■

Théorème 3.10 : *Soit A une matrice orthogonale d'ordre $n \geq 2$. Il existe une matrice orthogonale P telle que :*

$$P^{-1}AP = \begin{pmatrix} I_p & 0 & 0 & 0 & \dots & 0 \\ 0 & -I_q & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & R_1 & 0 & \ddots & 0 \\ 0 & \ddots & 0 & R_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 0 & R_r \end{pmatrix},$$

où, pour tout $k \in \{1, \dots, r\}$, on a noté :

$$R_k = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix}$$

avec $\theta_k \in]0, 2\pi[- \{\pi\}$.

Démonstration : On procède par récurrence sur $n \geq 2$.

Pour $n = 2$, dire que $A = \begin{pmatrix} a & c \\ b & d \end{pmatrix}$ est orthogonale équivaut à :

$$\begin{cases} a^2 + b^2 = c^2 + d^2 = 1, \\ ac + bd = 0, \end{cases} \quad \begin{cases} a^2 + c^2 = b^2 + d^2 = 1, \\ ab + cd = 0. \end{cases}$$

On déduit alors que $b^2 = 1 - a^2 = c^2$, $a^2 = 1 - b^2 = d^2$, donc $c = \epsilon b$, $d = \eta a$, avec ϵ, η dans $\{-1, 1\}$ tels que $\epsilon\eta = -1$. Ce qui entraîne $A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ ou $A = \begin{pmatrix} a & b \\ b & -a \end{pmatrix}$ avec $a^2 + b^2 = 1$. Réciproquement, les matrices de l'une de ces deux formes sont orthogonales.

Dans le premier cas on peut poser $a = \cos(\theta)$ avec $\theta \in [0, 2\pi[$ et $A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$. Dans le deuxième cas la matrice A est symétrique de valeurs propres 1 et -1, est donc orthogonalement semblable à $D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$.

Supposons le résultat acquis pour toute matrice orthogonale d'ordre p compris entre 2 et $n - 1$ et soit A une matrice orthogonale d'ordre $n > 2$.

On désigne par u l'endomorphisme orthogonal ayant A pour matrice dans la base canonique de $E = \mathbb{R}^n$ muni de sa structure euclidienne canonique.

Si u admet 1 ou -1 comme valeur propre, pour tout vecteur propre x associé à cette valeur propre, le sous-espace vectoriel $(\mathbb{R}x)^\perp$ est stable par u. En effet, pour tout y dans $(\mathbb{R}x)^\perp$, on a :

$$\langle u(y) \mid x \rangle = \pm \langle u(y) \mid u(x) \rangle = \pm \langle y \mid x \rangle = 0$$

et il existe une base orthonormée $\mathcal{B}$ de $(\mathbb{R}x)^{\perp}$ dans laquelle la matrice de la restriction de u à $(\mathbb{R}x)^{\perp}$ est de la forme :

$$A' = \begin{pmatrix} I_p & 0 & 0 & 0 & \dots & 0 \\ 0 & -I_q & 0 & \ddots & \ddots & \vdots \\ 0 & 0 & R_1 & 0 & \ddots & 0 \\ 0 & \ddots & 0 & R_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & 0 & R_r \end{pmatrix}.$$

Dans la base orthonormée $\left\{\frac{1}{\|x\|_2}x\right\} \cup \mathcal{B}$ la matrice de u est $A'' = \begin{pmatrix} \pm 1 & 0 \\ 0 & A' \end{pmatrix}$, qui est bien de la forme souhaitée.

Si toutes les valeurs propres de u sont complexes non réelles, alors dans une décomposition $E = \overset{r}{\underset{k=1}{\oplus}} P_k$ du type lemme 3.3 tous les sous-espaces vectoriels P_k sont de dimension 2 et la restriction de u à chacun des P_k a ses valeurs propres complexes non réelles. D'après l'étude du cas $n = 2$ (les P_k sont stables par u), on déduit alors qu'il existe une base orthonormée de P_k dans laquelle la matrice de u est de la forme :

$$R_k = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix},$$

avec $\theta_k \in \left]0, 2\pi\right[- \{\pi\}$. En réunissant toutes ces bases, on obtient une base orthonormée de E dans laquelle la matrice de u est

$$A' = \begin{pmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R_r \end{pmatrix}.$$

■

Remarque 3.6 : *On a* $p = \dim(\ker(A - I_n))$ *et* $q = \dim(\ker(A + I_n))$ *avec* $p + q + 2r = n$. *De plus* $A \in \mathcal{O}_n^+(\mathbb{R})$ *[resp.* $A \in \mathcal{O}_n^-(\mathbb{R})$*] si et seulement si* q *est pair [resp. impair].*

5. Réduction des matrices symétriques réelles

On garde les notations du paragraphe 4, page 89.

Définition 3.10 : *Une matrice réelle* A *est dite symétrique si* ${}^tA = A$. *On dit qu'elle est symétrique positive si on a de plus* $\langle Ax \mid x \rangle \geq 0$ *pour tout* $x \in E$. *On dit qu'elle est symétrique définie positive si on a de plus* $\langle Ax \mid x \rangle > 0$ *pour tout* $x \in E - \{0\}$.

On peut associer une forme quadratique à une matrice symétrique A en posant :

$$\forall x \in E, \quad q(x) = \langle Ax \mid x \rangle .$$

La forme polaire associée est alors définie par :

$$\forall x, y \in E, \quad \varphi(x,y) = \frac{1}{2}\big(q(x+y) - q(x) - q(y)\big) = \langle Ax \mid y \rangle .$$

Si A est positive, on alors l'inégalité de Cauchy-Schwarz :

$$\forall x, y \in E^2, \quad |\varphi(x,y)| \leq q(x)\, q(y) .$$

Le cône isotrope de q est définie par :

$$q^{-1}\{0\} = \{x \in E \mid q(x) = 0\} .$$

Pour A positive, on a $q^{-1}\{0\} = \ker(A)$. En effet, si $q(x) = 0$, on a alors :

$$\|Ax\|_2^2 = \langle Ax \mid Ax \rangle = \langle A(Ax) \mid x \rangle = \varphi(Ax, x) \leq q(Ax)\, q(x) = 0$$

et $Ax = 0$. La réciproque est claire.

Théorème 3.11 : *Soit A une matrice symétrique réelle. Les valeurs propres de A sont toutes réelles et A se diagonalise dans une base orthonormée, c'est-à-dire qu'il existe une matrice orthogonale P et une matrice diagonale D telles que $A = PD\,{}^tP$, les coefficients diagonaux de D étant les valeurs propres dans $\mathbb{R}$ de la matrice A.*

Démonstration : Soit $q : x \longmapsto q(x) = \langle Ax \mid x \rangle$ la forme quadratique associée à la matrice A. Elle définie une application continue de E dans E et elle atteint son maximum sur la sphère unité de E (qui est compacte puisque E est de dimension finie). On peut donc poser :

$$\lambda = \sup\{q(x) \mid x \in E, \quad \|x\| = 1\}$$

et on a $\lambda = q(x_0)$ avec $\|x_0\|_2 = 1$.

On en déduit alors que la forme quadratique

$$q_1 : x \longmapsto q_1(x) = \lambda \|x\|_2^2 - q(x) = \langle \lambda x - Ax | x \rangle$$

est positive.

Le vecteur x_0 qui est dans le cône isotrope de q_1 est alors dans le noyau de $\lambda I_n - A$. On a donc ainsi montré que λ est une valeur propre réelle de A et x_0 un vecteur propre réel associé.

Une récurrence sur la dimension $n \geq 1$ de E permet alors de montrer que A se diagonalise dans une base orthonormée. En effet, du fait que :

$$\langle Ax \mid x_0 \rangle = \langle x \mid Ax_0 \rangle = \lambda \langle x \mid x_0 \rangle ,$$

on déduit que l'hyperplan $H = (\mathbb{R}x_0)^\perp$ est stable par A et que, dans une base adaptée à la décomposition $E = \mathbb{R}x_0 \oplus H$, la matrice de l'application linéaire associée à la matrice A dans la base canonique est de la forme $A' = \begin{pmatrix} \lambda & 0 \\ 0 & B \end{pmatrix}$ avec B symétrique réelle d'ordre $n-1$. On applique alors l'hypothèse de récurrence à la matrice B pour conclure. ■

Remarque 3.7 : *La démonstration de ce théorème de diagonalisation des matrices symétriques réelles est une adaptation au cas de la dimension finie d'un résultat plus général de diagonalisation d'un opérateur autoadjoint compact sur un espace de Hilbert. Précisément, si u est un endomorphisme autoadjoint compact sur un espace de Hilbert E alors u a toutes ses valeurs propres réelles et E est somme hilbertienne des espaces propres de u (voir ref. [6], p. 125).*

Avec l'exercice 3.7 on propose une autre démonstration du théorème précédent.

Corollaire 3.9 : *Une matrice symétrique réelle est définie positive [resp. positive] si et seulement si toutes ses valeurs propres sont strictement positives [resp. positives].*

Une autre caractérisation des matrices réelles symétriques définies positives est donnée avec l'exercice 3.6, page 111.

Théorème 3.12 : *Une matrice symétrique réelle à diagonale strictement dominante $A = \left((a_{ij})\right)_{1\leq i,j\leq n}$ est définie positive si et seulement si $a_{ii} > 0$ pour tout i compris entre 1 et n.*

Démonstration : Supposons que $a_{ii} > 0$, pour tout i compris entre 1 et n. On sait déjà que les valeurs propres de A sont réelles. En reprenant la démonstration du théorème 2.3, on a :

$$|\lambda - a_{ii}| \leq \sum_{j\neq i} |a_{ij}| < a_{ii}$$

et nécessairement $\lambda > 0$. La matrice A a donc toutes ses valeurs propres strictement positives, ce qui équivaut à dire qu'elle est définie positive.

La réciproque provient de l'égalité $a_{ii} = \langle Ae_i \mid e_i \rangle$, où $\{e_1, \ldots, e_n\}$ désigne la base canonique de $\mathbb{R}^n$. ∎

6. Tridiagonalisation des matrices symétriques réelles. Méthode de Householder

On a vu que toute matrice symétrique réelle est orthogonalement semblable à une matrice diagonale, donc tridiagonale. L'avantage de la méthode de tridiagonalisation due à Householder, décrite dans ce paragraphe, est qu'elle fournit un procédé algorithmique simple de réduction à la forme tridiagonale d'une matrice symétrique réelle sans avoir à calculer ses valeurs propres. Cette réduction est utilisée pour calculer des valeurs approchées des valeurs propres d'une matrice symétrique réelle (méthode de Givens-Householder, paragraphe 4, page 248).

Pour $n \geq 2$, l'espace vectoriel $\mathbb{R}^n$ est muni du produit scalaire euclidien canonique $\langle \cdot \mid \cdot \rangle$ et de la norme euclidienne associée $\|\cdot\|$. On désigne par $\{e_1, e_2, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$. Pour tout réel α, le signe de α est défini par :

$$\operatorname{signe}(\alpha) = \begin{cases} 1 & \text{si } \alpha > 0, \\ 0 & \text{si } \alpha = 0, \\ -1 & \text{si } \alpha < 0. \end{cases}$$

Définition 3.11 : *On appelle matrice de Householder toute matrice réelle de la forme :*

$$P_u = I_n - 2u\,{}^t u,$$

où u est un vecteur unitaire dans $\mathbb{R}^n$.

Si, pour tout vecteur unitaire $u \in \mathbb{R}^n$, on note P_u la matrice de Householder qui lui est associée et H_u l'hyperplan orthogonal à u dans $\mathbb{R}^n$, on a alors :

$$\forall x \in H_u, \quad P_u x = x - 2\left(u\,{}^t u\right) x = x - 2\langle u \mid x\rangle u = x$$

et, avec $P_u u = -u$, on déduit que P_u est la symétrie orthogonale par rapport à l'hyperplan H_u. En particulier l'inverse de P_u est donné par $P_u^{-1} = {}^t P_u = P_u$.

Lemme 3.4 : *Soit M une matrice symétrique réelle d'ordre $n \geq 2$. En utilisant les notations qui précèdent, on a :*

$$P_u^{-1} M P_u = M - 2\left(v\,{}^t u + u\,{}^t v\right)$$

où $v = Mu - \langle u \mid Mu\rangle u$.

Démonstration : On a :

$$P_u^{-1} M P_u = P_u M P_u = M - 2u\,{}^t u M - 2Mu\,{}^t u + 4u\,{}^t u M u\,{}^t u,$$

avec :

$$u\,{}^t u M - u\left({}^t u M u\right)^t u = u^t\,(Mu) - u\,\langle u \mid Mu\rangle^t\, u = u\,{}^t v$$

et

$$Mu\,{}^t u - u\left({}^t u M u\right)^t u = \left(Mu - \langle u \mid Mu\rangle\, u\right)^t u = v\,{}^t u,$$

d'où le résultat. ■

Dans ce qui suit, on se donne une matrice symétrique réelle $A = \left((a_{ij})\right)_{1\leq i,j\leq n}$ d'ordre $n \geq 3$.

On note :

$$x = \begin{pmatrix} a_{11} \\ \vdots \\ a_{n1} \end{pmatrix},$$

la première colonne de A et :

$$\begin{cases} x^{(2)} = x - a_{11} e_1 = \begin{pmatrix} 0 \\ a_{21} \\ \vdots \\ a_{n1} \end{pmatrix}, \\ y^{(2)} = x^{(2)} + \operatorname{signe}(a_{21}) \left\| x^{(2)} \right\| e_2, \\ P_1 = \begin{cases} I_n & \text{si } y^{(2)} = 0, \\ P_{u^{(2)}} & \text{si } y^{(2)} \neq 0, \end{cases} \end{cases}$$

où $u^{(2)} = \dfrac{1}{\left\| y^{(2)} \right\|} y^{(2)}$ pour $y^{(2)} \neq 0$. Enfin on note $A_1 = P_1^{-1} A P_1 = \left(\left(a_{ij}^1\right)\right)_{1\leq i,j\leq n}$.

Lemme 3.5 : *Avec les notations qui précèdent, la matrice A_1 est symétrique avec $a_{i1}^1 = 0$ pour $i = 3, 4, \ldots, n$.*

Démonstration : Si $y^{(2)} = 0$, alors $a_{i1} = 0$ pour tout $i = 3, 4, \ldots, n$ et $A_1 = A$ vérifie bien la condition voulue. On suppose donc que $y^{(2)} \neq 0$.

On a $A_1 = P_1 A P_1$ et A_1 est symétrique puisque P_1 et A le sont.

La première colonne de A_1 est donnée par $A_1 e_1 = P_1 A P_1 e_1$. Avec $e_1 \in \left(\mathbb{R}y^{(2)}\right)^{\perp}$, on a $P_1 e_1 = e_1$ et $A_1 e_1 = P_1 A e_1 = P_1 x$.

Si on note $y^{(2)} = x^{(2)} + z^{(2)}$, avec $z^{(2)} = \text{signe}\,(a_{21}) \left\|x^{(2)}\right\| e_2$, on a :

$$\left\langle y^{(2)} \mid x^{(2)} - z^{(2)} \right\rangle = \left\langle x^{(2)} + z^{(2)} \mid x^{(2)} - z^{(2)} \right\rangle = \left\|x^{(2)}\right\|^2 - \left\|x^{(2)}\right\|^2 = 0,$$

donc :

$$\begin{cases} P_1 \left(x^{(2)} + z^{(2)}\right) = -x^{(2)} - z^{(2)}, \\ P_1 \left(x^{(2)} - z^{(2)}\right) = x^{(2)} - z^{(2)}, \end{cases}$$

et $P_1 x^{(2)} = -z^{(2)} = -\,\text{signe}\,(a_{21}) \left\|x^{(2)}\right\| e_2$. Ce qui donne en définitive :

$$P_1 x = P_1 \left(a_{11} e_1 + x^{(2)}\right) = a_{11} e_1 - \text{signe}\,(a_{21}) \left\|x^{(2)}\right\| e_2.$$

C'est-à-dire que :

$$a_{i1}^1 = \begin{cases} a_{11} & \text{si } i = 1, \\ -\,\text{signe}\,(a_{21}) \left\|x^{(2)}\right\| & \text{si } i = 2, \\ 0 & \text{si } i = 3, \ldots, n. \end{cases}$$ ■

Théorème 3.13 : *Soit A une matrice symétrique réelle d'ordre $n \geq 3$. Il existe $n-2$ matrices orthogonales $P_1, P_2, \ldots, P_{n-2}$ telles que :*

$$T = P_{n-2}^{-1} P_{n-1}^{-1} \cdots P_1^{-1} A P_1 P_2 \cdots P_{n-2}$$

soit tridiagonale.

Démonstration : On raisonne par récurrence sur $n \geq 3$. Pour $n = 3$, la matrice A est déjà tridiagonale. On suppose donc le résultat vrai pour les matrices symétriques d'ordre $n - 1 \geq 3$.

Avec les notations du lemme 3.5, on a :

$$A_1 = \begin{pmatrix} a_{11} & a_{12}^1 & 0 & \cdots & 0 \\ a_{21}^1 & a_{22}^1 & a_{23}^1 & \cdots & a_{2n}^1 \\ 0 & a_{32}^1 & a_{33}^1 & \cdots & a_{3n}^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^1 & a_{n3}^1 & \cdots & a_{nn}^1 \end{pmatrix},$$

cette matrice étant symétrique ainsi que la matrice d'ordre $n-1$:

$$B = \begin{pmatrix} a_{22}^1 & \dots & a_{2n}^1 \\ \vdots & \ddots & \vdots \\ a_{n2}^1 & \dots & a_{nn}^1 \end{pmatrix}.$$

Avec l'hypothèse de récurrence, on peut trouver des matrices orthogonales Q_2, $Q_3, \dots, Q_{n-2}$ telles que $T_1 = Q_{n-2}^{-1} \dots Q_2^{-1} B Q_2 \dots Q_{n-2}$ soit tridiagonale. On pose alors :

$$P_j = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & Q_j & \\ 0 & & & \end{pmatrix},$$

pour $j = 2, 3, \dots, n-2$ et :

$$T = \begin{pmatrix} a_{11} & a_{12}^1 & 0 & \dots & 0 \\ a_{21}^1 & & & & \\ 0 & & T_1 & & \\ \vdots & & & & \\ 0 & & & & \end{pmatrix}.$$

On a alors :

$$T = P_{n-2}^{-1} \dots P_2^{-1} A_1 P_2 \dots P_{n-2} = P_{n-2}^{-1} \dots P_2^{-1} P_1^{-1} A P_1 P_2 \dots P_{n-2}.$$ ■

7. Espaces vectoriels hermitiens

Pour ce paragraphe, E est un espace vectoriel sur $\mathbb{C}$ de dimension $n \geq 1$.

Définition 3.12 : *On appelle produit scalaire hermitien sur E toute application*

$$\varphi : \begin{array}{ccc} E \times E & \to & \mathbb{C} \\ (x, y) & \mapsto & \varphi(x, y) \end{array}$$

qui vérifie les propriétés suivantes :

- *pour tout y dans E l'application $x \mapsto \varphi(x, y)$ est linéaire ;*
- *$\varphi(y, x) = \overline{\varphi(x, y)}$ pour tous x, y dans E ;*
- *$\varphi(x, x) \geq 0$ pour tout x dans E ;*
- *pour x dans E, $\varphi(x, x) = 0$ équivaut à $x = 0$.*

On note en général :

$$(x,y) \longmapsto \langle x \mid y\rangle$$

un tel produit scalaire.

Définition 3.13 : *Un espace hermitien est un $\mathbb{C}$-espace vectoriel de dimension finie muni d'un produit scalaire hermitien.*

Dans la suite de ce paragraphe E désigne un espace hermitien et on note, pour tout x dans E, $\|x\| = \sqrt{\langle x|x\rangle}$.

Théorème 3.14 (inégalité de Cauchy-Schwarz) : *Pour tous x, y dans E on a :*

$$|\langle x \mid y\rangle| \leq \|x\| \, \|y\| \, ,$$

l'égalité étant réalisée si et seulement si x et y sont liés.

Démonstration : Pour x, y fixés dans E on pose :

$$\langle x \mid y\rangle = \rho e^{i\theta}$$

avec $\rho = |\langle x \mid y\rangle| \in \mathbb{R}^+$ et $\theta \in [0, 2\pi[$. On désigne par P la fonction polynomiale définie par :

$$\forall t \in \mathbb{R}, \quad P(t) = \left\|x + te^{i\theta}y\right\|^2 = \|y\|^2 t^2 + 2t\,|\langle x \mid y\rangle| + \|x\|^2 .$$

On a alors $P(t) \geq 0$ pour tout réel t et nécessairement le discriminant de ce polynôme de degré 2 est toujours négatif ou nul, soit :

$$|\langle x \mid y\rangle|^2 - \|x\|^2 \|y\|^2 \leq 0,$$

ce qui équivaut à $|\langle x \mid y\rangle| \leq \|x\| \, \|y\|$.

Si x et y sont liés il est clair que l'égalité est réalisée. Réciproquement, si l'égalité est réalisée alors le polynôme P admet une racine réelle λ et $x + \lambda e^{i\theta} y = 0$, c'est-à-dire que x et y sont liés. ■

Théorème 3.15 (inégalité de Minkowski) : *Pour tous x, y dans E on a :*

$$\|x + y\| \leq \|x\| + \|y\| \, ,$$

l'égalité étant réalisée si et seulement si $x = \lambda y$ avec $\lambda \geq 0$ (on dit que x et y sont positivement liés).

Démonstration : On a :

$$\|x + y\|^2 = \|x\|^2 + 2\mathcal{R}\langle x \mid y\rangle + \|y\|^2$$

et avec l'inégalité de Cauchy-Schwarz :

$$\|x + y\|^2 \leq \|x\|^2 + 2\,\|x\| \, \|y\| + \|y\|^2 = \left(\|x\| + \|y\|\right)^2,$$

ce qui équivaut à $\|x + y\| \leq \|x\| + \|y\|$.

L'égalité est réalisée si et seulement si :

$$\mathcal{R}\langle x \mid y\rangle = |\langle x \mid y\rangle| = \|x\| \, \|y\| ,$$

ce qui entraîne $|\langle x \mid y\rangle| = \|x\| \, \|y\|$ et il existe un nombre complexe λ tel que $x = \lambda y$. Si $x = 0$ l'égalité est réalisée et si $x \neq 0$ on a :

$$\mathcal{R}\langle x \mid y\rangle = \mathcal{R}\left(\lambda \|y\|^2\right) = |\lambda| \, \|y\|^2$$

avec $\|y\| \neq 0$, donc $\mathcal{R}(\lambda) = |\lambda|$ et λ est un réel positif. ■

Corollaire 3.10 : *L'application* $x \mapsto \|x\| = \sqrt{\langle x|x\rangle}$ *définit une norme sur* E.

Comme dans le cas euclidien, on dit que deux vecteurs x et y dans E sont orthogonaux si $\langle x \mid y\rangle = 0$.

Le théorème de Pythagore est encore vrai sous la forme : si les vecteurs x et y sont orthogonaux dans E alors :

$$\|x+y\|^2 = \|x\|^2 + \|y\|^2 .$$

La réciproque étant fausse dans le cas hermitien (l'égalité $\|x+y\|^2 = \|x\|^2 + \|y\|^2$ entraîne seulement $\mathcal{R}\langle x \mid y\rangle = 0$).

On définit, comme dans le cas euclidien, les notions de familles orthogonale et orthonormale.

Une famille orthogonale de vecteurs non nuls de E est libre.

Théorème 3.16 (procédé d'orthonormalisation de Gram-Schmidt) : *Pour toute famille libre* $\{x_1, x_2, \ldots, x_p\}$ *dans* E, *il existe une unique famille orthonormée* $\{e_1, e_2, \ldots, e_p\} \subset E$ *telle que :*

$$\forall k \in \{1, 2, \ldots, p\}, \begin{cases} \operatorname{Vect}\{e_1, \ldots, e_k\} = \operatorname{Vect}\{x_1, \ldots, x_k\}, \\ \langle x_k \mid e_k\rangle > 0. \end{cases}$$

Démonstration : La démonstration est analogue à celle du théorème 3.8. ■

Si on désigne par $\mathcal{E} = \{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{C}^n$ qui est muni de sa structure hermitienne canonique avec le produit scalaire :

$$(x, y) \mapsto \langle x \mid y\rangle = \sum_{k=1}^{n} x_k \overline{y_k},$$

alors tout système orthonormé $\mathcal{F} = \{f_1, \ldots, f_n\}$ est une base de $\mathbb{C}^n$ (ce système est libre formé de n vecteurs) et la matrice de passage U de la base canonique $\mathcal{E}$ à la base $\mathcal{F}$ est telle que $U^*U = I_n$, où $U^* = \left((u^*_{ij})\right)_{1\leq i,j\leq n}$ désigne la matrice adjointe de la matrice complexe $U = ((u_{ij}))_{1\leq i,j\leq n}$, définie par :

$$u^*_{ij} = \overline{u_{ji}} \quad (1 \leq i, j \leq n).$$

En effet les colonnes de U sont formées des composantes des vecteurs f_j dans la base $\mathcal{E}$ et le coefficient d'indice (i, j) de U^*U est $\langle f_j \mid f_i\rangle = \delta_{ij}$.

Définition 3.14 : *On dit qu'une matrice carrée* A *à coefficients complexes d'ordre* n *est unitaire si* $U^*U = I_n$. *On note* $\mathcal{U}_n(\mathbb{C})$ *l'ensemble des matrices unitaires.*

On peut également dire qu'une matrice unitaire est la matrice de passage de la base canonique de $\mathbb{C}^n$ à une base orthonormée où $\mathbb{C}^n$ est muni de sa structure hermitienne canonique.

Remarque 3.8 : *$\mathcal{U}_n(\mathbb{C})$ est un sous-groupe multiplicatif de $GL_n(\mathbb{C})$ et l'application $A \mapsto \det(A)$ réalise un morphisme de groupe de $\mathcal{U}_n(\mathbb{C})$ sur le groupe multiplicatif des nombres complexes de module 1.*

Corollaire 3.11 : *Toute matrice $A \in GL_n(\mathbb{C})$ s'écrit de manière unique $A = UT$ où U est une matrice unitaire et T une matrice triangulaire supérieure à coefficients diagonaux strictement positifs.*

Démonstration : La démonstration est analogue à celle du corollaire 3.7. ■

Théorème 3.17 : *Dans un espace vectoriel E hermitien de dimension n tout système orthonormé peut se prolonger en une base orthonormée.*

Démonstration : La démonstration est analogue à celle du théorème 3.9. ■

Corollaire 3.12 (Schur) : *Soit A une matrice complexe d'ordre $n > 0$. Il existe alors une matrice unitaire U telle que U^*AU soit triangulaire supérieure. C'est-à-dire que la matrice A se trigonalise dans une base orthonormée.*

Démonstration : La démonstration est analogue à celle du corollaire 3.8. ■

8. Réduction des matrices normales

On note $\{e_1, e_2, \ldots, e_n\}$ la base canonique de $\mathbb{C}^n$ et le produit scalaire hermitien canonique de $\mathbb{C}^n$ est noté :

$$(x, y) = \left(\sum_{i=1}^{n} x_i e_i, \sum_{i=1}^{n} y_i e_i\right) \longmapsto \langle x \mid y \rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$

La norme associée est la norme hermitienne notée $x \longmapsto \|x\|_2$.

Définition 3.15 : *Une matrice complexe A est dite hermitienne si $A^* = A$. On dit qu'elle est hermitienne positive si on a de plus $\langle Ax \mid x \rangle \geq 0$ pour tout $x \in E$. On dit qu'elle est hermitienne définie positive si on a de plus $\langle Ax \mid x \rangle > 0$ pour tout $x \in E - \{0\}$.*

Définition 3.16 : *Une matrice complexe A est dite normale si $A^*A = AA^*$.*

Les matrices hermitiennes et unitaires sont des cas particuliers de matrices normales.

Lemme 3.6 : *Les valeurs propres d'une matrice hermitienne [resp. hermitienne positive] sont réelles [resp. réelles et positives].*

Les valeurs propres d'une matrice unitaire sont de module égal à 1.

Démonstration : Soient $A \in \mathcal{M}_n(\mathbb{C})$ hermitienne, $\lambda \in \mathbb{C}$ une valeur propre de A et $x \in \mathbb{C}^n$ un vecteur propre non nul associé. On a :

$$\lambda \|x\|_2^2 = \langle Ax \mid x \rangle = \langle x \mid A^*x \rangle = \langle x \mid Ax \rangle = \bar{\lambda} \|x\|_2^2,$$

donc $\lambda = \bar{\lambda}$ et $\lambda \in \mathbb{R}$. Si de plus A est positive alors $\lambda \|x\|_2^2 = \langle Ax \mid x \rangle \geq 0$ avec $x \neq 0$ entraîne $\lambda \geq 0$.

Soient $A \in \mathcal{M}_n(\mathbb{C})$ unitaire, $\lambda \in \mathbb{C}$ une valeur propre de A et $x \in \mathbb{C}^n - \{0\}$ un vecteur propre associé. On a :

$$|\lambda|^2 \|x\|_2^2 = \|Ax\|_2^2 = \langle Ax \mid Ax \rangle = \langle x \mid A^*Ax \rangle = \|x\|_2^2$$

et nécessairement $|\lambda| = 1$. ■

Théorème 3.18 : *Si A est une matrice complexe normale, alors elle se diagonalise dans une base orthonormée, c'est-à-dire qu'il existe une matrice unitaire U et une matrice diagonale D telles que $A = UDU^*$, les coefficients diagonaux de D étant les valeurs propres dans $\mathbb{C}$ de la matrice A.*

Démonstration : On montre tout d'abord que pour toute valeur propre $\lambda \in \mathbb{C}$ de A, le sous-espace propre associé $E_\lambda = \ker(A - \lambda I_n)$ est stable par A^* et que son orthogonal $E_\lambda^\perp$ est stable par A et par A^*, ce qui permet de raisonner par récurrence sur l'ordre n de la matrice A.

Pour tout $x \in E_\lambda$ on a, puisque A est normale :

$$AA^*x = A^*Ax = \lambda A^*x$$

et $A^*x \in E_\lambda$. C'est-à-dire que E_λ est stable par A^*.

Soient $x \in E_\lambda^\perp$ et $y \in E_\lambda$. Si A est normale, on a alors $A^*y \in E_\lambda$ et :

$$\langle Ax \mid y \rangle = \langle x \mid A^*y \rangle = 0,$$

c'est-à-dire que $Ax \in E_\lambda^\perp$ et $E_\lambda^\perp$ est stable par A. Puis en écrivant que :

$$\langle A^*x \mid y \rangle = \langle x \mid Ay \rangle = \langle x \mid \lambda y \rangle = \bar{\lambda} \langle x \mid y \rangle = 0,$$

on déduit que $A^*x \in E_\lambda^\perp$ et $E_\lambda^\perp$ est stable par A^*.

On raisonne ensuite par récurrence sur la dimension n de l'espace vectoriel euclidien E.

Pour $n = 1$ le résultat est évident. Supposons-le acquis pour toute matrice normale d'ordre $p \in \{1, \ldots, n-1\}$. Soient λ une valeur propre de A et E_λ l'espace propre associé. Si $E_\lambda = E$ alors A est une homothétie et toute base orthonormée de E convient. Si $E_\lambda \neq E$, alors $E_\lambda^\perp$ est de dimension $p \in \{1, \ldots, n-1\}$ et la restriction B de A à $E_\lambda^\perp$ est un endomorphisme normal de $E_\lambda^\perp$ (une matrice est identifiée à l'application linéaire qu'elle définie dans la base canonique de E). Il existe donc une base orthonormée de $E_\lambda^\perp$ formée de vecteurs propres de B, donc de A. En complétant cette base par une base orthonormée de E_λ on obtient une base orthonormée de E formée de vecteurs propres de A. ∎

Remarque 3.9 *: Si λ est une valeur propre de A alors la stabilité de $E_\lambda^\perp$ par A^* est en fait vraie pour toute matrice A.*

Du fait qu'une matrice hermitienne ou unitaire est normale on déduit les résultats suivants.

Corollaire 3.13 : *Une matrice hermitienne [resp. unitaire] a ses valeurs propres réelles [resp. de module 1] et se diagonalise dans une base orthonormée.*

De ce résultat on déduit l'existence d'une racine carrée pour une matrice complexe hermitienne positive.

Corollaire 3.14 : *Si A est une matrice complexe hermitienne positive, alors il existe une unique matrice hermitienne positive B telle que $A = B^2$.*

Démonstration : La matrice A étant hermitienne positive, elle a toutes ses valeurs propres réelles positives et se diagonalise dans une base orthonormée, c'est-à-dire qu'il existe $U \in U_n(\mathbb{C})$ et $(\lambda_1, \ldots, \lambda_n) \in (\mathbb{R}^+)^n$ tels que $A = UDU^*$, avec :

$$D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}.$$

En posant :

$$\Delta = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \sqrt{\lambda_n} \end{pmatrix},$$

et $B = U\Delta U^*$, on a $B^2 = A$, la matrice B étant hermitienne positive (ses valeurs propres sont positives).

Si P est le polynôme d'interpolation de Lagrange défini par $P(\lambda_i) = \sqrt{\lambda_i}$ pour tout $i \in \{1, \ldots, n\}$ (le degré de P est $p-1$ où p est le nombre de valeurs propres distinctes de A), alors

$$P(A) = UP(D)U^* = U\Delta U^* = B.$$

C'est-à-dire que B est polynomiale en A.

Si C est une autre racine carrée de A hermitienne positive, on a alors $C^2 = A$ et C commute avec A, donc avec B. En définitive les matrices B et C commutent et sont diagonalisables, et on sait alors qu'elles sont simultanément diagonalisables (exercice 3.3), c'est-à-dire qu'il existe une matrice P dans $GL_n(\mathbb{C})$ telle que $C = P\Gamma P^{-1}$ et $B = P\Lambda P^{-1}$ avec Γ et Λ diagonales à coefficients réels positifs. De $C^2 = A = B^2$, on déduit alors que $\Gamma^2 = \Lambda^2$ et $\Gamma = \Lambda$ du fait que ces matrices sont diagonales à coefficients réels positifs. Et en définitive $B = C$. ■

Remarque 3.10 :

(*i*) *Une matrice réelle symétrique et positive admet également une unique racine carrée symétrique positive.*

(*ii*) *On montre de manière analogue que si A est une matrice complexe hermitienne [resp. réelle symétrique] positive, alors pour tout entier $k > 0$, il existe une unique matrice complexe hermitienne [resp. réelle symétrique] positive B telle que $A = B^k$. De plus, si A est définie positive, il en est de même de B.*

De ce résultat on peut déduire l'existence de la décomposition polaire d'une matrice réelle ou complexe inversible. Précisément on a le résultat suivant.

Corollaire 3.15 : *Toute matrice complexe [resp. réelle] inversible A peut s'écrire de manière unique $A = UH$ [resp. $A = \Omega S$] où U [resp. Ω] est une matrice unitaire [resp. orthogonale] et H [resp. S] une matrice hermitienne [resp. symétrique] définie positive.*

Démonstration : On démontre le résultat dans le cas complexe, le cas des matrices réelles se traitant de manière analogue.

Si $A = UH$, alors $A^*A = HU^*UH = H^2$ et H est la racine carrée de la matrice hermitienne définie positive A^*A ($\langle A^*Ax \mid x\rangle = \|Ax\|_2^2 > 0$ pour x non nul). La matrice U est alors donnée par $U = AH^{-1}$ (A inversible entraîne H inversible). On a donc, en cas d'existence, l'unicité des matrices U et H.

Si $A \in GL_n(\mathbb{C})$, alors A^*A est hermitienne définie positive et elle admet une unique racine carrée hermitienne définie positive H. En posant $U = AH^{-1}$, on a $A = UH$ et :

$$U^*U = \left(H^{-1}\right)^* (A^*A)\, H^{-1} = (H^*)^{-1} H^2 H^{-1} = H^{-1}H = I_n,$$

c'est-à-dire que U est unitaire. ∎

9. Forme réduite de Jordan des matrices complexes

Dans le cas complexe on peut affiner le théorème 3.1 de trigonalisation comme on va le voir dans ce paragraphe.

On suppose ici que E est un espace vectoriel de dimension $n \geq 2$ sur $\mathbb{C}$ et on désigne par E^* le dual algébrique de E, c'est-à-dire l'espace vectoriel des formes linéaires sur E.

L'orthogonal dans E d'une partie non vide Y de E^* est le sous-espace vectoriel Y° de E défini par :

$$Y^\circ = \{x \in E \mid \forall \varphi \in Y, \quad \varphi(x) = 0\}.$$

La base duale de $\{e_1, e_2, \ldots, e_n\}$ est la base de E^*, notée $\{e_1^*, e_2^*, \ldots, e_n^*\}$ définie par :

$$\forall j \in \{1, 2, \ldots, n\}, \quad \forall x = \sum_{i=1}^{n} x_i e_i, \quad e_j^*(x) = x_j.$$

Pour tout endomorphisme $u \in \mathcal{L}(E)$, on note ${}^t u \in \mathcal{L}(E^*)$ le transposé de u défini par :

$$\forall \varphi \in E^*, \quad {}^t u(\varphi) = \varphi \circ u.$$

Si $u \in \mathcal{L}(E)$ a pour matrice A dans la base canonique alors la matrice de ${}^t u$ dans la base duale est ${}^t A$.

Lemme 3.7 : *Si $u \in \mathcal{L}(E)$ est nilpotent d'ordre $p > 0$ alors ${}^t u \in \mathcal{L}(E^*)$ est aussi nilpotent d'ordre p.*

Démonstration : Pour u, v dans $\mathcal{L}(E)$, on a ${}^t(u \circ v) = {}^t v \circ {}^t u$. On a donc $({}^t u)^p = {}^t(u^p) = 0$ et $({}^t u)^{p-1} = {}^t\left(u^{p-1}\right) \neq 0$, c'est-à-dire que ${}^t u$ est nilpotent d'ordre p. ∎

Lemme 3.8 : *Soit u un endomorphisme de E nilpotent d'ordre $p > 0$. Il existe x dans E tel que le système $\left\{x, u(x), \ldots, u^{p-1}(x)\right\}$ soit libre dans E.*

Démonstration : On a $u^{p-1} \neq 0$. Il existe donc un vecteur non nul x dans E tel que $u^{p-1}(x) \neq 0$. Si $\sum\limits_{k=0}^{p-1} \lambda_k u^k(x) = 0$, alors :

$$0 = u^{p-1}\left(\sum_{k=0}^{p-1} \lambda_k u^k(x)\right) = \lambda_0 u^{p-1}(x)$$

$(u^p = 0)$ et $\lambda_0 = 0$. Si $p = 1$ c'est fini, sinon, en supposant que $\lambda_0 = \ldots = \lambda_j = 0$ pour $0 \leq j \leq p-2$, on a $\sum\limits_{k=j+1}^{p-1} \lambda_k u^k(x) = 0$, et, en appliquant u^{p-2-j} à cette dernière égalité, on obtient $\lambda_{j+1} u^{p-1}(x) = 0$ et $\lambda_{j+1} = 0$. D'où le résultat. ■

Lemme 3.9 : *Soit u un endomorphisme de E nilpotent d'ordre $p > 0$. Il existe $\varphi \in E^*$ et $x \in E$ tels que l'espace vectoriel $F = \operatorname{Vect}\{x, u(x), \ldots, u^{p-1}(x)\}$ et l'orthogonal G dans E de $H = \operatorname{Vect}\left\{\varphi, {}^t u(\varphi), \ldots, ({}^t u)^{p-1}(\varphi)\right\}$ sont stables par u avec $E = F \oplus G$.*

Démonstration : L'endomorphisme ${}^t u$ étant nilpotent d'ordre p, il existe une forme linéaire $\varphi \in E^*$ telle que $({}^t u)^{p-1}(\varphi) \neq 0$, c'est-à-dire que $\varphi \circ u^{p-1} \neq 0$ et qu'il existe donc x dans E tel que $\varphi\left(u^{p-1}(x)\right) \neq 0$.

On a nécessairement $u^{p-1}(x) \neq 0$ et avec le lemme 3.8 on déduit que les espaces vectoriels F et H sont de dimension p.

Du fait que u et ${}^t u$ sont nilpotents d'ordre p, on déduit que F est stable par u et que H est stable par ${}^t u$.

Pour tout y dans G et tout $k \in \{0, 1, \ldots, p-1\}$, on a :

$${}^t u^k(\varphi)(y) = \varphi\left(u^k(y)\right) = 0$$

et ${}^t u^k(\varphi)\left(u(y)\right) = \varphi\left(u^{k+1}(y)\right) = 0$ $(u^p = 0)$, donc $u(y) \in G$. C'est-à-dire que G est stable par u.

Soit $y = \sum\limits_{k=0}^{p-1} \lambda_k u^k(x) \in F \cap G$. On a $0 = \varphi\left(u^{p-1}(y)\right) = \lambda_0 \varphi\left(u^{p-1}(x)\right)$ et $\lambda_0 = 0$. En supposant que $\lambda_0 = \ldots = \lambda_j = 0$ pour $0 \leq j \leq p-2$ (si $p = 1$ c'est fini), on a $y = \sum\limits_{k=j+1}^{p-1} \lambda_k u^k(x)$, $0 = \varphi\left(u^{p-j-2}(y)\right) = \lambda_{j+1} \varphi\left(u^{p-1}(x)\right)$ et $\lambda_{j+1} = 0$. On a donc $F \cap G = \{0\}$ et avec les dimensions $\left(\dim(F) = p, \dim(G) = n - \dim(H) = n - p\right)$ on déduit que $E = F \oplus G$. ■

De ces lemmes on déduit les résultats suivants.

Théorème 3.19 : *Si $u \in \mathcal{L}(E)$ est nilpotent d'ordre $p > 0$, alors il existe une base de E :*

$$\mathcal{B} = \mathcal{B}_1 \cup \ldots \cup \mathcal{B}_r,$$

telle que chaque sous-espace vectoriel $E_i = \mathrm{Vect}(\mathcal{B}_i)$ soit stable par u et la matrice de la restriction de u à E_i est :

$$J_i = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix} \in M_{p_i}(\mathbb{C}),$$

avec $p_i = \dim(E_i)$ $(1 \leq i \leq r)$.

Démonstration : On procède par récurrence sur $n = \dim(E)$. Pour $n = 1$, on a $u = 0$ et le résultat est trivial. Supposons-le acquis pour les espaces vectoriels de dimension strictement inférieure à n. Avec les notations du lemme 3.9, la matrice de $u_{|F}$ (F est stable par u) dans la base $\mathcal{B}_1 = \{x, u(x), \ldots, u^{p-1}(x)\}$ est :

$$J_p = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ 1 & 0 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & 0 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix} \in M_p(\mathbb{C}),$$

et, en complétant cette base par une base $\mathcal{B}_G$ de G, la matrice de u dans la base $\mathcal{B}_1 \cup \mathcal{B}_G$ est $A = \begin{pmatrix} J_p & 0 \\ 0 & A_{n-p} \end{pmatrix}$, où $A_{n-p} \in \mathcal{M}_{n-p}(\mathbb{C})$ est la matrice de $u_{|G}$ dans $\mathcal{B}_G$. L'hypothèse de récurrence permet alors de conclure. ■

Théorème 3.20 : *Soit $u \in \mathcal{L}(E) - \{0\}$ tel que son polynôme caractéristique s'écrive :*

$$P_u(X) = (-1)^n \prod_{k=1}^{p} (X - \lambda_k)^{\alpha_k},$$

avec $\alpha_k > 0$ et les λ_k distinctes deux à deux.

Il existe une base $\mathcal{B}$ de E dans laquelle la matrice de u est de la forme :

$$A = \begin{pmatrix} J_1 & 0 & \ldots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & J_p \end{pmatrix}, \tag{3.1}$$

avec :

$$\forall k \in \{1,2,\ldots,p\}, \quad J_k = \begin{pmatrix} \lambda_k & 0 & 0 & \cdots & 0 \\ \epsilon_{k,2} & \lambda_k & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \epsilon_{k,\alpha_k-1} & \lambda_k & 0 \\ 0 & \cdots & 0 & \epsilon_{k,\alpha_k} & \lambda_k \end{pmatrix} \in M_{\alpha_k}(\mathbb{C}),$$

où $\epsilon_{k,i} \in \{0,1\}$ (forme réduite de Jordan).

Démonstration : On sait que si on désigne, pour tout $k \in \{1,2,\ldots,p\}$, par $N_k = \ker(u-\lambda_k Id)^{\alpha_k}$ les sous-espaces caractéristiques de u où α_k est la multiplicité de λ_k comme racine du polynôme minimal, on a $E = \bigoplus_{k=1}^{p} N_k$. Chaque sous-espace vectoriel N_k est de dimension α_k stable par u, λ_k est la seule valeur propre de la restriction de u à N_k et la restriction de $u - \lambda_k Id$ à N_k est nilpotente d'indice α_k.

On déduit alors qu'il existe une base $\mathcal{B}_k$ de N_k dans laquelle la matrice de la restriction de $u - \lambda_k Id$ à N_k est de la forme :

$$J_k = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ \epsilon_{k,2} & 0 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \epsilon_{k,\alpha_k-1} & 0 & 0 \\ 0 & \cdots & 0 & \epsilon_{k,\alpha_k} & 0 \end{pmatrix} \in M_{\alpha_k}(\mathbb{C}).$$

Dans la réunion de ces bases, la matrice de u a la forme indiquée. ■

Corollaire 3.16 : *Toute matrice non nulle A d'ordre n à coefficients complexes est semblable à une matrice triangulaire de la forme (3.1).*

10. Exercices

Exercice 3.1 : *Soit u un endomorphisme de E diagonalisable. Montrer que si F est un sous-espace vectoriel de E stable par u alors la restriction de u à F est aussi diagonalisable.*

Solution : Dire que u est diagonalisable équivaut à dire que son polynôme minimal π_u est scindé sur $\mathbb{K}$ à racines simples. Si F est stable par u alors la restriction v de u à F est un endomorphisme de F annulé par le polynôme π_u qui est scindé sur $\mathbb{K}$ à racines simples. Il en résulte que v est diagonalisable.

Exercice 3.2 : *Soit u un endomorphisme de E diagonalisable de valeurs propres deux à deux distinctes $\lambda_1, \ldots, \lambda_p$ $(1 \leq p \leq n)$. Montrer que pour $1 \leq k \leq p$ la projection de E sur le sous-espace propre* $\ker(u - \lambda_k Id)$ *est donnée par :*

$$p_k = \alpha_k \prod_{\substack{j=1 \\ j \neq k}}^{p} (u - \lambda_j Id),$$

avec $\alpha_k = \dfrac{1}{\prod\limits_{\substack{j=1 \\ j \neq k}}^{p} (\lambda_k - \lambda_j)}$ *(utiliser la décomposition en éléments simples de* $\dfrac{1}{\pi_u}$*).*

Solution : Si u est diagonalisable alors son polynôme minimal s'écrit :

$$\pi_u(X) = \prod_{k=1}^{p} (X - \lambda_k),$$

où $\lambda_1, \ldots, \lambda_p$ $(1 \leq p \leq n)$ sont les valeurs propres deux à deux distinctes de u. La décomposition en éléments simples de $\dfrac{1}{\pi_u}$:

$$\frac{1}{\pi_u(X)} = \sum_{k=1}^{p} \frac{\alpha_k}{X - \lambda_k},$$

où $\alpha_k = \dfrac{1}{\prod\limits_{\substack{j=1 \\ j \neq k}}^{p} (\lambda_k - \lambda_j)}$, permet d'obtenir l'égalité de Bézout :

$$\sum_{k=1}^{p} \alpha_k \prod_{\substack{j=1 \\ j \neq k}}^{p} (X - \lambda_j) = 1.$$

De cette égalité, on déduit que :

$$\forall x \in E, \quad x = \sum_{k=1}^{p} \alpha_k \prod_{\substack{j=1 \\ j \neq k}}^{p} (u - \lambda_j Id)\, x.$$

En considérant que pour tout $k \in \{1, \ldots, p\}$ et tout $x \in E$ on a :

$$(u - \lambda_k Id) \prod_{\substack{j=1 \\ j \neq k}}^{p} (u - \lambda_j Id)\, x = \pi_u(u)\, x = 0,$$

on déduit que $\alpha_k \prod_{\substack{j=1 \\ j\neq k}}^{p} (u - \lambda_j Id)\, x \in \ker(u - \lambda_k Id)$ c'est-à-dire que les projecteurs p_k s'écrivent :

$$p_k = \alpha_k \prod_{\substack{j=1 \\ j\neq k}}^{p} (u - \lambda_j Id) \text{ avec } \alpha_k = \frac{1}{\prod_{\substack{j=1 \\ j\neq k}}^{p} (\lambda_k - \lambda_j)}.$$

En particulier ces projecteurs sont des polynômes en u.

Exercice 3.3 : *Soit $\{u_i \mid i \in I\}$ une famille d'endomorphismes de E (I est un ensemble d'indices non nécessairement fini). On suppose que les u_i commutent deux à deux et sont diagonalisables. Montrer qu'il existe une base commune de diagonalisation.*

Solution : On procède par récurrence sur la dimension $n \geq 1$ de E. Pour $n = 1$ le résultat est évident. On suppose que E est de dimension $n+1$ et que le résultat est acquis pour les espaces vectoriels de dimension inférieure ou égale à n. Si tous les u_i sont des homothéties alors le résultat est clair. Sinon, soit j dans I tel que u_j ne soit pas une homothétie. On a alors la décomposition en sous-espaces propres :

$$E = \bigoplus_{k=1}^{p} \ker(u_j - \lambda_k Id).$$

L'endomorphisme u_j n'étant pas une homothétie chaque sous-espace propre est de dimension inférieure ou égale à n. Comme tous les u_i commutent à u_j, chaque sous-espace propre est stable par u_i pour tout i dans I et la restriction de chaque u_i à chaque $\ker(u_j - \lambda_k Id)$ est diagonalisable. On peut donc appliquer, pour tout entier k compris entre 1 et p, l'hypothèse de récurrence à la famille des restrictions des u_i à $\ker(u_j - \lambda_k Id)$, ce qui permet de construire une base de diagonalisation de $\ker(u_j - \lambda_k Id)$ commune à toutes les restrictions de u_i à cet espace. La réunion de ces bases donne alors une base de diagonalisation commune à tous les u_i.

Exercice 3.4 : *Soit $(E, \|\cdot\|)$ un espace vectoriel normé réel. On note :*

$$\mu(E) = \sup_{(x,y)\neq(0,0)} \frac{\|x+y\|^2 + \|x-y\|^2}{2\left(\|x\|^2 + \|y\|^2\right)}.$$

Montrer que les assertions suivantes sont équivalentes :

(i) $\forall (x,y) \in E^2, \quad \|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right)$ *(identité du parallélogramme).*
(ii) *La norme $\|\cdot\|$ dérive d'un produit scalaire sur E.*
(iii) $\mu(E) = 1$.
(iv) *Pour x, y fixés dans E, $\|x+ty\|^2$ est un trinôme en t.*

Solution :

(i) $\Rightarrow$ (ii) On définit la fonction $\langle \cdot \mid \cdot \rangle$ sur $E \times E$ par :

$$(x,y) \mapsto \langle x \mid y \rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right).$$

Il est facile de vérifier que cette application est symétrique ($\langle x \mid y \rangle = \langle y \mid x \rangle$) et définie positive ($\langle x \mid x \rangle = \|x\|^2 \geq 0$ avec égalité si et seulement si $x = 0$). En utilisant l'identité du parallélogramme on va montrer que cette application est bilinéaire. C'est donc un produit scalaire et la norme $\|\cdot\|$ en dérive.
Avec l'identité du parallélogramme on a, pour (x, y, z) dans E^3 :

$$\begin{aligned}
&4\left(\langle x+z \mid y \rangle + \langle x-z \mid y \rangle\right) \\
&\qquad = \|x+z+y\|^2 - \|x+z-y\|^2 + \|x-z+y\|^2 - \|x-z-y\|^2 \\
&\qquad = \left(\|x+y+z\|^2 + \|x+y-z\|^2\right) - \left(\|x-y+z\|^2 + \|x-y-z\|^2\right) \\
&\qquad = 2\left(\|x+y\|^2 + \|z\|^2\right) - 2\left(\|x-y\|^2 + \|z\|^2\right) \\
&\qquad = 8\langle x \mid y \rangle.
\end{aligned}$$

Soit :

$$\langle x+z \mid y \rangle + \langle x-z \mid y \rangle = 2\langle x \mid y \rangle.$$

Ce qui donne pour $z = x$:

$$\forall (x,y) \in E^2, \quad \langle 2x \mid y \rangle = 2\langle x \mid y \rangle.$$

En posant $u = \frac{1}{2}(x+y)$, $v = \frac{1}{2}(x-y)$, on a alors :

$$\begin{aligned}
\langle x+y \mid z \rangle = \langle 2u \mid z \rangle = 2\langle u \mid z \rangle &= \langle u+v \mid z \rangle + \langle u-v \mid z \rangle \\
&= \langle x \mid z \rangle + \langle y \mid z \rangle.
\end{aligned}$$

On a donc, pour tout (x, y, z) dans E^3 :

$$\langle x+y \mid z \rangle = \langle x \mid z \rangle + \langle y \mid z \rangle.$$

Par récurrence on déduit alors que :

$$\forall (x,y) \in E^2, \quad \forall n \in \mathbb{N} - \{0\}, \quad \langle nx \mid y \rangle = n\langle x \mid y \rangle. \tag{3.2}$$

Avec $\langle 0 \mid y \rangle = 0$ et $\langle -x \mid y \rangle = -\langle x \mid y \rangle$, on déduit que (3.2) est valable pour tout entier n dans $\mathbb{Z}$. Puis, avec $\langle x \mid y \rangle = n\left\langle \frac{1}{n}x \mid y \right\rangle$ pour tout entier $n > 0$, on déduit que (3.2) est valable pour tout $r \in \mathbb{Q}$. Enfin, avec la densité de $\mathbb{Q}$ dans $\mathbb{R}$ et la continuité de l'application :

$$(x,y) \longmapsto \langle x \mid y \rangle = \frac{1}{4}\left(\|x+y\|^2 - \|x-y\|^2\right),$$

on déduit que (3.2) est valable pour tout $r \in \mathbb{R}$. Ce qui achève de prouver que l'application $\langle \cdot \mid \cdot \rangle$ est bilinéaire.

(ii) $\Rightarrow$ (iii) Si la norme dérive d'un produit scalaire on vérifie facilement qu'on a l'identité du parallélogramme et l'égalité $\mu(E) = 1$.

(iii) $\Rightarrow$ (i) Si $\mu(E) = 1$, on a alors :

$$\forall (x, y) \in E^2, \quad \|x+y\|^2 + \|x-y\|^2 \leq 2\left(\|x\|^2 + \|y\|^2\right).$$

En posant $u = x + y$, $v = x - y$, on a :

$$\begin{aligned}\|2x\|^2 + \|2y\|^2 &= \|u+v\|^2 + \|u-v\|^2 \\ &\leq 2\left(\|u\|^2 + \|v\|^2\right) = 2\left(\|x+y\|^2 + \|x-y\|^2\right),\end{aligned}$$

soit $2\left(\|x\|^2 + \|y\|^2\right) \leq \|x+y\|^2 + \|x-y\|^2$ et l'égalité.

On a donc les équivalences (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii).

L'implication (ii) $\Rightarrow$ (iv) est évidente. Pour conclure, on va montrer que (iv) $\Rightarrow$ (i).

Pour x, y fixés dans E on peut écrire, pour tout réel t :

$$P(t) = \|x+ty\|^2 = at^2 + 2bt + c.$$

Le coefficient c est donné en faisant $t = 0$, soit $c = \|x\|^2$, et le coefficient a est donné par :

$$a = \lim_{t \to +\infty} \frac{P(t)}{t^2} = \lim_{t \to +\infty} \left\|\frac{1}{t}x + y\right\|^2 = \|y\|^2.$$

On déduit alors que :

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= P(1) + P(-1) = 2(a+c) \\ &= 2\left(\|x\|^2 + \|y\|^2\right),\end{aligned}$$

c'est-à-dire l'identité du parallélogramme.

Exercice 3.5 :

1. Montrer que l'application $(A, B) \longmapsto \langle A \mid B\rangle = \operatorname{Tr}(A^tB)$ définit un produit scalaire sur $\mathcal{M}_n(\mathbb{R})$.

2. Montrer que pour toute forme linéaire $\varphi \in \mathcal{M}_n(\mathbb{R})^*$ il existe une unique matrice $C \in \mathcal{M}_n(\mathbb{R})$ telle que :

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \quad \varphi(A) = \operatorname{Tr}\left(A^tC\right).$$

3. Montrer que l'application déterminant, $\det : \mathcal{M}_n(\mathbb{R}) \to \mathbb{R}$, est différentiable et que sa différentielle u est donnée par :

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \forall H \in \mathcal{M}_n(\mathbb{R}), \quad u(A)(H) = \operatorname{Tr}\left(H^tC\right),$$

où C est la comatrice de la matrice A.

4. Soit Y une fonction définie sur un intervalle réel I à valeurs dans $\mathcal{M}_n(\mathbb{R})$ et dérivable. Montrer que la fonction $f : t \longmapsto \det(Y(t))$ est dérivable et calculer sa dérivée.

Solution :

1. Pour toutes matrice A, B dans $\mathcal{M}_n(\mathbb{R})$, on a :

$$\langle A \mid B\rangle = \sum_{1\leq i,j\leq n} a_{ij}b_{ij},$$

c'est-à-dire que $\langle\cdot \mid \cdot\rangle$ est le produit scalaire euclidien de $\mathcal{M}_n(\mathbb{R}) \backsimeq \mathbb{R}^{n^2}$.

2. La base canonique $\{E_{i,j} \mid 1 \leq i, j \leq n\}$ de $\mathcal{M}_n(\mathbb{R})$ est orthonormée pour ce produit scalaire. Pour toute forme linéaire $\varphi \in \mathcal{M}_n(\mathbb{R})^*$ et toute matrice $A \in \mathcal{M}_n(\mathbb{R})$, on a :

$$\varphi(A) = \varphi\left(\sum_{1\leq i,j\leq n} a_{ij}E_{i,j}\right) = \sum_{1\leq i,j\leq n} a_{ij}\varphi(E_{i,j}) = \langle X \mid C\rangle,$$

où $C \in \mathcal{M}_n(\mathbb{R})$ est définie par $C = \sum\limits_{1\leq i,j\leq n} \varphi(E_{i,j})\, E_{i,j}$. On a donc ainsi montré que l'application :

$$C \longmapsto (\varphi_C : A \longmapsto \langle A \mid C\rangle)$$

est surjective de $\mathcal{M}_n(\mathbb{R})$ sur $\mathcal{M}_n(\mathbb{R})^*$. Cette application étant linéaire et les espaces vectoriels $\mathcal{M}_n(\mathbb{R})$ et $\mathcal{M}_n(\mathbb{R})^*$ étant de même dimension on en déduit que c'est un isomorphisme. C'est-à-dire que pour toute forme linéaire $\varphi \in \mathcal{M}_n(\mathbb{R})^*$ il existe une unique matrice $C \in \mathcal{M}_n(\mathbb{R})$ telle que :

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \quad \varphi(A) = \operatorname{Tr}\left(A\,{}^tC\right).$$

***Remarque 3.11** : Ce résultat est vrai sur tout corps commutatif $\mathbb{K}$.*

3. L'application det est continûment dérivable comme fonction polynomiale des coefficients, donc différentiable. Notons u sa différentielle. Pour toute matrice $A \in \mathcal{M}_n(\mathbb{R})$, $u(A)$ est une forme linéaire sur $\mathcal{M}_n(\mathbb{R})$, et il existe donc une unique matrice C telle que $u(A)(H) = \operatorname{Tr}(H\,{}^tC)$ pour tout $H \in \mathcal{M}_n(\mathbb{R})$.

Pour $H = E_{i,j}$ $(1 \leq i, j \leq n)$ on a $\operatorname{Tr}(H\,{}^tC) = c_{ij}$, c'est-à-dire que $C = \Big((u(A)(E_{i,j}))\Big)_{1\leq i,j\leq n}$. Il s'agit donc de calculer les $u(A)(E_{i,j})$, pour $1 \leq i, j \leq n$.

La différentielle de l'application det en A s'écrit :

$$u(A) = \sum_{1\leq i,j\leq n} \frac{\partial \det}{\partial x_{ij}}(A)\, dx_{ij},$$

où on a noté $dx_{ij} = E^*_{i,j}$ (base duale de la base canonique de $\mathcal{M}_n(\mathbb{R})$). On a donc :

$$u(A)(E_{i,j}) = \frac{\partial \det}{\partial x_{ij}}(A) = \lim_{\substack{t\to 0\\ t\neq 0}} \frac{\det(A + tE_{i,j}) - \det(A)}{t}.$$

En notant A_j la colonne j de la matrice A, $\{e_1, \dots, e_n\}$ la base canonique de $\mathbb{R}^n$ et en utilisant la n-linéarité du déterminant, on a :

$$\begin{aligned}\det(A + tE_{i,j}) &= \det(A_1, \dots, A_{j-1}, A_j + te_i, A_{j+1}, \dots, A_n) \\ &= \det(A) + t\det(A_1, \dots, A_{j-1}, e_i, A_{j+1}, \dots, A_n)\end{aligned}$$

et

$$c_{ij} = \frac{\partial \det}{\partial x_{ij}}(A) = \det(A_1, \dots, A_{j-1}, e_i, A_{j+1}, \dots, A_n)$$

(cofacteur (i, j) de A). On a donc :

$$\forall H \in \mathcal{M}_n(\mathbb{R}), \quad u(A)(H) = \mathrm{Tr}\left(H\,{}^tC\right),$$

où C est la matrice des cofacteurs de la matrice A.

4. La fonction f est dérivable sur I comme composée de fonctions dérivables, avec :

$$\forall t \in I, \quad f'(t) = u(Y)\left(Y'(t)\right) = \sum_{1 \leq i,j \leq n} \frac{\partial \det}{\partial x_{ij}}(Y)(t)\, y'_{ij}(t).$$

C'est-à-dire que :

$$\begin{aligned}f' &= \sum_{1 \leq i,j \leq n} \det(Y_1, \dots, Y_{j-1}, e_i, Y_{j+1}, \dots, Y_n)\, y'_{ij} \\ &= \sum_{j=1}^{n} \det\left(Y_1, \dots, Y_{j-1}, \sum_{i=1}^{n} y'_{ij} e_i, Y_{j+1}, \dots, Y_n\right) \\ &= \sum_{j=1}^{n} \det\left(Y_1, \dots, Y_{j-1}, Y'_j, \dots, Y_n\right).\end{aligned}$$

Exercice 3.6 : *Soit $A \in \mathcal{M}_n(\mathbb{R})$. On appelle sous-matrices principales de A les matrices :*

$$A_k = \left((a_{ij})\right)_{1 \leq i,j \leq k}, \quad (k = 1, \dots, n),$$

et déterminants principaux les réels $\Delta_k = \det(A_k)$.

Montrer qu'une matrice symétrique réelle est définie positive si et seulement si tous ses déterminants principaux sont strictement positifs.

Solution : Si la matrice A est symétrique définie positive alors il en est de même de toutes ses sous-matrices principales. En effet, pour tout entier k compris entre 1 et n, si x est un vecteur non nul de $\mathbb{R}^k$, le vecteur $x' = \begin{pmatrix} x \\ 0 \end{pmatrix} \in \mathbb{R}^n$ est non nul et

$$\langle A_k x \mid x \rangle = \langle A x' \mid x' \rangle > 0.$$

On déduit alors que toutes les valeurs propres de la matrice A_k sont strictement positives et $\det(A_k) > 0$ pour tout entier k compris entre 1 et n.

Pour la réciproque on procède par récurrence sur $n \geq 1$. Pour $n = 1$ le résultat est évident. Supposons-le acquis pour les matrices symétriques d'ordre $n - 1 \geq 1$ et soit A une matrice symétrique réelle d'ordre n ayant tous ses déterminants principaux strictement positifs. On désigne par $\{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$. Avec l'hypothèse de récurrence on déduit que la matrice symétrique A_{n-1} est définie positive et la restriction à l'espace vectoriel H engendré par $\{e_1, \ldots, e_{n-1}\}$ de la forme quadratique $q : x \mapsto \langle Ax \mid x \rangle$ est définie positive. Il existe alors une base $\{f_1, \ldots, f_{n-1}\}$ de H orthonormée pour $q_{|H}$ et la matrice de q dans la base $\{f_1, \ldots, f_{n-1}, e_n\}$ s'écrit :

$$A' = \begin{pmatrix} 1 & 0 & \ldots & \alpha_1 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & \alpha_{n-1} \\ \alpha_1 & \ldots & \alpha_{n-1} & \alpha_n \end{pmatrix}.$$

En posant :

$$f_n = e_n - \sum_{i=1}^{n-1} \alpha_i f_i,$$

le système $\{f_1, \ldots, f_{n-1}, f_n\}$ est une base de $\mathbb{R}^n$ orthogonale pour q et la matrice de q dans cette base s'écrit :

$$A'' = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \ldots & 0 & \beta_n \end{pmatrix}.$$

Le déterminant de la matrice A'' est du même signe que celui de A, c'est-à-dire positif. En conséquence, la forme quadratique q est définie positive et la matrice A symétrique définie positive.

Exercice 3.7 : *On se place dans $(\mathbb{R}^n, \langle \cdot \mid \cdot \rangle)$ euclidien, on note $\|\cdot\|_2$ la norme associée et on désigne par S_1 la sphère unité de $\mathbb{R}^n$ pour cette norme.*

Soit A une matrice réelle symétrique d'ordre n. On lui associe la forme quadratique q définie par :

$$\forall x \in \mathbb{R}^n, \quad q(x) = \langle Ax \mid x \rangle .$$

1. Montrer que l'application q est bornée sur S_1 et atteint ses bornes. On note $\lambda_0 = q(x_0) = \sup_{x \in S_1} q(x)$, où x_0 dans $\mathbb{R}^n$ est tel que $\|x_0\|_2 = 1$.

2. On se donne un vecteur x_1 dans $(\mathbb{R}x_0)^\perp \cap S_1$ et pour tout réel $\theta \in [-\pi, \pi]$ on note :

$$x_\theta = \cos(\theta)\, x_0 + \sin(\theta)\, x_1.$$

En utilisant les inégalités $q(x_\theta) \leq q(x_0)$ montrer que :

$$\forall \theta \in [-\pi, \pi], \quad 2\cos(\theta)\sin(\theta)\langle Ax_0 \mid x_1\rangle \leq \sin^2(\theta)\big(q(x_0) - q(x_1)\big)$$

et en déduire que $\langle Ax_0 \mid x_1 \rangle = 0$.

3. Déduire de ce qui précède que A se diagonalise dans une base orthonormée.

Solution :

1. L'application q étant continue en tant qu'application polynomiale, elle est donc bornée et atteint ses bornes sur la sphère unité S_1 qui est compacte dans $\mathbb{R}^n$.

2. On a $x_\theta \in S^1$ pour tout $\theta \in [-\pi, \pi]$, donc :

$$q(x_\theta) = \cos^2(\theta)\, q(x_0) + 2\cos(\theta)\sin(\theta)\langle Ax_0 \mid x_1\rangle + \sin^2(\theta)\, q(x_1) \leq q(x_0),$$

ce qui entraîne :

$$2\cos(\theta)\sin(\theta)\langle Ax_0 \mid x_1\rangle \leq \sin^2(\theta)\big(q(x_0) - q(x_1)\big).$$

On déduit alors que :

$$\begin{cases} \forall \theta \in \,]0, \pi[, \quad 2\cos(\theta)\langle Ax_0 \mid x_1\rangle \leq \sin(\theta)\big(q(x_0) - q(x_1)\big), \\ \forall \theta \in \,]-\pi, 0[, \ 2\cos(\theta)\langle Ax_0 \mid x_1\rangle \geq \sin(\theta)\big(q(x_0) - q(x_1)\big). \end{cases}$$

En faisant tendre θ vers 0 par valeurs positives et négatives, on déduit que $\langle Ax_0 \mid x_1 \rangle = 0$, c'est-à-dire que $x_1 \in \{Ax_0\}^\perp$.

3. On déduit de ce qui précède que pour tout x appartenant à $\{x_0\}^\perp - \{0\}$, x appartient à $\{Ax_0\}^\perp$ (il suffit de prendre $x_1 = \frac{1}{\|x\|}x$), c'est-à-dire que $\{x_0\}^\perp \subset \{Ax_0\}^\perp$. Il en résulte que :

$$\mathbb{R}(Ax_0) = \{Ax_0\}^{\perp\perp} \subset \{x_0\}^{\perp\perp} = \mathbb{R}x_0$$

et en particulier il existe $\lambda \in \mathbb{R}$ tel que $Ax_0 = \lambda x_0$, ce qui signifie que x_0 est un vecteur propre de A.

La matrice A étant symétrique, l'orthogonal de x_0 dans $\mathbb{R}^n$ est stable par A et il suffit d'appliquer le raisonnement précédent à cet orthogonal pour conclure par récurrence.

Exercice 3.8 : *Soit $A = (C_1, \ldots, C_n) \in \mathcal{M}_n(\mathbb{C})$ où C_k désigne la colonne k de A. Montrer l'inégalité d'Hadamard :*

$$|\det(A)| \leq \prod_{k=1}^{n} \|C_k\|,$$

en notant $\|\cdot\|$ la norme hermitienne sur $\mathbb{C}^n$ (on peut utiliser le procédé d'orthogonalisation de Gram-Schmidt).

Solution : Si $\det(A) = 0$, le résultat est alors évident.

On suppose que $\det(A) \neq 0$. Dans ce cas le système $\{C_1, \ldots, C_n\}$ est une base de $\mathbb{C}^n$ et le procédé de Gram-Schmidt nous permet alors de construire une base orthogonale $\{E_1, \ldots, E_n\}$ de $\mathbb{C}^n$ telle que :

$$E_1 = C_1\,; \quad \forall k \in \{2, \ldots, n\}, \quad E_k = C_k + \sum_{j=1}^{k-1} \lambda_{j,k} E_j.$$

Un déterminant étant inchangé si on ajoute à une colonne une combinaison linéaire des autres, en notant $N = (E_1, \ldots, E_n) \in \mathcal{M}_n(\mathbb{C})$, on a $\det(A) = \det(N)$.

On note $D = N^* N$ et on a :

$$D = \begin{pmatrix} \|E_1\|^2 & 0 & \ldots & 0 \\ 0 & \|E_2\|^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \|E_n\|^2 \end{pmatrix},$$

de sorte que :

$$|\det(A)|^2 = |\det(N)|^2 = \det(D) = \prod_{k=1}^{n} \|E_k\|^2.$$

On a donc :

$$|\det(A)| = \prod_{k=1}^{n} \|E_k\|.$$

Enfin, avec :

$$\|C_k\|^2 = \|E_k\|^2 + \sum_{j=1}^{k-1} |\lambda_{j,k}|^2 \|E_j\|^2 \geq \|E_k\|^2,$$

on déduit que :

$$|\det(A)| \leq \prod_{k=1}^{n} \|C_k\|.$$

Exercice 3.9 : *Montrer qu'une matrice A d'ordre $n > 0$ à coefficients complexes est normale si et seulement si il existe un polynôme P à coefficients complexes tel que $A^* = P(A)$.*

Solution : Une matrice normale se diagonalise dans une base orthonormée, c'est-à-dire qu'il existe une matrice unitaire U et une matrice diagonale D telles que $D = U^*AU$. En notant $\lambda_1, \lambda_2, \ldots, \lambda_p$ les éléments diagonaux deux à deux distincts de D (ce sont les valeurs propres de A) avec $1 \leq p \leq n$, et en désignant par P le polynôme d'interpolation de Lagrange défini par :

$$P \in \mathbb{C}_{p-1}[X], \quad P(\lambda_k) = \overline{\lambda_k} \quad (1 \leq k \leq p),$$

on a :

$$D^* = P(D) = U^*P(A)U$$

et

$$P(A) = UD^*U^* = (UDU^*)^* = A^*.$$

La réciproque est évidente.

Exercice 3.10 : *Soit A une matrice d'ordre $n > 0$ à coefficients complexes.*

1. Montrer que A est la matrice nulle si et seulement si $\mathrm{Tr}(AA^*) = 0$.

2. Montrer que la matrice A est hermitienne si et seulement si $AA^* = A^2$.

Solution :

1. Si $A = ((a_{ij}))_{1 \leq i,j \leq n}$, on a alors :

$$\mathrm{Tr}(AA^*) = \sum_{1 \leq i,j \leq n} |a_{ij}|^2$$

(carré de la norme hermitienne de $\mathbb{C}^{n^2}$), et $\mathrm{Tr}(AA^*) = 0$ équivaut bien à $A = 0$.

2. Soit $A \in \mathcal{M}_n(\mathbb{C})$ telle que $AA^* = A^2$. En posant $B = A - A^*$, on a :

$$BB^* = A^*A - (A^*)^2$$

et

$$\begin{aligned}\mathrm{Tr}(BB^*) &= \mathrm{Tr}(A^*A) - \mathrm{Tr}\left((A^*)^2\right) \\ &= \mathrm{Tr}(AA^*) - \mathrm{Tr}\left((A^*)^2\right) = \mathrm{Tr}\left(AA^* - (A^*)^2\right),\end{aligned}$$

avec :

$$AA^* = (AA^*)^* = \left(A^2\right)^* = (A^*)^2 .$$

On déduit donc que $\mathrm{Tr}\,(BB^*) = 0$ et $B = 0$, c'est-à-dire $A = A^*$. La réciproque est évidente.

Exercice 3.11 : *Soit $A \in GL_n(\mathbb{C})$. Montrer qu'il existe deux matrices unitaires U, V et une matrice diagonale D à coefficients réels strictement positifs telles que $A = UDV^*$ (décomposition singulière de la matrice A).*

Solution : La matrice A^*A est hermitienne définie positive (en effet, pour x non nul on a $\langle A^*Ax \mid x\rangle = \|Ax\|_2^2 > 0$). Il existe donc une matrice unitaire V telle que $V^*(A^*A)V = \Delta$, où :

$$\Delta = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix},$$

avec $\lambda_i \in \mathbb{R}_+^*$ pour tout $i \in \{1, \dots, n\}$ (la matrice A est inversible). En notant $W = AV$, on a alors $W^*W = D^2$, où :

$$D = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & 0 \\ 0 & \sqrt{\lambda_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \sqrt{\lambda_n} \end{pmatrix}$$

et, avec $U^*U = D^{-1}W^*WD^{-1} = D^{-1}D^2D^{-1} = I_n$, on déduit que la matrice $U = WD^{-1}$ est unitaire.

Ce qui donne en définitive $U^*AV = D^{-1}W^*W = D^{-1}D^2 = D$, avec D diagonale à coefficients réels strictement positifs.

Exercice 3.12 : *Soient A et B deux matrices hermitiennes positives dans $\mathcal{M}_n(\mathbb{C})$. Montrer que :*

$$0 \leq \mathrm{Tr}\,(AB) \leq \mathrm{Tr}\,(A)\,\mathrm{Tr}\,(B)$$

(on peut d'abord considérer le cas où la matrice A est diagonale).

Solution : Si $\{e_1, \ldots, e_n\}$ désigne la base canonique de $\mathbb{C}^n$, alors pour toute matrice $A = ((a_{ij}))_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{C})$ on a $a_{ij} = \langle Ae_j \mid e_i \rangle$ pour $1 \leq i, j \leq n$. Dans le cas particulier où A est hermitienne positive on a alors $a_{ii} = \langle Ae_i \mid e_i \rangle \geq 0$ pour $1 \leq i \leq n$.

Si A est diagonale hermitienne positive, alors :

$$A = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix},$$

avec $\lambda_i \in \mathbb{R}^+$ pour tout i. Et pour toute matrice $B \in \mathcal{M}_n(\mathbb{C})$ hermitienne positive, on a $b_{ii} \geq 0$ pour tout i et

$$0 \leq \operatorname{Tr}(AB) = \sum_{i=1}^{n} \lambda_i b_{ii} \leq \left(\sum_{i=1}^{n} \lambda_i \right) \left(\sum_{i=1}^{n} b_{ii} \right) = \operatorname{Tr}(A) \operatorname{Tr}(B).$$

Dans le cas général, on sait que si A est hermitienne positive, alors toutes les valeurs propres de A sont réelles positives et il existe une matrice unitaire U telle que :

$$A = U \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix} U^* = UDU^*,$$

avec $\lambda_i \in \mathbb{R}^+$ pour tout i.

En remarquant que $AB = UDU^*B = U(DU^*BU)U^*$ est semblable à $D(U^*BU) = DC$, on a $\operatorname{Tr}(AB) = \operatorname{Tr}(DC)$. Mais D est diagonale hermitienne positive et C hermitienne positive unitairement semblable à B. Donc :

$$\operatorname{Tr}(AB) = \operatorname{Tr}(DC) \leq \operatorname{Tr}(D) \operatorname{Tr}(C) = \operatorname{Tr}(A) \operatorname{Tr}(B).$$

Exercice 3.13 : *Montrer que l'application définie sur $\mathcal{M}_n(\mathbb{C})$ par :*

$$A \mapsto \|A\|_s = \sqrt{\operatorname{Tr}(A^*A)},$$

est norme (norme de Schur) et qu'elle est sous-multiplicative, c'est-à-dire que $\|AB\|_s \leq \|A\|_s \|B\|_s$ pour toutes matrices A et B appartenant à $\mathcal{M}_n(\mathbb{C})$.

Solution : Pour toute matrice A dans $\mathcal{M}_n(\mathbb{C})$, on a :

$$\|A\|_s = \sqrt{\operatorname{Tr}(A^*A)} = \sqrt{\sum_{i,j=1}^{n} |a_{ij}|^2},$$

c'est-à-dire que $\|\cdot\|_s$ est la norme hermitienne de $\mathcal{M}_n(\mathbb{C})$ identifié à $\mathbb{C}^{n^2}$.

Pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{C})$, on a :

$$\|AB\|_s^2 = \operatorname{Tr}\big(B^*(A^*AB)\big) = \operatorname{Tr}\big((A^*AB)B^*\big) = \operatorname{Tr}\big((A^*A)(BB^*)\big)$$

et, les matrices A^*A et BB^* étant hermitiennes positives, on peut utiliser le résultat de l'exercice 3.12 pour écrire que :

$$\|AB\|_s^2 \leq \operatorname{Tr}(A^*A)\operatorname{Tr}(BB^*) = \|A\|_s^2 \|B^*\|_s^2 = \|A\|_s^2 \|B\|_s^2 .$$

Ce qui prouve que la norme de Schur est sous-multiplicative.

Remarque 3.12 : *On peut aussi montrer ce résultat directement en écrivant que :*

$$\|AB\|_s^2 = \sum_{i,j=1}^{n} \left|\sum_{k=1}^{n} a_{ik}b_{kj}\right|^2 \leq \sum_{i,j=1}^{n} \left(\sum_{k=1}^{n} |a_{ik}|\,|b_{kj}|\right)^2$$

et en utilisant l'inégalité de Cauchy-Schwarz :

$$\|AB\|_s^2 \leq \sum_{i,j=1}^{n} \left(\sum_{k=1}^{n} |a_{ik}|^2\right)\left(\sum_{k=1}^{n} |b_{kj}|^2\right) = \|A\|_s^2 \|B\|_s^2 .$$

Exercice 3.14 : *Montrer que toute matrice à coefficients réels ou complexes est semblable à sa transposée (utiliser la décomposition de Jordan sur $\mathbb{C}$).*

Solution : On se place tout d'abord dans le cas où A est un bloc de Jordan complexe de la forme :

$$A = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \ddots & 0 \\ 0 & 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 0 & \lambda \end{pmatrix}.$$

L'endomorphisme u associé à la matrice A est alors défini dans la base canonique $\{e_1, \dots, e_n\}$ de $\mathbb{C}^n$ par :

$$\begin{cases} u(e_1) = \lambda e_1, \\ u(e_j) = e_{j-1} + \lambda e_j \quad (2 \leq j \leq n). \end{cases}$$

Dans la base $\{e_n, \ldots, e_1\}$ l'endomorphisme u a pour matrice ${}^t A$.

Si A est une matrice complexe quelconque, on sait qu'elle est semblable à une matrice de la forme :

$$A' = \begin{pmatrix} J_1 & 0 & \ldots & 0 \\ 0 & J_2 & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & J_p \end{pmatrix},$$

où les J_k sont des blocs de Jordan du type précédent. Chacun de ces J_k étant semblable à sa transposée on en déduit facilement que A est semblable à sa transposée.

Le résultat est encore valable pour les matrices réelles du fait que deux matrices réelles qui sont semblables sur $\mathbb{C}$ le sont sur $\mathbb{R}$ (exercice 2.1).

4 L'espace vectoriel normé $\mathcal{M}_n(\mathbb{K})$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$)

Dans ce chapitre, on note $\mathbb{K}$ le corps des réels ou des complexes.

Pour tout entier n strictement positif on note $\mathcal{M}_n(\mathbb{K})$ l'espace vectoriel des matrices à n lignes et n colonnes à coefficients dans $\mathbb{K}$ et $GL_n(\mathbb{K})$ le groupe multiplicatif des éléments inversibles de $\mathcal{M}_n(\mathbb{K})$.

On note I_n la matrice identité d'ordre n.

Une matrice A est identifiée à l'endomorphisme de $\mathbb{K}^n$ qu'elle définit dans la base canonique.

Pour tous i, j compris entre 1 et n, on note E_{ij} la matrice dont tous les coefficients sont nuls sauf celui d'indice (i, j) qui vaut 1. La famille $\{E_{i,j} \mid 1 \leq i, j \leq n\}$ est alors une base de $\mathcal{M}_n(\mathbb{K})$.

1. Norme matricielle induite par une norme vectorielle

Dans ce paragraphe, on se donne une norme $x \mapsto \|x\|$ sur $\mathbb{K}^n$ et on note :

$$B = \{x \in \mathbb{K}^n \mid \|x\| \leq 1\},$$
$$S = \{x \in \mathbb{K}^n \mid \|x\| = 1\},$$

la boule unité et la sphère unité de $(\mathbb{K}^n, \|\cdot\|)$.

On sait que sur un espace vectoriel normé une application linéaire est continue si et seulement si elle est bornée sur la sphère unité et qu'en dimension finie la sphère unité est compacte et tout endomorphisme est continu. On peut donc définir pour toute matrice A la quantité :

$$\|A\| = \sup_{x \in S} \|Ax\|.$$

Théorème 4.1 : *Pour toute norme $x \mapsto \|x\|$ sur $\mathbb{K}^n$, l'application :*

$$A \mapsto \|A\| = \sup_{x \in S} \|Ax\|$$

définit une norme sur $\mathcal{M}_n(\mathbb{K})$.

Démonstration : L'égalité $\|A\| = 0$ équivaut à $\|Ax\| = 0$ pour tout x dans S. En remarquant que pour tout x dans $\mathbb{K}^n$ on a $\frac{1}{\|x\|}x \in S$ et en utilisant la linéarité de A, on déduit que $A = 0$.

La vérification des autres propriétés d'une norme ne pose pas de problèmes particuliers. ∎

Définition 4.1 : *L'application :*

$$A \mapsto \|A\| = \sup_{x \in S} \|Ax\|$$

est la norme matricielle induite par (ou subordonnée à) la norme vectorielle $x \longmapsto \|x\|$.

L'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ étant de dimension finie, on sait que toutes les normes sur cet espace sont équivalentes, c'est-à-dire qu'on a une seule topologie sur cet espace compatible avec la structure d'espace vectoriel.

Les deux exemples qui suivent nous permettent de calculer effectivement les normes matricielles induites par les normes vectorielles $\|\cdot\|_\infty$ et $\|\cdot\|_1$.

Lemme 4.1 : *La norme matricielle induite par $\|\cdot\|_\infty$ est définie par :*

$$\forall A \in \mathcal{M}_n(\mathbb{K}), \quad \|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|.$$

Démonstration : Posons

$$\alpha = \max_{1 \le i \le n} \sum_{j=1}^{n} |a_{ij}|.$$

Si $A = 0$ alors $\alpha = 0$ et le résultat est évident. On suppose donc A non nulle.

Pour x dans S_∞ (sphère unité de $(\mathbb{K}^n, \|\cdot\|_\infty)$), on a :

$$\|Ax\|_\infty = \max_{1 \le i \le n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \le \alpha.$$

Pour montrer que $\alpha \le \|A\|$, il suffit de trouver un vecteur x dans S_∞ tel que $\|Ax\|_\infty = \alpha$. Soit k compris entre 1 et n tel que $\alpha = \sum_{j=1}^{n} |a_{kj}|$. On pose, pour tout j compris entre 1 et n :

$$x_j = \begin{cases} \dfrac{\overline{a_{kj}}}{|a_{kj}|} & \text{si } a_{kj} \neq 0, \\ 0 & \text{si } a_{kj} = 0. \end{cases}$$

On a alors $\|Ax\|_\infty = \alpha$. En effet, comme A est non nulle il existe un indice j compris entre 1 et n tel que $a_{kj} \neq 0$, donc $\|x\|_\infty = 1$ et, en notant $y = Ax$, on a :

$$\forall i \in \{1, 2, \ldots, n\}, \quad |y_i| = \left| \sum_{j=1}^{n} a_{ij} x_j \right| \le \alpha,$$

soit, avec pour $i = k$:

$$|y_k| = \left|\sum_{j=1}^{n} a_{kj} x_j\right| = \sum_{j=1}^{n} |a_{kj}| = \alpha.$$

On a donc bien $\|Ax\|_\infty = \alpha$ et $\|A\|_\infty = \alpha$. ∎

Lemme 4.2 : *La norme matricielle induite par $\|\cdot\|_1$ est définie par :*

$$\forall A \in \mathcal{M}_n(\mathbb{K}), \quad \|A\|_1 = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|.$$

Démonstration : Posons :

$$\beta = \max_{1 \le j \le n} \sum_{i=1}^{n} |a_{ij}|.$$

Pour x dans S_1 (sphère unité de $(\mathbb{K}^n, \|\cdot\|_1)$), on a :

$$\|Ax\|_1 = \sum_{i=1}^{n} \left|\sum_{j=1}^{n} a_{ij} x_j\right| \le \sum_{j=1}^{n} |x_j| \sum_{i=1}^{n} |a_{ij}| \le \beta \sum_{j=1}^{n} |x_j| = \beta.$$

Pour montrer que $\beta \le \|A\|_1$, il suffit de trouver un vecteur x dans S_1 tel que $\|Ax\|_1 = \beta$. Soit k compris entre 1 et n tel que $\beta = \sum_{i=1}^{n} |a_{ik}|$. Pour $x = e_k$ (k-ième vecteur de la base canonique de $\mathbb{K}^n$) on a :

$$\|Ax\|_1 = \sum_{i=1}^{n} \left|\sum_{j=1}^{n} a_{ij} x_j\right| = \sum_{i=1}^{n} |a_{ik}| = \beta.$$

On a donc bien $\|A\|_1 = \beta$. ∎

Remarque 4.1 *: On a $\|A\|_1 = \|A^*\|_\infty$ où A^* désigne la matrice adjointe de A.*

Lemme 4.3 : *Pour toute matrice P dans $GL_n(\mathbb{K})$ l'application :*

$$x \mapsto \|x\|_P = \left\|P^{-1} x\right\|_\infty$$

définit une norme sur $\mathbb{K}^n$ et la norme matricielle induite par cette norme est définie par :

$$\forall A \in \mathcal{M}_n(\mathbb{K}), \quad \|A\|_P = \left\|P^{-1} A P\right\|_\infty.$$

Démonstration : Il est facile de vérifier que $\|\cdot\|_P$ définit une norme sur $\mathbb{K}^n$.

On peut aussi remarquer que P est la matrice de passage de la base canonique de $\mathbb{K}^n$, $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$, à une base $\mathcal{B}' = \{e'_1, e'_2, \ldots, e'_n\}$ et, en notant :

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad X' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

les coordonnées du vecteur x dans les bases respectives $\mathcal{B}$ et $\mathcal{B}'$, on a :

$$\|x\|_P = \left\|P^{-1}X\right\|_\infty = \|X'\|_\infty .$$

Pour toute matrice A dans $\mathcal{M}_n(\mathbb{K})$, on a :

$$\|A\|_P = \sup_{x \in \mathbb{K}^n - \{0\}} \frac{\|Ax\|_P}{\|x\|_P} = \sup_{X \in \mathbb{K}^n - \{0\}} \frac{\left\|P^{-1}AX\right\|_\infty}{\|P^{-1}X\|_\infty},$$

soit, en posant $X = PX'$:

$$\|A\|_P = \sup_{X' \in \mathbb{K}^n - \{0\}} \frac{\left\|P^{-1}APX'\right\|_\infty}{\|X'\|_\infty} = \left\|P^{-1}AP\right\|_\infty . \qquad \blacksquare$$

Le théorème suivant nous donne quelques propriétés des normes matricielles induites par une norme vectorielle.

Théorème 4.2 : *Soit $A \longmapsto \|A\|$ une norme sur $\mathcal{M}_n(\mathbb{K})$ induite par une norme $x \longmapsto \|x\|$ sur $\mathbb{K}^n$. On a :*

$$\forall A \in \mathcal{M}_n(\mathbb{K}), \quad \|A\| = \sup_{x \in B} \|Ax\| = \sup_{x \in \mathbb{K}^n - \{0\}} \frac{\|Ax\|}{\|x\|}. \tag{4.1}$$

$$\forall x \in \mathbb{K}^n, \quad \forall A \in \mathcal{M}_n(\mathbb{K}), \quad \|Ax\| \leq \|A\| \|x\| . \tag{4.2}$$

$$\forall A \in \mathcal{M}_n(\mathbb{K}), \quad \forall B \in \mathcal{M}_n(\mathbb{K}), \quad \|AB\| \leq \|A\| \|B\| . \tag{4.3}$$

$\|I_n\| = 1$.

Il existe x dans S tel que $\|A\| = \|Ax\|$.

$$\|A\| = \inf\left\{\alpha \in \mathbb{R}^+ \mid \forall x \in \mathbb{K}^n, \quad \|Ax\| \leq \alpha \|x\|\right\}. \tag{4.4}$$

Démonstration : L'égalité (4.1) résulte de l'inclusion $S \subset B$ et de la linéarité de A.

L'inégalité (4.2) se déduit facilement de (4.1).

De (4.2) on déduit que pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$ on a :

$$\forall x \in \mathbb{K}^n, \quad \|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\| .$$

D'où l'inégalité (4.3).

Pour tout x dans S on a $\|I_n x\| = \|x\| = 1$. On déduit donc que $\|I_n\| = 1$.

Du fait qu'en dimension finie la sphère unité est compacte et qu'une fonction continue sur un compact admet une borne supérieure qui est atteinte on déduit qu'il existe x dans S tel que $\|A\| = \|Ax\|$.

On note

$$\mathcal{D} = \left\{\alpha \in \mathbb{R}^+ \mid \forall x \in \mathbb{K}^n, \quad \|Ax\| \leq \alpha \|x\|\right\}.$$

On a $\mathcal{D} \neq \emptyset$, car $\|A\| \in \mathcal{D}$. Donc $\mathcal{D}$ admet une borne inférieure comme partie non vide et minorée de $\mathbb{R}^+$ et $\inf(\mathcal{D}) \leq \|A\|$.

Soient $\alpha \in \mathcal{D}$ et x dans S tel que $\|A\| = \|Ax\|$. On a alors :

$$\|A\| = \|Ax\| \leq \alpha \|x\| = \alpha,$$

soit $\|A\| \leq \alpha$. On en déduit donc que $\|A\| \leq \inf(\mathcal{D})$ et $\|A\| = \inf(\mathcal{D})$. ∎

La propriété (4.3) se traduit en disant que toute norme matricielle induite par une norme vectorielle est sous-multiplicative.

En fait, pour toute norme $A \longmapsto \|A\|$ sur $\mathcal{M}_n(\mathbb{K})$ il existe une constante réelle $\lambda > 0$ telle que :

$$\forall (A, B) \in M_n(\mathbb{K}) \times M_n(\mathbb{K}), \quad \|AB\| \leq \lambda \|A\| \|B\|$$

(exercice 4.1).

Il existe quand même des normes qui ne sont pas sous-multiplicatives (exercice 4.2).

Une norme sous-multiplicative n'est pas nécessairement induite par une norme vectorielle. Par exemple la norme de Schur est sous-multiplicative (exercice 3.13) et n'est pas induite par une norme vectorielle du fait que pour tout entier $n \geq 2$ on a $\|I_n\|_s = \sqrt{n} \neq 1$.

2. Le groupe topologique $\mathbf{GL_n(\mathbb{K})}$

L'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ est muni d'une norme quelconque (elles sont toutes équivalentes).

Lemme 4.4 : *L'application* det : $\mathcal{M}_n(\mathbb{K}) \to \mathbb{K}$ *est continue.*

Démonstration : L'application déterminant est continue comme fonction polynomiale des coefficients m_{ij} d'une matrice M. ∎

Théorème 4.3 : *L'ensemble $GL_n(\mathbb{K})$ est un ouvert dense de $\mathcal{M}_n(\mathbb{K})$.*

Démonstration : L'ensemble $GL_n(\mathbb{K})$ est un ouvert de $\mathcal{M}_n(\mathbb{K})$ en tant qu'image réciproque de l'ouvert $\mathbb{K}^*$ par l'application continue déterminant.

La fonction polynomiale $z \longmapsto \det(A - zI_n)$ a au plus n racines dans $\mathbb{K}$. Il existe donc un entier k_0 tel que :

$$\forall k > k_0, \quad \det\left(A - \frac{1}{k} I_n\right) \neq 0$$

et $A = \lim\limits_{\substack{k \to +\infty \\ k > k_0}} A_k$, avec les $A_k = A - \dfrac{1}{k} I_n$ inversibles pour tout $k > k_0$. ∎

Ce théorème peut aussi se montrer en utilisant la caractérisation des matrices de rang r donné (exercice 4.3).

Théorème 4.4 : *L'application $A \longmapsto A^{-1}$ est continue de $GL_n(\mathbb{K})$ dans $GL_n(\mathbb{K})$.*

Démonstration : L'application $M \longmapsto C(M)$ qui associe à une matrice M sa comatrice $C(M)$ est continue de $\mathcal{M}_n(\mathbb{K})$ dans $\mathcal{M}_n(\mathbb{K})$ car ses composantes sont des fonctions polynomiales des coefficients m_{ij}. On en déduit alors que l'application $M \longmapsto M^{-1} = \dfrac{1}{\det(M)} {}^tC(M)$ est continue de $GL_n(\mathbb{K})$ dans $GL_n(\mathbb{K})$. ■

***Remarque 4.2 :** La continuité de $(U, V) \longmapsto UV$ et $M \longmapsto M^{-1}$ peut se traduire en disant que la topologie de $GL_n(\mathbb{K})$ est compatible avec sa structure de groupe. On dit aussi que $GL_n(\mathbb{K})$ est un groupe topologique.*

Théorème 4.5 : *Le centre de $\mathcal{M}_n(\mathbb{K})$ est formé des homothéties et celui de $GL_n(\mathbb{K})$ est formé des homothéties non nulles.*

Démonstration : Soit $A = ((a_{ij}))_{1 \leq i,j \leq n}$ dans le centre de $\mathcal{M}_n(\mathbb{K})$, c'est-à-dire commutant avec toutes les matrices. On a $AE_{ij} = E_{ij}A$ pour tous i, j et, en désignant par $\{e_i \mid 1 \leq i \leq n\}$ la base canonique de $\mathbb{K}^n$:

$$AE_{ij}e_j = Ae_i = \sum_{k=1}^{n} a_{ki}e_k = E_{ij}Ae_j = E_{ij}\left(\sum_{k=1}^{n} a_{kj}e_k\right) = a_{jj}e_i.$$

Donc $a_{ki} = 0$ pour $k \in \{1, \ldots, n\} - \{i\}$ et $a_{ii} = a_{jj}$. C'est-à-dire que $A = \lambda I_n$.

On peut aussi remarquer que si A est dans le centre de $\mathcal{M}_n(\mathbb{K})$, alors A commute à tout projecteur p_x sur la droite $\mathbb{K}x$. On en déduit alors que toutes les droites sont stables par A et A est une homothétie.

Avec la densité de $GL_n(\mathbb{K})$ dans $\mathcal{M}_n(\mathbb{K})$ et la continuité du produit matriciel, on déduit que toute matrice dans le centre de $GL_n(\mathbb{K})$ est aussi dans le centre de $\mathcal{M}_n(\mathbb{K})$. C'est donc une homothétie et son rapport est non nulle puisqu'elle est inversible. ■

***Remarque 4.3 :** Ces résultats sont en fait valables sur tout corps commutatif. Pour le cas de $\mathcal{M}_n(\mathbb{K})$, la démonstration est identique.*

Soit A dans le centre de $GL_n(\mathbb{K})$. Pour tout $B \in \mathcal{M}_n(\mathbb{K})$, le polynôme $\det(B - \lambda I_n)$ a au plus n racines dans $\mathbb{K}$ et il existe un scalaire $\lambda \in \mathbb{K}$ tel que $B - \lambda I_n$ soit inversible. On a alors $A(B - \lambda I_n) = (B - \lambda I_n)A$, c'est-à-dire $AB - \lambda A = BA - \lambda A$ et $AB = BA$. Donc A est dans le centre de $\mathcal{M}_n(\mathbb{K})$.

Corollaire 4.1 : *Les groupes multiplicatifs $GL_n(\mathbb{R})$ et $GL_n(\mathbb{C})$ ne sont pas isomorphes.*

Démonstration : Si $\varphi : GL_n(\mathbb{R}) \to GL_n(\mathbb{C})$ est un isomorphisme de groupes multiplicatifs, alors il induit un isomorphisme du centre de $GL_n(\mathbb{R})$ sur celui de $GL_n(\mathbb{C})$ (dire que A est dans le centre de $GL_n(\mathbb{R})$ équivaut à dire que $AM = MA$ pour tout $M \in GL_n(\mathbb{R})$, ce qui équivaut à :

$$\forall M \in GL_n(\mathbb{R}), \quad \varphi(A)\varphi(M) = \varphi(M)\varphi(A),$$

encore équivalent à :

$$\forall M \in GL_n(\mathbb{C}), \quad \varphi(A) M = M\varphi(A)$$

et donc équivalent au fait que $\varphi(A)$ est dans le centre de $GL_n(\mathbb{C})$). On aurait alors un isomorphisme de groupes multiplicatifs de $\mathbb{R}^*$ sur $\mathbb{C}^*$, ce qui est impossible (i est d'ordre 4 dans $\mathbb{C}^*$ et il n'y a pas d'élément d'ordre 4 dans $\mathbb{R}^*$). ■

Corollaire 4.2 : *Il existe une base de $\mathcal{M}_n(\mathbb{K})$ formée de matrices inversibles.*

Démonstration : $V = \mathrm{Vect}(GL_n(\mathbb{K}))$ est un sous-espace vectoriel fermé de $\mathcal{M}_n(\mathbb{K})$ (on est en dimension finie) qui contient $GL_n(\mathbb{K})$. Il contient donc son adhérence, c'est-à-dire $\mathcal{M}_n(\mathbb{K})$. Donc $V = \mathcal{M}_n(\mathbb{K})$ et du système générateur $GL_n(\mathbb{K})$ on peut extraire une base. ■

Remarque 4.4 *: Ce résultat est encore valable sur un corps commutatif infini. En effet, il existe une infinité de scalaires $\lambda \in \mathbb{K}$ tels que, pour tous i, j dans $\{1, \dots, n\}$, $E_{ij} + \lambda I_n \in GL_n(\mathbb{K})$. Pour un tel $\lambda \in \mathbb{K}$ le système :*

$$\mathcal{B} = \{E_{ij} + \lambda I_n \mid 1 \le i, j \le n\}$$

est une base de $\mathcal{M}_n(\mathbb{K})$. En effet l'égalité :

$$\sum_{1 \le i,j \le n} a_{ij}(E_{ij} + \lambda I_n) = 0$$

entraîne $a_{ij} = 0$ pour $1 \le i \ne j \le n$ et

$$\sum_{i=1}^{n} \left(a_{ii} + \lambda \sum_{j=1}^{n} a_{jj} \right) E_{ii} = 0$$

($I_n = \sum_{i=1}^{n} E_{ii}$), donc $a_{ii} + \lambda \sum_{j=1}^{n} a_{jj} = 0$ pour tout $i \in \{1, \dots, n\}$ et en sommant on a $(1 + n\lambda) \sum_{j=1}^{n} a_{jj} = 0$, ce qui donne $\sum_{j=1}^{n} a_{jj} = 0$, en choisissant $\lambda \in \mathbb{K}$ tel que $1 + n\lambda \ne 0$ et $a_{ii} = 0$ pour tout i. Le système $\{E_{ij} + \lambda I_n \mid 1 \le i, j \le n\}$ est donc libre et c'est une base car il est formé de n^2 éléments.

Corollaire 4.3 : *Pour $n > 1$, il n'existe pas de norme sur $\mathcal{M}_n(\mathbb{K})$ telle que $\left\|P^{-1}AP\right\| = \|A\|$ pour tout A dans $\mathcal{M}_n(\mathbb{K})$ et tout P dans $GL_n(\mathbb{K})$.*

Démonstration : Supposons que pour toutes matrices A dans $\mathcal{M}_n(\mathbb{K})$ et P dans $GL_n(\mathbb{K})$ on ait $\|P^{-1}AP\| = \|A\|$. En appliquant cette égalité à PA et P, on a $\|AP\| = \|PA\|$. Par densité de $GL_n(\mathbb{K})$ dans $\mathcal{M}_n(\mathbb{K})$ et continuité du produit matriciel, on déduit que $\|AP\| = \|PA\|$ pour tous A, P dans $\mathcal{M}_n(\mathbb{K})$. Mais ce dernier résultat est impossible pour $n > 1$. En effet $E_{12}E_{11} = 0$ et $E_{11}E_{12} \neq 0$. ■

De la densité de $GL_n(\mathbb{K})$ dans $\mathcal{M}_n(\mathbb{K})$, on peut déduire une généralisation à $\mathcal{M}_n(\mathbb{K})$ du théorème de décomposition polaire des matrices inversibles (corollaire 3.15). Pour ce faire on utilise le lemme suivant.

Lemme 4.5 : *L'ensemble $\mathcal{U}_n(\mathbb{C})$ [resp. $\mathcal{O}_n(\mathbb{R})$] des matrices complexes unitaires [resp. réelles orthogonales] est compact dans $\mathcal{M}_n(\mathbb{C})$ [resp. $\mathcal{M}_n(\mathbb{R})$].*

Démonstration : On munit l'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ de la norme matricielle $\|\cdot\|_2$ induite par la norme hermitienne ou euclidienne de $\mathbb{K}^n$. Du fait qu'une transformation unitaire ou orthogonale conserve la norme hermitienne ou euclidienne de $\mathbb{K}^n$ on déduit que pour toute matrice A unitaire ou orthogonale on a $\|A\|_2 = 1$. On déduit donc que $\mathcal{U}_n(\mathbb{C})$ [resp. $\mathcal{O}_n(\mathbb{R})$] est borné dans $(\mathcal{M}_n(\mathbb{K}), \|\cdot\|_2)$. De plus cet ensemble est fermé comme image réciproque du fermé $\{I_n\}$ de $\mathcal{M}_n(\mathbb{K})$ par l'application continue $A \mapsto A^*A$. En conclusion, $\mathcal{U}_n(\mathbb{C})$ [resp. $\mathcal{O}_n(\mathbb{R})$] est compact dans $(\mathcal{M}_n(\mathbb{K}), \|\cdot\|_2)$ en tant que fermé borné (on est en dimension finie). ■

Corollaire 4.4 : *Toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ [resp. $A \in \mathcal{M}_n(\mathbb{R})$] peut s'écrire $A = UH$ [resp. $A = \Omega S$] où U [resp. Ω] est une matrice unitaire [resp. orthogonale] et H [resp. S] une matrice hermitienne [resp. symétrique] positive.*

Démonstration : Toute matrice $A \in \mathcal{M}_n(\mathbb{K})$ peut s'écrire $A = \lim\limits_{k\to+\infty} A_k$ où $(A_k)_{k\in\mathbb{N}}$ est une suite de matrices inversibles. Avec le théorème de décomposition polaire (corollaire 3.15), on peut écrire, pour tout entier k, $A_k = U_k H_k$ [resp. $A_k = \Omega_k S_k$] où $(U_k)_{k\in\mathbb{N}}$ [resp. $(\Omega_k)_{k\in\mathbb{N}}$] est une suite de matrices unitaires [resp. orthogonales] et $(H_k)_{k\in\mathbb{N}}$ [resp. $(S_k)_{k\in\mathbb{N}}$] une suite de matrices hermitiennes [resp. symétriques] définies positives. De la suite $(U_k)_{k\in\mathbb{N}}$ [resp. $(\Omega_k)_{k\in\mathbb{N}}$] dans le compact $\mathcal{U}_n(\mathbb{C})$ [resp. $\mathcal{O}_n(\mathbb{R})$] on peut extraire une sous-suite $\left(U_{\varphi(k)}\right)_{k\in\mathbb{N}}$ [resp. $\left(\Omega_{\varphi(k)}\right)_{k\in\mathbb{N}}$] qui converge vers une matrice $U \in \mathcal{U}_n(\mathbb{C})$ [resp. $\Omega \in \mathcal{O}_n(\mathbb{R})$]. De $H_k = U_k^{-1}A_k = U_k^* A_k$ [resp. $S_k = \Omega_k^{-1}A_k = {}^t\Omega_k A_k$] et de la continuité du produit matriciel, on déduit que la suite $\left(H_{\varphi(k)}\right)_{k\in\mathbb{N}}$ [resp. $\left(S_{\varphi(k)}\right)_{k\in\mathbb{N}}$] est convergente. La limite H [resp. S] de cette suite est une matrice hermitienne [resp. symétrique] positive et $A = UH$ [resp. $A = \Omega S$]. ■

Remarque 4.5 *: Si A est de rang $r < n$, alors la décomposition ci-dessus n'est pas unique. En effet, on peut diagonaliser la matrice hermitienne [resp. symétrique] positive H [resp. S] dans une base orthonormée $\{e_i \mid 1 \leq i \leq n\}$ avec $He_i = \lambda_i e_i$ [resp. $Se_i = \lambda_i e_i$] pour $1 \leq i \leq n$ où $\lambda_i = 0$ pour $1 \leq i \leq n-r$ et $\lambda_i > 0$ sinon (si A n'est pas inversible alors il en est de même de H [resp. S] et 0 est valeur propre de H [resp. S]). Les Ue_i [resp. Ωe_i] sont alors uniquement déterminés pour $n-r+1 \leq i \leq n$, mais pour $1 \leq i \leq n-r$ il n'y a pas unicité.*

En désignant par $\mathcal{H}_n^{++}(\mathbb{C})$ [resp. $\mathcal{S}_n^{++}(\mathbb{R})$] l'ensemble des matrices complexes [resp. réelles] d'ordre n hermitiennes [resp. symétriques] définies positives, le théorème de décomposition polaire des matrices inversibles peut s'exprimer comme suit.

Théorème 4.6 : *L'application $(U,H) \longmapsto UH$ [resp. $(\Omega, S) \longmapsto \Omega S$] réalise un homéomorphisme de $\mathcal{U}_n(\mathbb{C}) \times \mathcal{H}_n^{++}(\mathbb{C})$ [resp. $\mathcal{O}_n(\mathbb{R}) \times \mathcal{S}_n^{++}(\mathbb{R})$] sur $GL_n(\mathbb{C})$ [resp. $GL_n(\mathbb{R})$].*

Démonstration : On fait la démonstration dans le cas des matrices réelles inversibles, le cas complexe se traitant de manière analogue.

On sait déjà que toute matrice $A \in GL_n(\mathbb{R})$ s'écrit de manière unique $A = \Omega S$ avec $\Omega \in \mathcal{O}_n(\mathbb{R})$ et $S \in S_n^{++}(\mathbb{R})$. On déduit que l'application $\varphi : (\Omega, S) \longmapsto \Omega S$ réalise une bijection de $\mathcal{O}_n(\mathbb{R}) \times S_n^{++}(\mathbb{R})$ sur $GL_n(\mathbb{R})$.

Cette application est continue car ses composantes sont des fonctions polynomiales des coefficients ω_{ij} de Ω et s_{rp} de S. Il reste à montrer que φ^{-1} est continue.

Soit $(A_k)_{k\in\mathbb{N}}$ une suite de matrices dans $GL_n(\mathbb{R})$ qui converge vers A. On note $\varphi^{-1}(A_k) = (\Omega_k, S_k)$ pour tout $k \in \mathbb{N}$ et $\varphi^{-1}(A) = (\Omega, S)$. De la suite $(\Omega_k)_{k\in\mathbb{N}}$ dans le compact $\mathcal{O}_n(\mathbb{R})$ on peut extraire une sous-suite $\left(\Omega_{\varphi(k)}\right)_{k\in\mathbb{N}}$ qui converge vers une matrice $\Omega' \in \mathcal{O}_n(\mathbb{R})$. De $S_k = {}^t\Omega_k A_k$, on déduit que la suite $\left(S_{\varphi(k)}\right)_{k\in\mathbb{N}}$ converge vers $S' = {}^t\Omega' A$. La matrice S' est symétrique positive comme limite d'une suite de matrices symétriques positives et elle est définie puisque inversible.

On a alors la décomposition polaire $A = \Omega' S'$. Cette dernière décomposition étant unique, on a nécessairement $\Omega' = \Omega$.

On a donc ainsi montré que la suite $(\Omega_k)_{k\in\mathbb{N}}$ a une unique valeur d'adhérence dans le compact $\mathcal{O}_n(\mathbb{R})$. Elle converge donc vers Ω et $(S_k)_{k\in\mathbb{N}} = ({}^t\Omega_k A_k)_{k\in\mathbb{N}}$ converge vers ${}^t\Omega A = S$. C'est-à-dire que la suite $\left((\Omega_k, S_k)\right)_{k\in\mathbb{N}} = \left(\varphi^{-1}(A_k)\right)_{k\in\mathbb{N}}$ converge vers $(\Omega, S) = \varphi^{-1}(A)$ et φ^{-1} est continue. ■

Théorème 4.7 : *L'ensemble $GL_n(\mathbb{C})$ est connexe par arcs.*

Démonstration : Soient $A \neq B$ dans $GL_n(\mathbb{C})$. Le polynôme

$$P(z) = \det\left((1-z)A + zB\right)$$

est de degré inférieur ou égal à n. Il a donc au plus n racines distinctes $z_1, \ldots, z_p$ dans $\mathbb{C}$. On note $\Omega = \mathbb{C} - \{z_1, \ldots, z_p\}$.

On a $\{0,1\} \subset \Omega$ car $P(0) = \det(A) \neq 0$ et $P(1) = \det(B) \neq 0$. De plus parmi les droites $D_\alpha = \left\{z = \rho e^{i\alpha}; \rho \in \mathbb{R}\right\}$ et $\Delta_\beta = \left\{z = 1 + \rho e^{i\beta}; \rho \in \mathbb{R}\right\}$ avec $0 \leq \alpha, \beta < \pi$, il en existe une infinité qui ne rencontrent pas $\{z_1, \ldots, z_p\}$ et telles que $D_\alpha \cap \Delta_\beta \neq \emptyset$ (Fig. 4.1, page suivante).

En choisissant deux telles droites, le chemin formé de la juxtaposition des segments $[0,\gamma]$ et $[\gamma, 1]$ permet alors de joindre continûment les points 0 et 1 dans Ω. On note $\varphi : [0,1] \to \Omega$ ce chemin et on définit $\psi : [0,1] \to \mathcal{M}_n(\mathbb{C})$ par :

$$\forall t \in [0,1], \quad \psi(t) = \left(1 - \varphi(t)\right)A + \varphi(t) B.$$

Cette application est continue et $P\left(\psi(t)\right) \neq 0$ pour tout $t \in [0,1]$ ($\psi(t)$ est dans Ω), c'est-à-dire que ψ est une fonction continue qui relie A et B dans $GL_n(\mathbb{C})$ ($\psi(0) = A$ et $\psi(1) = B$).

On a donc ainsi prouvé que $GL_n(\mathbb{C})$ est connexe par arcs. ■

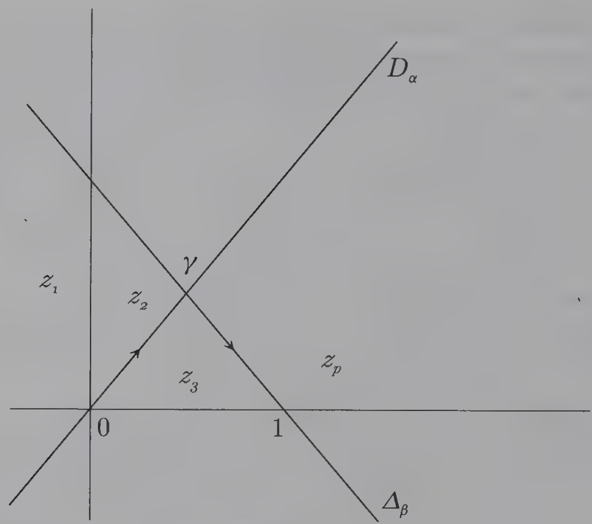

Figure 4.1.

Remarque 4.6 : *On peut aussi montrer la connexité de $GL_n(\mathbb{C})$ en utilisant le théorème de trigonalisation des matrices complexes (exercice 4.8).*

Remarque 4.7 : *L'ensemble $GL_n(\mathbb{R})$ n'est pas connexe (exercice 4.9).*

Des résultats sur la réduction des matrices unitaires complexes et orthogonales réelles on déduit les résultats de connexité suivants.

Théorème 4.8 : *Le sous-groupe $U_n(\mathbb{C})$ de $GL_n(\mathbb{C})$ formé des matrices unitaires est connexe par arcs.*

Démonstration : Une matrice unitaire a toutes ses valeurs propres de module 1 et se diagonalise dans une base orthonormée. Pour tout $A \in U_n(\mathbb{C})$ il existe donc une matrice diagonale :

$$D = \begin{pmatrix} e^{i\theta_1} & 0 & \dots & 0 \\ 0 & e^{i\theta_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & e^{i\theta_n} \end{pmatrix}$$

et une matrice unitaire $U \in U_n(\mathbb{C})$ telles que $A = UDU^*$. En posant, pour tout $t \in [0,1]$, $\gamma(t) = UD(t)U^*$, où :

$$D(t) = \begin{pmatrix} e^{it\theta_1} & 0 & \dots & 0 \\ 0 & e^{it\theta_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & e^{it\theta_n} \end{pmatrix},$$

on définit un chemin continu dans $U_n(\mathbb{C})$ qui relie I_n et A. Ce qui prouve la connexité par arcs de $U_n(\mathbb{C})$. ∎

Théorème 4.9 : *Les composantes connexes du sous-groupe $\mathcal{O}_n(\mathbb{R})$ de $GL_n(\mathbb{R})$ formé des matrices orthogonales sont $\mathcal{O}_n^+(\mathbb{R})$ et $\mathcal{O}_n^-(\mathbb{R})$.*

Démonstration : On sait que pour toute matrice orthogonale A, il existe une matrice Ω dans $\mathcal{O}_n(\mathbb{R})$ et une matrice :

$$D = \begin{pmatrix} I_p & 0 & 0 & \dots & 0 \\ 0 & -I_q & 0 & \ddots & \vdots \\ 0 & 0 & R(\theta_1) & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 & R(\theta_r) \end{pmatrix},$$

où pour tout $k \in \{1, \dots, r\}$, $R(\theta_k) = \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix}$ avec θ_k dans $]0, 2\pi[- \{\pi\}$, telles que $A = \Omega D^t\Omega$. Si de plus $A \in \mathcal{O}_n^+(\mathbb{R})$ alors q est nécessairement pair et la matrice D peut s'écrire :

$$D = \begin{pmatrix} I_p & 0 & 0 \\ 0 & D_1 & 0 \\ 0 & 0 & D_2 \end{pmatrix},$$

avec :

$$D_1 = \begin{pmatrix} R(\alpha_1) & 0 & \dots & 0 \\ 0 & R(\alpha_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R(\alpha_{q'}) \end{pmatrix}, \ D_2 = \begin{pmatrix} R(\theta_1) & 0 & \dots & 0 \\ 0 & R(\theta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R(\theta_r) \end{pmatrix},$$

avec $\alpha_j = \pi$ pour tout j. En posant, pour tout $t \in [0, 1]$, $\gamma(t) = \Omega D(t)^t\,\Omega$, où :

$$D(t) = \begin{pmatrix} I_p & 0 & 0 \\ 0 & D_1(t) & 0 \\ 0 & 0 & D_2(t) \end{pmatrix},$$

avec :

$$D_1(t) = \begin{pmatrix} R(t\alpha_1) & 0 & \dots & 0 \\ 0 & R(t\alpha_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R(t\alpha_{q'}) \end{pmatrix}, \ D_2(t) = \begin{pmatrix} R(t\theta_1) & 0 & \dots & 0 \\ 0 & R(t\theta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & R(t\theta_r) \end{pmatrix},$$

on définit un chemin continu dans $\mathcal{O}_n^+(\mathbb{R})$ qui relie I_n et A. Ce qui suffit à prouver la connexité par arcs de $\mathcal{O}_n^+(\mathbb{R})$.

Pour toute matrice $A \in \mathcal{O}_n^-(\mathbb{R})$ (par exemple $A = \begin{pmatrix} I_{n-1} & 0 \\ 0 & -1 \end{pmatrix}$) l'application $M \longmapsto AM$ réalise un homéomorphisme de $\mathcal{O}_n^+(\mathbb{R})$ sur $\mathcal{O}_n^-(\mathbb{R})$. On en déduit alors que $\mathcal{O}_n^-(\mathbb{R})$ est connexe par arcs.

On a $\mathcal{O}_n(\mathbb{R}) = \mathcal{O}_n^+(\mathbb{R}) \cup \mathcal{O}_n^-(\mathbb{R})$, avec $\mathcal{O}_n^+(\mathbb{R})$ et $\mathcal{O}_n^-(\mathbb{R})$ fermés (images réciproques de 1 et -1 respectivement par l'application déterminant) connexes disjoints dans $\mathcal{O}_n(\mathbb{R})$. Ce sont donc les composantes connexes de $\mathcal{O}_n(\mathbb{R})$. ∎

3. Propriétés topologiques de l'ensemble des matrices diagonalisables de $\mathcal{M}_n(\mathbb{C})$

Définition 4.2 : *Si $P(X) = \sum_{k=0}^{n} a_k X^k$ et $Q(X) = \sum_{k=0}^{m} b_k X^k$ sont deux polynômes non nuls dans $\mathbb{K}[X]$ avec $a_n \neq 0$ et $b_m \neq 0$, on appelle matrice de Sylvester de P et Q la matrice du système de vecteurs :*

$$\{P, XP, \ldots, X^{m-1}P, Q, XQ, \ldots, X^{n-1}Q\}$$

dans la base canonique de $\mathbb{K}_{n+m-1}[X]$. On note $S(P,Q)$ cette matrice, et son déterminant est appelé le résultant de P et Q et est noté $\operatorname{res}(P,Q)$.

Lemme 4.6 : *Soient P et Q deux polynômes non nuls dans $\mathbb{C}[X]$. Ces polynômes ont une racine commune dans $\mathbb{C}$ si et seulement si il existe deux polynômes non nuls U et V tels que $\deg(U) < \deg(Q)$, $\deg(V) < \deg(P)$ et $UP + VQ = 0$.*

Démonstration : Si λ est une racine commune à P et Q, il existe alors deux polynômes non nuls P_0 et Q_0 tels que :

$$P(X) = (X-\lambda) P_0(X), \quad Q(X) = (X-\lambda) Q_0(X).$$

En posant $U = Q_0$, $V = -P_0$, on a $UP + VQ = 0$ avec $\deg(U) = \deg(Q) - 1$ et $\deg(V) = \deg(P) - 1$.

Réciproquement, supposons qu'il existe deux polynômes non nuls U et V tels que $UP = -VQ$ avec $\deg(U) < \deg(Q)$ et $\deg(V) < \deg(P)$. Si P et Q sont premiers entre eux alors P divise V (théorème de Gauss), ce qui est impossible du fait que $\deg(V) < \deg(P)$. P et Q ne sont donc pas premiers entre eux et nécessairement ils ont une racine commune dans $\mathbb{C}$. ∎

Théorème 4.10 : *Deux polynômes non nuls P et Q ont une racine commune dans $\mathbb{C}$ si et seulement si $\operatorname{res}(P,Q) = 0$.*

Démonstration : En reprenant la démonstration du lemme précédent, on pose :

$$U(X) = \sum_{k=0}^{m-1} u_k X^k, \quad V(X) = \sum_{k=0}^{n-1} v_k X^k,$$

où n est le degré de P et m celui de Q. La condition $UP + VQ = 0$ avec U et V non nuls est équivalente à :

$$\sum_{k=0}^{m-1} u_k X^k P(X) + \sum_{k=0}^{n-1} v_k X^k Q(X) = 0,$$

avec les u_k et v_k non tous nuls, ce qui équivaut à dire que le système :

$$\left\{X^k P(X),\ X^j Q(X) \mid 0 \leq k \leq m-1,\ 0 \leq j \leq n-1\right\}$$

est lié dans $\mathbb{C}_{n+m-1}[X]$, ce qui est encore équivalent à dire que le résultant de P et Q est nul.

En conclusion, P et Q ont une racine commune dans $\mathbb{C}$ si et seulement si $\operatorname{res}(P, Q) = 0$. ■

On désigne par $\mathcal{D}'_n(\mathbb{C})$ l'ensemble des matrices $M \in \mathcal{M}_n(\mathbb{C})$ ayant n valeurs propres distinctes dans $\mathbb{C}$ et par $\mathcal{D}_n(\mathbb{C})$ l'ensemble des matrices diagonalisables de $\mathcal{M}_n(\mathbb{C})$.

Théorème 4.11 : *L'ensemble $\mathcal{D}'_n(\mathbb{C})$ est l'intérieur de $\mathcal{D}_n(\mathbb{C})$.*

Démonstration : Une matrice $M \in \mathcal{M}_n(\mathbb{C})$ est dans $\mathcal{D}'_n(\mathbb{C})$ si et seulement si son polynôme caractéristique P_M n'a que des racines simples dans $\mathbb{C}$, ce qui équivaut à dire que $\varphi(M) = \operatorname{res}(P_M, P'_M) \neq 0$. L'ensemble $\mathcal{D}'_n(\mathbb{C})$ est donc un ouvert de $\mathcal{M}_n(\mathbb{C})$ comme image réciproque de l'ouvert $\mathbb{C}^*$ par l'application continue φ (φ est une fonction polynomiale des coefficients de M).

Une matrice ayant n valeurs propres distinctes est diagonalisable, on a donc les inclusions $\mathcal{D}'_n(\mathbb{C}) \subset \mathcal{D}_n(\mathbb{C})$ et $\mathcal{D}'_n(\mathbb{C}) \subset \overset{\circ}{\mathcal{D}}_n(\mathbb{C})$ puisque $\mathcal{D}'_n(\mathbb{C})$ est ouvert.

Supposons qu'il existe $A \in \overset{\circ}{\mathcal{D}}_n(\mathbb{C})$ ayant une valeur propre λ d'ordre supérieur ou égal à 2. On peut alors trouver une matrice inversible P telle que $A = PDP^{-1}$ où :

$$D = \begin{pmatrix} \lambda & 0 & 0 & \dots & 0 \\ 0 & \lambda & 0 & \dots & 0 \\ 0 & 0 & \lambda_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Pour tout entier $k > 0$, on pose :

$$\Delta_k = \begin{pmatrix} \lambda & \dfrac{1}{k} \\ 0 & \lambda \end{pmatrix}, \quad D_k = \begin{pmatrix} \Delta_k & 0 & \dots & 0 \\ 0 & \lambda_3 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Le polynôme minimal de D_k est un multiple de celui de Δ_k, c'est-à-dire de $(X-\lambda)^2$ (si $P(D_k)=0$, alors $P(\Delta_k)=0$ et P est un multiple de π_{Δ_k}). En conséquence, la matrice D_k et la matrice $A_k = PD_kP^{-1}$ ne sont pas diagonalisables (une matrice est diagonalisable si et seulement si son polynôme minimal est scindé à racines simples). Comme $A = \lim\limits_{k\to+\infty} A_k$, on ne peut pas avoir $A \in \overset{\circ}{\mathcal{D}}_n(\mathbb{C})$. Donc toutes les matrices de $\overset{\circ}{\mathcal{D}}_n(\mathbb{C})$ ont n valeurs propres distinctes et $\overset{\circ}{\mathcal{D}}_n(\mathbb{C}) \subset \mathcal{D}'_n(\mathbb{C})$.

En définitive, $\overset{\circ}{\mathcal{D}}_n(\mathbb{C}) = \mathcal{D}'_n(\mathbb{C})$. ■

Théorème 4.12 : *Les ensembles $\mathcal{D}'_n(\mathbb{C})$ et $\mathcal{D}_n(\mathbb{C})$ sont denses dans $\mathcal{M}_n(\mathbb{C})$.*

Démonstration : Toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ est semblable à une matrice triangulaire supérieure, c'est-à-dire qu'il existe une matrice P inversible et une matrice triangulaire $T = \begin{pmatrix} t_{11} & t_{12} & \dots & t_{1n} \\ 0 & t_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & t_{n-1,n-1} \\ 0 & \dots & 0 & t_{nn} \end{pmatrix}$ telles que $A = PTP^{-1}$. On pose alors :

$$\alpha = \begin{cases} 1 & \text{si } t_{ii} = t_{jj} \\ & \text{pour tous } i \neq j \text{ dans } \{1,\dots,n\}, \\ \inf\{|t_{ii}-t_{jj}|\,;1 \leq i,j \leq n,\ t_{ii} \neq t_{jj}\} & \text{sinon,} \end{cases}$$

et on définit la suite de matrices $(T_k)_{k\geq 1}$ par $T_k = T + \Delta_k$, où :

$$\Delta_k = \begin{pmatrix} \frac{\alpha}{k} & 0 & \dots & 0 \\ 0 & \frac{\alpha}{2k} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \frac{\alpha}{nk} \end{pmatrix}.$$

Pour tout entier $k>0$ la matrice T_k a toutes ses valeurs propres distinctes (si $t_{ii} + \dfrac{\alpha}{ik} = t_{jj} + \dfrac{\alpha}{jk}$ avec $t_{ii} \neq t_{jj}$, alors :

$$|t_{ii}-t_{jj}| = \frac{\alpha}{k}\left|\frac{1}{i}-\frac{1}{j}\right| < \frac{\alpha}{k} \leq \alpha,$$

ce qui contredit la définition de α) et donc $T_k \in \mathcal{D}'_n(\mathbb{C})$ et elle est en particulier diagonalisable. On a alors, avec la continuité du produit matriciel, $A = \lim\limits_{k\to+\infty} A_k$, où pour tout $k>0$ la matrice $A_k = PT_kP^{-1}$ est dans $\mathcal{D}'_n(\mathbb{C})$ et diagonalisable. D'où la densité de $\mathcal{D}'_n(\mathbb{C})$ et $\mathcal{D}_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$. ■

Remarque 4.8 : *L'ensemble $\mathcal{D}_n(\mathbb{R})$ des matrices diagonalisables de $\mathcal{M}_n(\mathbb{R})$ n'est pas dense dans $\mathcal{M}_n(\mathbb{R})$ (exercice 4.11).*

De manière plus précise on peut montrer que l'adhérence de $\mathcal{D}_n(\mathbb{R})$ est l'ensemble $\mathcal{T}_n(\mathbb{R})$ des matrices trigonalisables de $\mathcal{M}_n(\mathbb{R})$ (exercice 4.12).

L'application qui associe à une matrice $M \in \mathcal{M}_n(\mathbb{K})$ son polynôme caractéristique est continue (les coefficients de ce polynôme sont des fonctions polynomiales des m_{ij}). On peut alors se poser la question de savoir s'il en est de même pour l'application qui associe à une matrice son polynôme minimal. De la densité de $\mathcal{D}'_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$ on déduit que la réponse est négative.

Corollaire 4.5 : *Pour $n \geq 2$, l'application qui associe à une matrice $A \in \mathcal{M}_n(\mathbb{C})$ son polynôme minimal n'est pas continue.*

Démonstration : Supposons que l'application $A \longmapsto \pi_A$, qui associe à une matrice A de $\mathcal{M}_n(\mathbb{C})$ son polynôme minimal, soit continue. Avec la densité de $\mathcal{D}'_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$, on peut écrire que $I_n = \lim\limits_{k \to +\infty} A_k$ où $(A_k)_{k \in \mathbb{N}}$ est une suite de matrices à valeurs propres deux à deux distinctes. Avec la continuité de $A \longmapsto \pi_A$, on aurait alors $\pi_{I_n} = \lim\limits_{k \to +\infty} \pi_{A_k}$. Mais pour toute matrice M dans $\mathcal{D}'_n(\mathbb{C})$ on a $\pi_M = (-1)^n P_M$ (polynôme caractéristique). On aurait alors, avec la continuité de $A \longmapsto P_A$:

$$(X-1) = \pi_{I_n} = \lim_{k \to +\infty} \pi_{A_k} = \lim_{k \to +\infty} P_{A_k} = P_{I_n} = (X-1)^n,$$

ce qui est impossible pour $n > 1$.

Donc, pour $n > 1$, l'application $A \longmapsto \pi_A$ n'est pas continue. ∎

4. Rayon spectral d'une matrice complexe

Définition 4.3 : *Soit A une matrice d'ordre n à coefficients complexes. Le rayon spectral de A est le réel :*

$$\rho(A) = \max_{\lambda \in \mathrm{sp}(A)} |\lambda|.$$

Le rayon spectral $\rho(A)$ est le rayon du plus petit disque centré en 0 du plan complexe contenant toutes les valeurs propres de la matrice A.

Lemme 4.7 : *Si $A \in \mathcal{M}_n(\mathbb{C})$ est une matrice normale alors :*

$$\|A\|_2 = \rho(A).$$

Démonstration : Une matrice normale $A \in \mathcal{M}_n(\mathbb{C})$ se diagonalise dans une base orthonormée. Il existe donc des scalaires $\{\lambda_1, \ldots, \lambda_n\}$ et une base orthonormée de $\mathbb{C}^n$ $\{e_1, \ldots, e_n\}$ tels que $Ae_k = \lambda_k e_k$ pour tout $k \in \{1, \ldots, n\}$. Pour tout $x = \sum\limits_{k=1}^{n} x_k e_k \in \mathbb{C}^n$ tel que $\|x\|_2^2 = \sum\limits_{k=1}^{n} |x_k|^2 = 1$, on a alors :

$$\|Ax\|_2^2 = \sum_{k=1}^{n} |x_k|^2 |\lambda_k|^2 \leq \rho(A)^2 \sum_{k=1}^{n} |x_k|^2 = \rho(A)^2.$$

On peut donc conclure que $\|A\|_2 \leq \rho(A)$.

Si $k \in \{1, \ldots, n\}$ est tel que $\rho(A) = |\lambda_k|$, alors $\rho(A) = |\lambda_k| = \|Ae_k\|_2$ avec $\|e_k\|_2 = 1$, Donc $\|A\|_2 = \rho(A)$. ∎

L'égalité $\|A\|_2 = \rho(A)$ est valable en particulier pour A complexe hermitienne ou unitaire et pour A réelle symétrique ou orthogonale.

Théorème 4.13 : *Pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, on a :*

$$\|A\|_2 = \sqrt{\|A^*A\|_2} = \sqrt{\rho(A^*A)}.$$

Démonstration : Pour tout $x \in \mathbb{C}^n$ tel que $\|x\|_2^2 = 1$, en utilisant l'inégalité de Cauchy-Schwarz, on a :

$$\begin{aligned}\|Ax\|_2^2 = \langle x \mid A^*Ax\rangle &\leq \|x\|_2 \|A^*Ax\|_2 \\ &\leq \|x\|_2 \|A^*A\|_2 \|x\|_2 = \|A^*A\|_2,\end{aligned}$$

ce qui entraîne $\|A\|_2^2 \leq \|A^*A\|_2 \leq \|A\|_2 \|A^*\|_2$ et $\|A\|_2 \leq \|A^*\|_2$. En appliquant cette inégalité à A^*, on obtient $\|A^*\|_2 \leq \|A\|_2$, ce qui donne :

$$\|A^*\|_2 = \|A\|_2.$$

On déduit donc que :

$$\|A\|_2^2 \leq \|A^*A\|_2 \leq \|A\|_2 \|A^*\|_2 = \|A\|_2^2,$$

ce qui entraîne $\|A\|_2^2 = \|A^*A\|_2$. La matrice A^*A étant normale (elle est hermitienne), on a aussi $\|A^*A\|_2 = \rho(A^*A)$. Donc :

$$\|A\|_2 = \sqrt{\|A^*A\|_2} = \sqrt{\rho(A^*A)}.$$ ■

Définition 4.4 : *Pour toute matrice $A \in \mathcal{M}_n(\mathbb{K})$, les racines carrées positives des valeurs propres de A^*A sont appelées les valeurs singulières de A.*

Remarque 4.9 : *On peut définir la norme matricielle subordonnée à la norme euclidienne $\|\cdot\|_{2,\mathbb{R}}$ sur $\mathbb{R}^n$ par :*

$$\forall A \in \mathcal{M}_n(\mathbb{R}), \quad \|A\|_{2,\mathbb{R}} = \sup_{\|x\|_{2,\mathbb{R}}=1} \|Ax\|_{2,\mathbb{R}}.$$

A priori *cette norme pourrait être différente de celle induite par la norme euclidienne $\|\cdot\|_{2,\mathbb{C}}$ de $\mathbb{C}^n$ que nous avons définie par :*

$$\|A\|_{2,C} = \sup_{\|z\|_{2,\mathbb{C}}=1} \|Az\|_{2,\mathbb{C}},$$

où $\|z\|_{2,\mathbb{C}}^2 = \|x\|_{2,\mathbb{R}}^2 + \|y\|_{2,\mathbb{R}}^2$ pour tout $z = x + iy \in \mathbb{C}^n$.

En fait, il n'en est rien. En effet, pour tout $x \in \mathbb{R}^n$ on a $\|x\|_{2,\mathbb{C}} = \|x\|_{2,\mathbb{R}}$ et $\|Ax\|_{2,\mathbb{C}} = \|Ax\|_{2,\mathbb{R}}$, donc $\|A\|_{2,\mathbb{R}} \leq \|A\|_{2,\mathbb{C}}$. Et pour $z = x + iy \in \mathbb{C}^n$, on a :

$$\|Az\|_{2,\mathbb{C}}^2 = \|Ax\|_{2,\mathbb{R}}^2 + \|Ay\|_{2,\mathbb{R}}^2 \leq \|A\|_{2,\mathbb{R}}^2 \left(\|x\|_{2,\mathbb{R}}^2 + \|y\|_{2,\mathbb{R}}^2\right) = \|A\|_{2,\mathbb{R}}^2 \|z\|_{2,\mathbb{C}}^2,$$

donc $\|A\|_{2,\mathbb{C}} \leq \|A\|_{2,\mathbb{R}}$.

Lemme 4.8 : *Si pour tout réel $\delta > 0$, on note :*

$$D_\delta = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \delta & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \delta^{n-1} \end{pmatrix}$$

alors pour toute matrice $A = ((a_{ij}))_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{K})$ on a :

$$D_\delta^{-1} A D_\delta = \left(\delta^{j-i} a_{ij}\right)_{1 \leq i,j \leq n}.$$

Démonstration : La multiplication à droite par D_δ a pour effet de multiplier la colonne numéro j par δ^{j-1} et la multiplication à gauche par D_δ^{-1} a pour effet de diviser la ligne numéro i par δ^{i-1}. Il en résulte que le coefficient d'indice (i,j) de $D_\delta^{-1} A D_\delta$ est $\delta^{j-1}\delta^{1-i} a_{ij} = \delta^{j-i} a_{ij}$. ∎

Lemme 4.9 : *Soit A une matrice d'ordre n à coefficients complexes. Pour tout réel $\epsilon > 0$ il existe une matrice d'ordre n à coefficients complexes et inversible P_ϵ telle que la matrice $T_\epsilon = P_\epsilon^{-1} A P_\epsilon$ soit triangulaire supérieure avec :*

$$T_\epsilon = ((t_{ij}))_{1 \leq i,j \leq n}, \quad \max_{1 \leq i \leq n-1} \sum_{j=i+1}^{n} |t_{ij}| < \epsilon.$$

Démonstration : On peut trouver une matrice inversible P à coefficients complexes telle que $T = P^{-1} A P$ soit triangulaire supérieure. Pour tout réel $\delta > 0$, on pose :

$$D_\delta = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \delta & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \delta^{n-1} \end{pmatrix}$$

et on a alors :

$$T_\delta = D_\delta^{-1} T D_\delta = \begin{pmatrix} t_{11} & \delta t_{12} & \dots & \delta^{n-1} t_{1n} \\ 0 & t_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \delta t_{n-1,n-1} \\ 0 & \dots & 0 & t_{nn} \end{pmatrix}.$$

La matrice T_δ est semblable à la matrice A et, en notant $T_\delta = \left((t'_{ij})\right)_{1 \leq i,j \leq n}$, on a $\lim\limits_{\delta \to 0} t'_{ij} = 0$ pour $1 \leq i < j \leq n$, et on peut donc choisir $\delta > 0$ tel que :

$$\forall i \in \{1, \dots, n\}, \quad \sum_{j=i+1}^{n} \left|t'_{ij}\right| < \epsilon.$$

Une autre façon de procéder est de considérer l'endomorphisme u de $\mathbb{K}^n$ canoniquement associé à la matrice A et une base $\{e_1, \ldots, e_n\}$ dans laquelle la matrice T de u est triangulaire supérieure. Pour tout réel $\delta > 0$, on note $\mathcal{B}_\delta = \{e'_1, \ldots, e'_n\}$ la base de $\mathbb{K}^n$ définie par $e'_j = \delta^{j-1} e_j$ et on a :

$$\forall j \in \{1, \ldots, n\}, \quad u\left(e'_j\right) = \delta^{j-1} \sum_{i=1}^{j} t_{ij} e_i = \sum_{i=1}^{j} \delta^{j-i} t_{ij} e'_i.$$

Pour $\delta > 0$ assez petit la matrice de u dans $\mathcal{B}_\delta$ a alors la forme souhaitée. ■

Théorème 4.14 : *Soit A une matrice d'ordre n à coefficients complexes.*

1. Pour toute norme matricielle induite par une norme vectorielle, on a :

$$\rho(A) \leq \|A\|,$$

l'inégalité pouvant être stricte.

2. Pour tout $\epsilon > 0$, il existe une norme matricielle induite par une norme vectorielle telle que :

$$\|A\| \leq \rho(A) + \epsilon.$$

3. $\rho(A) = \inf_{\|\cdot\| \in \mathcal{N}} \|A\|$, où $\mathcal{N}$ désigne l'ensemble de toutes les normes matricielles induites par une norme vectorielle.

Démonstration :

1. Soit λ une valeur propre de A qui vérifie $\rho(A) = |\lambda|$ et x un vecteur propre associé dans $\mathbb{C}^n$ de norme 1. On a alors :

$$\|Ax\| = \|\lambda x\| = |\lambda| \leq \|A\|,$$

d'où $\rho(A) \leq \|A\|$.

En prenant $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, on a $\rho(A) = 0$ et $\|A\| > 0$.

2. En utilisant le lemme 4.9, pour tout réel $\epsilon > 0$ on peut trouver une matrice inversible P_ϵ telle que $T_\epsilon = P_\epsilon^{-1} A P_\epsilon$ soit triangulaire supérieure avec :

$$T_\epsilon = \left((t_{ij})\right)_{1 \leq i,j \leq n}, \quad \max_{1 \leq i \leq n-1} \sum_{j=i+1}^{n} |t_{ij}| < \epsilon.$$

On associe alors à la matrice P_ϵ la norme matricielle :

$$M \mapsto \|M\|_P = \left\|P_\epsilon^{-1} M P_\epsilon\right\|_\infty$$

(lemme 4.3) et on a :

$$\begin{aligned}
\|A\|_P = \left\|P_\epsilon^{-1} A P_\epsilon\right\|_\infty &= \|T_\epsilon\|_\infty \\
&= \max \left\{ |t_{nn}|, \quad |t_{ii}| + \sum_{j=i+1}^{n} |t_{ij}| \ ; \ 1 \leq i \leq n-1 \right\} \\
&\leq \epsilon + \max_{1 \leq i \leq n} |t_{ii}|,
\end{aligned}$$

soit $\|A\|_P \leq \rho(A) + \epsilon$ puisque les t_{ii} sont les valeurs propres de A.

3. Résulte de ce qui précède. ■

Remarque 4.10 : *De l'équivalence des normes sur $\mathcal{M}_n(\mathbb{C})$ on déduit que pour toute norme sur $\mathcal{M}_n(\mathbb{C})$, il existe une constante $\alpha > 0$ telle que $\rho(A) \leq \alpha \|A\|$ pour tout $A \in \mathcal{M}_n(\mathbb{C})$. Mais on peut avoir $\rho(A) > \|A\|$. Par exemple, pour $n = 2$ avec la norme :*

$$A \longmapsto \|A\| = \max_{1 \leq i,j \leq n} |a_{ij}|$$

et la matrice $A = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}$, *on a :*

$$\|A\| = \max\{|\cos(\theta)|, |\sin(\theta)|\} < 1 = \rho(A),$$

si $\theta \in \left]0, \frac{\pi}{2}\right[$.

Théorème 4.15 : *L'application ρ, qui associe à toute matrice de $\mathcal{M}_n(\mathbb{C})$ son rayon spectral, est continue.*

Démonstration : On munit $\mathcal{M}_n(\mathbb{C})$ d'une norme matricielle induite par une norme vectorielle.

Si $(T_k)_{k\in\mathbb{N}}$ est une suite de matrices triangulaires supérieures qui converge vers une matrice T, alors T est également triangulaire supérieure et il est facile de vérifier que la suite $(\rho(T_k))_{k\in\mathbb{N}}$ converge vers $\rho(T)$.

Soit $(A_k)_{k\in\mathbb{N}}$ une suite de matrices qui converge vers la matrice A dans $\mathcal{M}_n(\mathbb{C})$. On veut montrer que la suite $(\rho(A_k))_{k\in\mathbb{N}}$ converge vers $\rho(A)$ dans $\mathbb{R}$. Pour ce faire on va montrer que cette suite est bornée et admet $\rho(A)$ pour unique valeur d'adhérence.

Avec les inégalités $\rho(A_k) \leq \|A_k\|$ et la convergence de la suite $(A_k)_{k\in\mathbb{N}}$ on déduit que la suite $(\rho(A_k))_{k\in\mathbb{N}}$ est bornée dans $\mathbb{R}$.

Soit $\left(\rho\left(A_{\varphi(k)}\right)\right)_{k\in\mathbb{N}}$ une sous-suite convergente de $(\rho(A_k))_{k\in\mathbb{N}}$. Dans $\mathcal{M}_n(\mathbb{C})$ on sait que toute matrice se trigonalise dans une base orthonormée (théorème de Schur), il existe donc, pour tout entier naturel k, une matrice unitaire U_k telle que la matrice $T_k = U_k^* A_k U_k$ soit triangulaire supérieure. Dans le compact $\mathcal{U}_n(\mathbb{C})$ on peut extraire de $\left(U_{\varphi(k)}\right)_{k\in\mathbb{N}}$ une sous-suite $\left(U_{\sigma(k)}\right)_{k\in\mathbb{N}}$ qui converge vers une matrice unitaire U. La suite $\left(T_{\sigma(k)}\right)_{k\in\mathbb{N}}$ converge alors vers la matrice $T = U^* AU$ qui est triangulaire supérieure. On a alors :

$$\rho(A) = \rho(T) = \lim_{k\to+\infty} \rho\left(T_{\sigma(k)}\right) = \lim_{k\to+\infty} \rho\left(A_{\sigma(k)}\right) = \lim_{k\to+\infty} \rho\left(A_{\varphi(k)}\right).$$

On a donc ainsi montré que la suite bornée $(\rho(A_k))_{k\in\mathbb{N}}$ admet $\rho(A)$ pour unique valeur d'adhérence. Cette suite converge donc vers $\rho(A)$. ■

Lemme 4.10 : *Soient $\theta_1, \ldots, \theta_p$ des réels deux à deux distincts dans $[0, 2\pi[$ et $a_1, \ldots, a_p$ des réels. On définit la suite $(u_k)_{k\in\mathbb{N}}$ par $u_k = \sum_{j=1}^{p} a_j e^{ik\theta_j}$. Si $\lim_{k\to+\infty} u_k = 0$, alors tous les a_j $(1 \leq j \leq p)$ sont nuls.*

Démonstration : Pour $z \in \mathbb{C}$ on a $\left|u_k z^k\right| \leq \left(\sum\limits_{j=1}^{p} |a_j|\right) |z|^k$ et du fait que $\sum\limits_{k=0}^{+\infty} |z|^k < +\infty$ pour $|z| < 1$ on déduit que l'on peut poser $f(z) = \sum\limits_{k=0}^{+\infty} u_k z^k$ pour $|z| < 1$. On a alors :

$$f(z) = \sum_{j=1}^{p} a_j \sum_{k=0}^{+\infty} \left(z e^{i\theta_j}\right)^k = \sum_{j=1}^{p} \frac{a_j}{1 - z e^{i\theta_j}} \qquad (|z| < 1).$$

Ce qui donne, pour tout $j \in \{1, \ldots, p\}$:

$$a_j = \left(1 - z e^{i\theta_j}\right) f(z) - \left(1 - z e^{i\theta_j}\right) \sum_{\substack{r=1 \\ r \neq j}}^{p} \frac{a_r}{1 - z e^{i\theta_r}} \quad (|z| < 1).$$

En prenant $z = \rho e^{-i\theta_j}$, avec $0 < \rho < 1$, on obtient en faisant tendre ρ vers 1 :

$$a_j = \lim_{\rho \to 1} (1 - \rho) f\left(\rho e^{-i\theta_j}\right),$$

puisque les θ_r sont deux à deux distincts dans $[0, 2\pi[$.

Comme $\lim\limits_{k \to +\infty} u_k = 0$, pour tout $\epsilon > 0$ on peut trouver un entier k_0 tel que $|u_k| < \epsilon$ pour tout $k > k_0$. Ce qui donne, pour $|z| < 1$:

$$\begin{aligned} |f(z)| &\leq \sum_{k=0}^{k_0} |u_k| + \epsilon \sum_{k=k_0+1}^{+\infty} |z|^k = \sum_{k=0}^{k_0} |u_k| + \frac{\epsilon |z|^{k_0+1}}{1 - |z|} \\ &\leq \sum_{k=0}^{k_0} |u_k| + \frac{\epsilon}{1 - |z|}. \end{aligned}$$

Et pour $z = \rho e^{-i\theta_j}$ avec $0 < \rho < 1$:

$$(1 - \rho) \left| f\left(\rho e^{-i\theta_j}\right) \right| \leq (1 - \rho) \left(\sum_{k=0}^{k_0} |u_k| \right) + \epsilon < 2\epsilon,$$

pour ρ voisin de 1. On a donc $\lim\limits_{\rho \to 1} (1 - \rho) f\left(\rho e^{-i\theta_j}\right) = 0$ et $a_j = 0$ pour tout $j \in \{1, \ldots, p\}$. ∎

Théorème 4.16 : *Soit A dans $\mathcal{M}_n(\mathbb{C})$, les conditions suivantes sont équivalentes :*

(i) $\lim\limits_{k \to +\infty} A^k = 0$.

(ii) *Pour toute valeur initiale x_0, la suite $(x_k)_{k \in \mathbb{N}}$ définie par $x_{k+1} = A x_k$, pour $k \geq 0$, converge vers le vecteur nul.*

(iii) $\rho(A) < 1$.

(iv) *Il existe au moins une norme matricielle induite telle que $\|A\| < 1$.*

(v) *La matrice $I_n - A$ est inversible et la série de terme général A^k est convergente de somme $(I_n - A)^{-1}$.*

(vi) *La matrice $I_n - A$ est inversible et la série de terme général $\mathrm{Tr}\left(A^k\right)$ est convergente de somme $\mathrm{Tr}\left((I_n - A)^{-1}\right)$.*

(vii) $\lim\limits_{k \to +\infty} \mathrm{Tr}\left(A^k\right) = 0$.

Démonstration :

(i) $\Rightarrow$ (ii) Résulte de :

$$\|x_k\| = \left\|A^k x_0\right\| \leq \left\|A^k\right\| \|x_0\| .$$

(ii) $\Rightarrow$ (iii) Supposons qu'il existe une valeur propre λ de A telle que $|\lambda| \geq 1$. Si x_0 est un vecteur propre non nul associé à λ, en écrivant que $x_k = A^k x_0 = \lambda^k x_0$, on voit que la suite $(x_k)_{k \in \mathbb{N}}$ ne converge pas vers 0.

(iii) $\Rightarrow$ (iv) Soit $\epsilon > 0$ tel que $\rho(A) + \epsilon < 1$. Il suffit de prendre une norme matricielle induite telle que $\|A\| < \rho(A) + \epsilon$ (théorème 4.14).

(iv) $\Rightarrow$ (i) En prenant une norme matricielle induite qui vérifie $\|A\| < 1$ et en écrivant que $\left\|A^k\right\| \leq \|A\|^k$, on déduit que $\lim\limits_{k \to +\infty} \left(A^k\right) = 0$.

On a donc montré que les assertions (i) à (iv) sont équivalentes.

(iii) $\Rightarrow$ (v) Si $\rho(A) < 1$ alors 1 n'est pas valeur propre de A et $I_n - A$ est inversible. En notant, pour tout entier p, $S_p = \sum\limits_{k=0}^{p} A^k$, on a :

$$(I_n - A) S_p = I_n - A^{p+1},$$

avec $\lim\limits_{p \to +\infty} \left(I_n - A^{p+1}\right) = I_n$. En utilisant la continuité du produit matriciel, on déduit alors que :

$$\lim_{p \to +\infty} S_p = \lim_{p \to +\infty} (I_n - A)^{-1} \left(I_n - A^{p+1}\right) = (I_n - A)^{-1},$$

c'est-à-dire que $\sum\limits_{k=0}^{+\infty} A^k = (I_n - A)^{-1}$.

(v) $\Rightarrow$ (vi) La convergence de la série $\sum\limits_{k=0}^{+\infty} A^k$ entraîne $\lim\limits_{k \to +\infty} A^k = 0$ et en conséquence $\rho(A) < 1$. En notant $\{\lambda_1, \ldots, \lambda_n\}$ les valeurs propres de A, on a pour tout entier p :

$$\sum_{k=0}^{p} \mathrm{Tr}\left(A^k\right) = \sum_{k=0}^{p} \sum_{j=1}^{n} \lambda_j^p = \sum_{j=1}^{n} \sum_{k=0}^{p} \lambda_j^p = \sum_{j=1}^{n} \frac{1 - \lambda_j^{p+1}}{1 - \lambda_j},$$

avec $|\lambda_j| < 1$ pour tout j. On déduit alors que la série de terme général $\mathrm{Tr}\left(A^k\right)$ est convergente avec :

$$\sum_{k=0}^{+\infty} \mathrm{Tr}\left(A^k\right) = \sum_{j=1}^{n} \frac{1}{1 - \lambda_j} \cdot$$

En considérant que les $\frac{1}{1-\lambda_j}$, pour $1 \leq j \leq n$, sont toutes les valeurs propres de $(I_n - A)^{-1}$, on déduit que :

$$\sum_{k=0}^{+\infty} \operatorname{Tr}\left(A^k\right) = \operatorname{Tr}\left(\left(I_n - A\right)^{-1}\right).$$

(vi) $\Rightarrow$ (vii) Résulte immédiatement du fait que le terme général d'une série convergente tend vers 0.

(vii) $\Rightarrow$ (iii) Supposons que $\rho(A) \geq 1$. On note $\{\lambda_1, \ldots, \lambda_n\}$ les valeurs propres de A avec :

$$|\lambda_1| = \ldots = |\lambda_p| = \rho(A) > |\lambda_j| \quad (j > p)$$

(dans le cas où $p < n$).
On a alors :

$$\begin{aligned}\sum_{j=1}^{p} \left(\frac{\lambda_j}{\rho(A)}\right)^k &= \frac{1}{(\rho(A))^k} \sum_{j=1}^{n} \lambda_j^k - \sum_{j=p+1}^{n} \left(\frac{\lambda_j}{\rho(A)}\right)^k \\ &= \frac{1}{(\rho(A))^k} \operatorname{Tr}\left(A^k\right) - \sum_{j=p+1}^{n} \left(\frac{\lambda_j}{\rho(A)}\right)^k \underset{k \to +\infty}{\to} 0.\end{aligned}$$

En notant $\left\{e^{i\theta_1}, \ldots, e^{i\theta_q}\right\}$ l'ensemble des valeurs prises par $\left\{\frac{\lambda_1}{\rho(A)}, \ldots, \frac{\lambda_p}{\rho(A)}\right\}$ avec les réels θ_j deux à deux distincts dans $[0, 2\pi[$, on a :

$$\sum_{j=1}^{p} \left(\frac{\lambda_j}{\rho(A)}\right)^k = \sum_{j=1}^{q} a_j e^{ik\theta_j} \underset{k \to +\infty}{\to} 0,$$

les coefficients a_j étant des entiers strictement positifs, ce qui est impossible d'après le lemme 4.10. On a donc $\rho(A) < 1$. ∎

Corollaire 4.6 : *Quelle que soit la norme choisie sur $\mathcal{M}_n(\mathbb{C})$ on a :*

$$\rho(A) = \lim_{k \to +\infty} \left(\left\|A^k\right\|^{\frac{1}{k}}\right).$$

Démonstration : On travaille tout d'abord avec une norme matricielle induite par une norme vectorielle.

Soit $\epsilon > 0$ et $A_\epsilon = \frac{1}{\rho(A) + \epsilon} A$.

On a $\rho(A_\epsilon) = \frac{\rho(A)}{\rho(A) + \epsilon} < 1$ donc $\lim_{k \to +\infty} \left(A_\epsilon^k\right) = 0$ et

$$\exists k_\epsilon \in \mathbb{N} \mid \forall k \geq k_\epsilon, \quad \left\|A_\epsilon^k\right\| < 1.$$

On a alors :

$$\forall k \geq k_\epsilon, \quad \left\|A^k\right\| < \left(\rho(A) + \epsilon\right)^k.$$

Puis, avec $\rho(A) = \left(\rho\left(A^k\right)\right)^{\frac{1}{k}} \leq \left\|A^k\right\|^{\frac{1}{k}}$, on déduit que :

$$\forall k \geq k_\epsilon, \quad \rho(A) \leq \left\|A^k\right\|^{\frac{1}{k}} \leq \rho(A) + \epsilon,$$

c'est-à-dire le résultat.

Pour toute norme sur $\mathcal{M}_n(\mathbb{C})$, il existe deux constantes $\alpha > 0$ et $\beta > 0$ telles que $\alpha \|A\|_1 \leq \|A\| \leq \beta \|A\|_1$ pour toute $A \in \mathcal{M}_n(\mathbb{C})$ (toutes les normes sont équivalentes sur un espace vectoriel de dimension finie). On a alors :

$$\forall k > 0, \quad \alpha^{\frac{1}{k}} \left\|A^k\right\|_1^{\frac{1}{k}} \leq \left\|A^k\right\|^{\frac{1}{k}} \leq \beta^{\frac{1}{k}} \left\|A^k\right\|_1^{\frac{1}{k}},$$

avec $\lim\limits_{k\to+\infty} \alpha^{\frac{1}{k}} = \lim\limits_{k\to+\infty} \beta^{\frac{1}{k}} = 1$ et $\lim\limits_{k\to+\infty} \left(\left\|A^k\right\|_1^{\frac{1}{k}}\right) = \rho(A)$. Donc :

$$\lim_{k\to+\infty} \left(\left\|A^k\right\|^{\frac{1}{k}}\right) = \rho(A). \qquad \blacksquare$$

Ce résultat peut aussi se montrer en utilisant la décomposition $D+N$ de Dunford-Schwarz (exercice 4.22).

5. Le théorème de Perron-Frobenius

Le théorème de Perron-Frobenius nous donne une condition suffisante sur les coefficients d'une matrice réelle pour que sa valeur propre dominante soit unique. Dans ce cas on dispose d'un algorithme de calcul de cette valeur propre dominante : la méthode de la puissance itérée (paragraphe 2, page 237).

La démonstration de ce théorème utilise les lemmes qui suivent.

Lemme 4.11 : *Si $z_1, \ldots, z_n$ sont des nombres complexes tels que :*

$$\left|\sum_{k=1}^{n} z_k\right| = \sum_{i=1}^{n} |z_k|, \tag{4.5}$$

alors il existe un réel $\theta \in \left]-\pi, \pi\right]$ tel que :

$$\forall k \in \{1, \ldots, n\}, \quad z_k = |z_k| e^{i\theta}.$$

Démonstration : Chaque z_k ($1 \leq k \leq n$) peut s'écrire $z_k = \rho_k e^{i\theta_k}$ avec $\rho_k = |z_k| \geq 0$ et $\theta_k \in \left]-\pi, \pi\right]$.

On a alors :

$$\begin{cases} \left|\sum\limits_{k=1}^{n} z_k\right|^2 = \sum\limits_{i=1}^{n} |z_k|^2 + 2 \sum\limits_{1\leq j<k\leq n} \rho_j \rho_k \cos(\theta_j - \theta_k), \\ \left(\sum\limits_{i=1}^{n} |z_k|\right)^2 = \sum\limits_{i=1}^{n} |z_k|^2 + 2 \sum\limits_{1\leq j<k\leq n} \rho_j \rho_k, \end{cases}$$

et la relation (4.5) est équivalente à :

$$\sum_{1\leq j<k\leq n} \rho_j\rho_k\big(1-\cos(\theta_j-\theta_k)\big)=0.$$

Tous les termes de cette somme étant positifs ou nuls, on en déduit que $\rho_j\rho_k=0$ ou $\cos(\theta_j-\theta_k)=1$ pour $1\leq j\neq k\leq n$. Pour $j\neq k$ tels que ρ_j et ρ_k soient non nuls, on a alors $\cos(\theta_j-\theta_k)=1$ avec $-\pi<\theta_j-\theta_k\leq\pi$, ce qui équivaut à $\theta_j=\theta_k$. En notant θ cette valeur commune on peut prendre $\theta_k=\theta$ pour les indices k tels que $\rho_k=0$ et on a alors $z_k=\rho_k e^{i\theta}=|z_k|\,e^{i\theta}$ pour tout entier k compris entre 1 et n. ■

Lemme 4.12 : *Si A est une matrice d'ordre $n\geq 2$ à coefficients réels strictement positifs alors $\rho(A)$ est non nul.*

Démonstration : Si $\rho(A)=0$ alors toutes les valeurs propres de A sont nulles et son polynôme caractéristique est donné par $P_A(X)=(-1)^n X^n$. Avec le théorème de Cayley-Hamilton on déduit alors que $A^n=0$, ce qui est contradictoire avec le fait que tous les coefficients de A sont strictement positifs. On a donc $\rho(A)>0$. ■

Théorème 4.17 (Perron-Frobenius) : *Si A est une matrice d'ordre $n\geq 2$ à coefficients réels strictement positifs alors $\rho(A)$ est l'unique valeur propre de A de module maximal et l'espace propre associé est une droite vectorielle engendrée par un vecteur à composantes réelles strictement positives.*

Démonstration : On montre tout d'abord que si $y\in\mathbb{C}^n-\{0\}$ est un vecteur propre associé à une valeur propre $\lambda\in\mathbb{C}$ telle que $\rho(A)=|\lambda|$, alors toutes les composantes de y sont non nulles et le vecteur $x\in\mathbb{R}^n$ de composantes $x_i=|y_i|$ $(1\leq i\leq n)$ est aussi un vecteur propre associé à λ.

Soit donc $y\in\ker(A-\lambda I_n)$. On a :

$$\sum_{j=1}^n a_{ij}y_j=\lambda y_i\quad(1\leq i\leq n)$$

et

$$|\lambda|\,|y_i|\leq\sum_{j=1}^n a_{ij}\,|y_j|\quad(1\leq i\leq n)$$

(les coefficients a_{ij} sont tous strictement positifs). Supposons que :

$$\forall i\in\{1,\ldots,n\},\quad |\lambda|\,|y_i|<\sum_{j=1}^n a_{ij}\,|y_j|.$$

On peut alors trouver un réel $\epsilon>0$ tel que :

$$\forall i\in\{1,\ldots,n\},\quad (|\lambda|+\epsilon)\,|y_i|<\sum_{j=1}^n a_{ij}\,|y_j|.$$

Si x est le vecteur de composantes $x_i = |y_i|$, on a alors, en notant $(Ax)_i$ la composante numéro i du vecteur Ax :

$$\forall i \in \{1, \ldots, n\}, \quad (|\lambda| + \epsilon) x_i < (Ax)_i .$$

Par récurrence sur $k \geq 1$, on déduit alors que :

$$\forall i \in \{1, \ldots, n\}, \quad (|\lambda| + \epsilon)^k x_i < \left(A^k x\right)_i ,$$

en notant $\left(A^k x\right)_i$ la composante numéro i du vecteur $A^k x$.

En effet, le résultat est vrai pour $k = 1$ et, en le supposant vrai pour $k \geq 1$, on a :

$$\left(A^{k+1} x\right)_i = \sum_{j=1}^{n} a_{ij} \left(A^k x\right)_j > (|\lambda| + \epsilon)^k \sum_{j=1}^{n} a_{ij} x_j,$$

soit :

$$\left(A^{k+1} x\right)_i > (|\lambda| + \epsilon)^k (Ax)_i > (|\lambda| + \epsilon)^{k+1} x_i.$$

On déduit alors que :

$$\forall k \geq 1, \quad (|\lambda| + \epsilon)^k \sum_{i=1}^{n} x_i < \sum_{i=1}^{n} \left(A^k x\right)_i ,$$

ce qui s'écrit aussi :

$$\forall k \geq 1, \quad (|\lambda| + \epsilon)^k \|x\|_1 < \left\|A^k x\right\|_1 \leq \left\|A^k\right\|_1 \|x\|_1 .$$

Le vecteur x étant non nul, on aboutit à :

$$\forall k \geq 1, \quad (|\lambda| + \epsilon)^k < \left\|A^k\right\|_1 ,$$

ou encore :

$$\forall k \geq 1, \quad (|\lambda| + \epsilon) < \left\|A^k\right\|_1^{\frac{1}{k}} ,$$

ce qui contredit $\lim\limits_{k \to +\infty} \left\|A^k\right\|_1^{\frac{1}{k}} = \rho(A) = |\lambda|$.

On a donc montré qu'il existe un indice $i \in \{1, 2, ..., n\}$ tel que :

$$\sum_{j=1}^{n} a_{ij} x_j = |\lambda| x_i,$$

ce qui peut encore s'écrire :

$$|\lambda y_i| = \left| \sum_{j=1}^{n} a_{ij} y_j \right| = \sum_{j=1}^{n} a_{ij} |y_j| .$$

Avec le lemme 4.11, on déduit alors que x est proportionnel à y, c'est-à-dire qu'on peut trouver un vecteur propre x associé à la valeur propre λ de composantes positives.

Si l'une des composantes x_i du vecteur x est nulle, avec $\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i = 0$, $a_{ij} > 0$ et $x_j \geq 0$ pour tout j, on déduit alors que $x = 0$, ce qui est faux. On a donc $x_i > 0$ pour tout $i \in \{1, 2, ..., n\}$.

Ensuite avec $\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i$, les coefficients a_{ij} et x_j étant tous strictement positifs, on déduit que λ est un réel strictement positif.

On montre ensuite que l'espace propre E_λ associé à la valeur propre λ est une droite vectorielle.

Si x, y sont deux éléments non nuls de E_λ alors toutes leurs composantes sont non nulles et il existe $\alpha \in \mathbb{C} - \{0\}$ tel que $x_1 - \alpha y_1 = 0$. D'après ce qui précède, on a nécessairement $x - \alpha y = 0$. En effet, $x - \alpha y$ est un vecteur propre associé à λ et s'il est non nul toutes ses composantes sont non nulles. On en déduit donc que E_λ est de dimension 1.

On montre enfin que $\lambda = \rho(A)$ est valeur propre simple de A.

Si λ est valeur propre de A de multiplicité supérieure ou égale à 2, en tenant compte de $\dim(E_\lambda) = 1$ et en réduisant la matrice A à la forme triangulaire supérieure, on déduit que pour tout $x \in \left(\mathbb{R}_+^*\right)^n \cap E_\lambda$ on peut trouver $y \in \mathbb{C}^n$ non colinéaire à x tel que $Ay = x + \lambda y$. Comme A, λ et x sont réels, on a aussi $A\bar{y} = x + \lambda\bar{y}$ et $A\dfrac{y + \bar{y}}{2} = x + \lambda\dfrac{y + \bar{y}}{2}$ avec $y + \bar{y} \in \mathbb{R}^n$. On peut donc trouver un vecteur $z \in \mathbb{R}^n$ tel que $Az = x + \lambda z$. Pour $\alpha > 0$ assez grand, on aura alors $u = z + \alpha x \in \left(\mathbb{R}_+^*\right)^n$ et $Au = x + \lambda u$. Les composantes des vecteurs considérés étant toutes positives, on a :

$$\|Au\|_1 = \|x\|_1 + \lambda \|u\|_1 .$$

Pour $\epsilon > 0$ tel que $\|x\|_1 > \epsilon \|u\|_1$, on a alors :

$$\|Au\|_1 > (\lambda + \epsilon) \|u\|_1 .$$

Par récurrence, on en déduit que :

$$\forall k \geq 1, \quad (\lambda + \epsilon)^k \|u\|_1 < \left\|A^k u\right\|_1 \leq \left\|A^k\right\|_1 \|u\|_1 .$$

Le vecteur u étant non nul, on déduit que $\left\|A^k\right\|_1^{\frac{1}{k}} > (\lambda + \epsilon)$, ce qui est en contradiction avec $\lim\limits_{k \to +\infty} \left\|A^k\right\|_1^{\frac{1}{k}} = \rho(A) = |\lambda|$.

On a donc ainsi montré que λ est valeur propre simple de A. ■

6. Conditionnement d'une matrice

Quand on étudie un système linéaire de n équations à n inconnues à coefficients réels ou complexes, on peut se poser la question suivante : si $x \in \mathbb{K}^n$ est solution du système $Ax = b$, comment sera modifiée cette solution si les coefficients du second membre ou de la matrice sont modifiés ?

Considérons par exemple le système $Ax = b$, avec :

$$A = \begin{pmatrix} 10 & 7 & 8 & 7 \\ 7 & 5 & 6 & 5 \\ 8 & 6 & 10 & 9 \\ 7 & 5 & 9 & 10 \end{pmatrix}, \quad b = \begin{pmatrix} 32 \\ 23 \\ 33 \\ 31 \end{pmatrix},$$

de solution $x = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$. Si on modifie le second membre en :

$$b + \delta b = \begin{pmatrix} 32{,}1 \\ 22{,}9 \\ 33{,}1 \\ 30{,}9 \end{pmatrix},$$

c'est-à-dire que b est donné avec une erreur relative de 0,3 %, alors la solution devient :

$$x + \delta x = \begin{pmatrix} 9{,}2 \\ -12{,}6 \\ 4{,}5 \\ -1{,}1 \end{pmatrix}.$$

On a donc une erreur relative sur x de l'ordre de 1 000 %.

De même si on perturbe la matrice en prenant :

$$A + \delta A = \begin{pmatrix} 10 & 7 & 8{,}1 & 7{,}2 \\ 7{,}08 & 5{,}04 & 6 & 5 \\ 8 & 5{,}98 & 9{,}81 & 9 \\ 6{,}99 & 4{,}99 & 9 & 9{,}98 \end{pmatrix}$$

et, en gardant le second membre initial, la solution devient :

$$x + \delta x = \begin{pmatrix} -8{,}1 \\ 137 \\ -34 \\ 22 \end{pmatrix}.$$

Le théorème et la définition qui suivent permettent d'étudier plus en détail cette question.

Théorème 4.18 : *Soient $x \mapsto \|x\|$ une norme sur $\mathbb{K}^n$, $A \mapsto \|A\|$ la norme matricielle induite, A une matrice dans $GL_n(\mathbb{K})$ et x dans $\mathbb{K}^n$ solution du système $Ax = b$. Si $x + \delta x$ est la solution du système perturbé $Ay = b + \delta b$, on a alors :*

$$\frac{\|\delta x\|}{\|x\|} \leq \|A\| \left\|A^{-1}\right\| \frac{\|\delta b\|}{\|b\|}. \tag{4.6}$$

Si $x + \delta x$ est la solution du système perturbé $(A + \delta A)\, y = b$, on a alors :

$$\frac{\|\delta x\|}{\|x + \delta x\|} \leq \|A\| \left\|A^{-1}\right\| \frac{\|\delta A\|}{\|A\|}. \tag{4.7}$$

Démonstration : Supposons que $A(x + \delta x) = b + \delta b$ et $Ax = b$. On en déduit alors que $\delta x = A^{-1}\delta b$ et $\|\delta x\| \leq \left\|A^{-1}\right\| \|\delta b\|$. De même $b = Ax$ donne $\|b\| \leq \|A\| \|x\|$. L'inégalité (4.6) s'en déduit alors immédiatement.

Supposons que $(A + \delta A)(x + \delta x) = b$ et $Ax = b$. On en déduit alors que $\delta x = -A^{-1}\delta A\,(x + \delta x)$ et $\|\delta x\| \leq \left\|A^{-1}\right\| \|\delta A\| \|x + \delta x\|$. D'où l'inégalité (4.7). ■

Ce qui nous amène à poser la définition suivante.

Définition 4.5 : *Soit $A \mapsto \|A\|$ une norme matricielle induite par une norme vectorielle $x \mapsto \|x\|$. Si A est une matrice réelle ou complexe inversible, alors le conditionnement de A relativement à cette norme est la quantité :*

$$\operatorname{cond}(A) = \frac{1}{\|A\| \left\|A^{-1}\right\|}.$$

Remarque 4.11 : *Le conditionnement n'est défini que pour une matrice inversible et dépend du choix d'une norme matricielle.*

On notera $\operatorname{cond}_\infty$, cond_1 et cond_2 les conditionnements associés respectivement aux trois normes classiques de $\mathbb{K}^n$.

Le mauvais conditionnement d'une matrice n'est pas lié à un déterminant voisin de 0. Dans le cas de l'exemple précédent, la matrice est symétrique définie positive de déterminant égal à 1.

Les valeurs propres de la matrice A sont 0,01, 0,84, 3,86 et 30,3. Le conditionnement, relativement à la norme euclidienne, est alors :

$$\operatorname{cond}_2(A) \cong \frac{0{,}01}{30{,}3} \cong 3{,}3 \times 10^{-4}.$$

Le théorème suivant résume quelques propriétés du conditionnement.

Théorème 4.19 : *Soit $A \mapsto \|A\|$ une norme matricielle induite par une norme vectorielle $x \mapsto \|x\|$. Pour toute matrice inversible A à coefficients réels ou complexes, on a :*

$$\operatorname{cond}(A) \in \left]0, 1\right].$$
$$\operatorname{cond}(A) = \operatorname{cond}\left(A^{-1}\right).$$
$$\forall \alpha \in \mathbb{K}^*, \quad \operatorname{cond}(\alpha A) = \operatorname{cond}(A).$$

Démonstration : En écrivant que $\left\|AA^{-1}\right\| = 1$, on déduit que $1 \leq \|A\| \left\|A^{-1}\right\|$. C'est-à-dire que le conditionnement est un réel non nul compris entre 0 et 1. Les autres propriétés sont des conséquences immédiates de la définition. ∎

Remarque 4.12 : *Un système sera bien conditionné si* $\operatorname{cond}(A)$ *est voisin de* 1 *et mal conditionné si ce conditionnement est proche de* 0.

Dans le cas de la norme hermitienne, le conditionnement d'une matrice peut se calculer en fonction de ses valeurs singulières (les valeurs propres de A^*A).

Théorème 4.20 : *Soit A une matrice inversible à coefficients réels ou complexes. On a alors :*

$$\operatorname{cond}_2(A) = \sqrt{\frac{\mu_{\min}}{\mu_{\max}}},$$

où $\mu_{\min}$ [resp. $\mu_{\max}$] est la plus petite [resp. plus grande] valeur propre de A^*A.

Pour A normale, on a :

$$\operatorname{cond}_2(A) = \frac{\min\limits_{1\leq i\leq n} |\lambda_i|}{\max\limits_{1\leq i\leq n} |\lambda_i|},$$

où les λ_i sont les valeurs propres de A.

Démonstration : On a $\|A\|_2^2 = \rho(A^*A) = \mu_{\max}$ et :

$$\left\|A^{-1}\right\|_2^2 = \rho\left(\left(A^{-1}\right)^* A^{-1}\right) = \rho\left((AA^*)^{-1}\right) = \frac{1}{\mu_{\min}}.$$

Ce qui donne :

$$\operatorname{cond}_2(A) = \frac{1}{\|A\|_2 \left\|A^{-1}\right\|_2} = \sqrt{\frac{\mu_{\min}}{\mu_{\max}}}.$$

Si A est normale alors A^{-1} est aussi normale et $\|A\|_2 = \rho(A)$,

$$\left\|A^{-1}\right\|_2 = \rho\left(A^{-1}\right) = \frac{1}{\min\limits_{1\leq i\leq n} |\lambda_i|}.$$

Ce qui donne :

$$\operatorname{cond}_2(A) = \frac{1}{\|A\|_2 \left\|A^{-1}\right\|_2} = \frac{\min\limits_{1\leq i\leq n} |\lambda_i|}{\max\limits_{1\leq i\leq n} |\lambda_i|}.$$ ∎

Remarque 4.13 : *Si la matrice A est unitaire complexe ou orthogonale réelle, alors $\|A\|_2 = \left\|A^{-1}\right\|_2 = 1$ et $\operatorname{cond}_2(A) = 1$.*

7. Quotient de Rayleigh-Ritz et hausdorffien

Un outil intéressant pour l'étude des matrices hermitiennes est le quotient de Rayleigh-Ritz. Il nous permettra d'étudier le problème du conditionnement des valeurs propres dans le cas des matrices hermitiennes.

Dans ce paragraphe, $\mathbb{C}^n$ est muni de sa structure hermitienne canonique et on note S_1 la sphère unité de $(\mathbb{C}^n, \|\cdot\|_2)$.

Définition 4.6 : *Soit A une matrice complexe. On appelle quotient de Rayleigh-Ritz associé à cette matrice l'application :*

$$R_A : x \in \mathbb{C}^n - \{0\} \longmapsto R_A(x) = \frac{\langle Ax \mid x\rangle}{\|x\|_2^2}.$$

Définition 4.7 : *Soit A une matrice complexe. On appelle hausdorffien de A la partie de $\mathbb{C}$ définie par :*

$$\mathcal{H}(A) = R_A(\mathbb{C}^n - \{0\}).$$

Par linéarité, on vérifie facilement que le hausdorffien d'une matrice A est aussi défini par :

$$\mathcal{H}(A) = R_A(S_1) = \{\langle Ax \mid x\rangle \mid x \in S_1\}.$$

Le résultat qui suit nous donne une définition variationnelle des valeurs propres d'une matrice hermitienne.

Théorème 4.21 (Rayleigh-Ritz) : *Soit A une matrice hermitienne de valeurs propres (réelles) :*

$$\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$$

et $\{e_1, e_2, \ldots, e_n\}$ une base orthonormée de vecteurs propres associés avec, pour tout entier k compris entre 1 et n, $Ae_k = \lambda_k e_k$. On a :

$$\begin{cases} \lambda_k = \sup\left\{R_A(x) \mid x \in \operatorname{Vect}\{e_{k+1}, \ldots, e_n\}^\perp - \{0\}\right\} & (1 \le k \le n-1), \\ \lambda_n = \sup\left\{R_A(x) \mid x \in \mathbb{C}^n - \{0\}\right\}. \end{cases}$$

$$\begin{cases} \lambda_1 = \inf\left\{R_A(x) \mid x \in \mathbb{C}^n - \{0\}\right\}, \\ \lambda_k = \inf\left\{R_A(x) \mid x \in \operatorname{Vect}\{e_1, \ldots, e_{k-1}\}^\perp - \{0\}\right\} & (2 \le k \le n). \end{cases}$$

Démonstration : Pour tout entier k compris entre 1 et n, on a :

$$R_A(e_k) = \langle Ae_k \mid e_k\rangle = \lambda_k.$$

On note $V_n = \mathbb{C}^n$ et pour k compris entre 1 et $n-1$:

$$V_k = \operatorname{Vect}\{e_{k+1}, \ldots, e_n\}^\perp.$$

Soit k un entier compris entre 1 et n. Tout vecteur x dans V_k s'écrit :

$$x = \sum_{j=1}^{k} x_j e_j,$$

et si ce vecteur est non nul on a :

$$R_A(x) = \frac{\sum\limits_{j=1}^{k} \lambda_j x_j^2}{\sum\limits_{j=1}^{k} x_j^2} \leq \lambda_k \frac{\sum\limits_{j=1}^{k} x_j^2}{\sum\limits_{j=1}^{k} x_j^2} = \lambda_k.$$

Avec $R_A(e_k) = \lambda_k$, on déduit que :

$$\lambda_k = \sup\left\{R_A(x) \mid x \in V_k - \{0\}\right\}.$$

Les identités avec les bornes inférieures se montrent de manière analogue. ■

Corollaire 4.7 : *Si A est une matrice hermitienne de valeurs propres :*

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n,$$

alors :

$$\mathcal{H}(A) = [\lambda_1, \lambda_n].$$

Démonstration : Du théorème de Rayleigh-Ritz, on déduit que $\mathcal{H}(A) \subset [\lambda_1, \lambda_n]$.

En écrivant tout $\lambda \in [\lambda_1, \lambda_n]$ sous la forme $\lambda = (1-t)\lambda_1 + t\lambda_n$ et en posant $x = \sqrt{1-t}e_1 + \sqrt{t}e_n$ (notations du théorème 4.21), on a $\|x\|_2 = 1$ et $R_A(x) = \lambda$. On a donc $[\lambda_1, \lambda_n] \subset \mathcal{H}(A)$.

Une autre façon de voir les choses est de dire que la sphère unité de $\mathbb{C}^n$ étant compacte et connexe et son image par la fonction continue R_A étant compacte et connexe dans $\mathbb{R}$, c'est donc un intervalle fermé borné. Les valeurs propres λ_1 et λ_n étant les bornes inférieure et supérieure de cette image, il en résulte que $\mathcal{H}(A) = [\lambda_1, \lambda_n]$. ■

Ce corollaire s'interprète en disant que si A est une matrice hermitienne, pour tout vecteur non nul x dans $\mathbb{C}^n$ il existe au moins une valeur propre de A inférieure ou égale à $R_A(x)$ et au moins une valeur propre supérieure ou égale à $R_A(x)$.

Dans le cas d'une matrice hermitienne on a vu que le hausdorffien est compact convexe et contient le spectre de A. En fait ce résultat est général.

Lemme 4.13 : *Pour toute matrice complexe A et tout scalaire α, on a :*

$$\mathcal{H}(\alpha A) = \alpha\mathcal{H}(A), \quad \mathcal{H}(A + \alpha I_n) = \mathcal{H}(A) + \{\alpha\}.$$

Démonstration : Il suffit de le vérifier. ■

Théorème 4.22 : *Pour toute matrice complexe A, le hausdorffien $\mathcal{H}(A)$ est une partie compacte et connexe de $\mathbb{C}$ qui contient le spectre de A.*

Démonstration : Le hausdorffien $\mathcal{H}(A)$ est compact comme image de la sphère unité S_1 (compacte dans $\mathbb{C}^n$) par l'application continue R_A.

Pour toute valeur propre λ de A, on peut trouver un vecteur propre associé unitaire x et on a :

$$\lambda = \langle Ax \mid x \rangle = R_A(x) \in \mathcal{H}(A).$$

On a donc $\mathrm{sp}(A) \subset \mathcal{H}(A)$.

Montrer que $\mathcal{H}(A)$ est convexe revient à montrer que pour tous x, y dans S_1 le segment $[R_A(x), R_A(y)]$ est contenu dans $\mathcal{H}(A)$.

Si $R_A(x) = R_A(y)$ ce segment est réduit au point $R_A(x)$ qui est bien dans $\mathcal{H}(A)$.

On suppose donc que $R_A(x) \neq R_A(y)$. Dans ces conditions le système $\{x, y\}$ est libre (en effet $y = \lambda x$ entraîne $|\lambda| = 1$ et $R_A(y) = |\lambda|^2 R_A(x) = R_A(x)$) et pour tout réel t compris entre 0 et 1, le vecteur $z(t) = tx + (1-t)y$ est non nul. On peut alors poser :

$$\varphi(t) = R_A(z(t))$$

et on définit ainsi une fonction continue de $[0,1]$ dans $\mathbb{C}$ à valeurs dans $\mathcal{H}(A)$.

On suppose dans un premier temps que $R_A(x) = 1$ et $R_A(y) = 0$.

La matrice A peut s'écrire $A = H_1 + iH_2$ avec H_1 et H_2 hermitiennes (prendre $H_1 = \frac{1}{2}(A - A^*)$ et $H_2 = \frac{1}{2i}(A - A^*)$) et pour tout z dans $\mathbb{C}^n$ on a :

$$\langle Az \mid z \rangle = \langle H_1 z \mid z \rangle + i \langle H_2 z \mid z \rangle$$

avec $\langle H_1 z \mid z \rangle$ et $\langle H_2 z \mid z \rangle$ réels.

Du fait que $R_A(x)$ et $R_A(y)$ sont réels on déduit alors que $\langle H_2 x \mid x \rangle$ et $\langle H_2 y \mid y \rangle$ sont nuls et pour tout réel t compris entre 0 et 1 on a :

$$\langle H_2 z(t) \mid z(t) \rangle = 2\Re \langle H_2 x \mid y \rangle t(1-t).$$

Quitte à remplacer x par λx avec $|\lambda| = 1$ (ce qui ne change pas $R_A(x)$), on peut supposer que $\langle H_2 x \mid y \rangle$ est imaginaire pur de sorte que :

$$\forall t \in \mathbb{R}, \quad \langle H_2 z(t) \mid z(t) \rangle = 0$$

et

$$\forall t \in \mathbb{R}, \quad \langle Az(t) \mid z(t) \rangle = \langle H_1 z(t) \mid z(t) \rangle \in \mathbb{R}.$$

La fonction φ est donc à valeurs réelles, continue sur $[0,1]$ avec $\varphi(0) = 0$ et $\varphi(1) = 1$. Le théorème des valeurs intermédiaires nous dit alors que φ prend toutes les valeurs entre 0 et 1, c'est-à-dire que :

$$[R_A(x), R_A(y)] = [0,1] \subset \varphi([0,1]) \subset \mathcal{H}(A).$$

Dans le cas général, avec $R_A(x) \neq R_A(y)$, on peut trouver des scalaires α, β tels que :

$$\begin{cases} \alpha R_A(x) + \beta = 1, \\ \alpha R_A(y) + \beta = 0. \end{cases}$$

Ce qui peut aussi s'écrire, en tenant compte de $\|x\|_2 = \|y\|_2 = 1$:

$$\begin{cases} \langle (\alpha A + \beta I_n) x \mid x \rangle = 1, \\ \langle (\alpha A + \beta I_n) y \mid y \rangle = 0. \end{cases}$$

En posant $B = \alpha A + \beta I_n$, on a alors $R_B(x) = 1$ et $R_B(y) = 0$ et ce qui précède nous dit que :

$$[0,1] \subset \mathcal{H}(B) = \alpha \mathcal{H}(A) + \{\beta\},$$

c'est-à-dire que tout réel t compris entre 0 et 1 s'écrit $t = \alpha\mu + \beta$ avec $\mu \in \mathcal{H}(A)$.

En écrivant que :

$$\alpha = \frac{1}{R_A(x) - R_A(y)}, \quad \beta = -\frac{R_A(y)}{R_A(x) - R_A(y)},$$

on obtient pour tout $t \in [0,1]$:

$$t = \frac{\mu - R_A(y)}{R_A(x) - R_A(y)}, \quad 1 - t = \frac{R_A(x) - \mu}{R_A(x) - R_A(y)},$$

et

$$tR_A(x) + (1-t) R_A(y) = \mu \in \mathcal{H}(A).$$

On a donc bien $[R_A(x), R_A(y)] \subset \mathcal{H}(A)$. Ce qui achève de prouver que $\mathcal{H}(A)$ est convexe. ∎

Corollaire 4.8 : *Soit A une matrice complexe de trace nulle. Il existe une matrice unitaire U telle U^*AU ait tous ses termes diagonaux nuls.*

Démonstration : On procède par récurrence sur $n \geq 1$. Pour $n = 1$ le résultat est évident.

On le suppose acquis pour $n - 1 \geq 1$ et on se donne une matrice $A = ((a_{ij}))_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{C})$ de trace nulle. On note u l'endomorphisme de $\mathbb{C}^n$ défini par la matrice A dans la base canonique.

Pour tout entier i compris entre 1 et n on a :

$$a_{ii} = \langle Ae_i \mid e_i \rangle \in \mathcal{H}(A),$$

c'est-à-dire que $\mathcal{H}(A)$ contient tous les termes diagonaux de A. Avec la convexité de $\mathcal{H}(A)$ on déduit alors que :

$$0 = \frac{1}{n} \operatorname{Tr}(A) = \frac{1}{n} \sum_{i=1}^{n} a_{ii} \in \mathcal{H}(A).$$

Il existe donc un vecteur unitaire f_1 tel que $\langle Af_1 \mid f_1 \rangle = 0$. On complète ce vecteur en une base orthonormée de $\mathbb{C}^n$, $\{f_1, \ldots, f_n\}$, et dans cette base la matrice de u est de la forme :

$$A_1 = U_1^* A U_1 = \begin{pmatrix} 0 & a \\ b & C \end{pmatrix},$$

où U_1 est une matrice unitaire, a une matrice à une ligne et $n-1$ colonnes, b une matrice à $n-1$ lignes et 1 colonne et C une matrice carrée d'ordre $n-1$. Comme :

$$0 = \mathrm{Tr}\,(A) = \mathrm{Tr}\,(A_1) = \mathrm{Tr}\,(C)\,,$$

on peut appliquer l'hypothèse de récurrence à la matrice C et il existe une matrice unitaire U_2 d'ordre $n-1$ telle que la matrice $U_2^* C U_2$ soit de trace nulle. En posant $U_3 = \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}$ on définit une matrice unitaire d'ordre n et la matrice :

$$U_3^* U_1^* A U_1 U_3 = \begin{pmatrix} 0 & a' \\ b' & U_2^* C U_2 \end{pmatrix}$$

est unitairement semblable à la matrice A et de trace nulle. ∎

8. Conditionnement du problème de valeurs propres

Dans ce paragraphe on va voir que les problèmes de valeurs propres sont bien conditionnés pour les matrices hermitiennes.

Le résultat qui suit nous donne une autre caractérisation variationnelle des valeurs propres d'une matrice hermitienne. Contrairement à la caractérisation de Rayleigh-Ritz elle ne fait pas intervenir les vecteurs propres.

Pour tout k compris entre 1 et n, on désigne par E_k l'ensemble des sous-espaces vectoriels de dimension k de $\mathbb{C}^n$.

Théorème 4.23 (Courant-Fischer) : *Soit A une matrice hermitienne de valeurs propres :*

$$\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n.$$

Pour tout entier k compris entre 1 et n on a :

$$\lambda_k = \inf \{\mu_A\,(V) \mid V \in E_k\}\,,$$

avec :

$$\mu_A\,(V) = \sup \{R_A(x) \mid x \in V - \{0\}\}\,.$$

Démonstration : On désigne par $\{e_1, e_2, \ldots, e_n\}$ une base orthonormée de vecteurs propres de la matrice A, avec, pour tout entier k compris entre 1 et n, $Ae_k = \lambda_k e_k$.

On note $V_n = \mathbb{C}^n$ et pour k compris entre 1 et $n-1$:

$$V_k = \mathrm{Vect}\,\{e_{k+1}, \ldots, e_n\}^{\perp}\,.$$

On a vu que $\lambda_k = \sup \{R_A(x) \mid x \in V_k - \{0\}\} = \mu_A\,(V_k)$. On déduit donc que :

$$\lambda_k \geq \alpha_k = \inf \{\mu\,(V) \mid V \in E_k\}\,.$$

Soit $V \in E_k$, on a :

$$\begin{aligned}\dim\left(V \cap V_{k-1}^{\perp}\right) &= \dim\,(V) + \dim\left(V_{k-1}^{\perp}\right) - \dim\left(V + V_{k-1}^{\perp}\right)\\ &= n + 1 - \dim\left(V + V_{k-1}^{\perp}\right) \geq 1.\end{aligned}$$

Donc $V \cap V_{k-1}^{\perp} \neq \{0\}$ et pour tout $y \in V \cap V_{k-1}^{\perp} - \{0\}$ on a :

$$\lambda_k = \inf\left\{R_A(x) \mid x \in V_{k-1}^{\perp} - \{0\}\right\} \leq R_A(y) \leq \sup\left\{R_A(x) \mid x \in V - \{0\}\right\},$$

c'est-à-dire que :

$$\forall V \in E_k, \quad \lambda_k \leq \sup\left\{R_A(x) \mid x \in V - \{0\}\right\} = \mu_A(V),$$

soit $\lambda_k \leq \alpha_k$ et $\lambda_k = \alpha_k$. ■

On note $\mathcal{H}_n(\mathbb{C})$ l'ensemble des matrices hermitiennes.

Corollaire 4.9 : *Soit $A : [a,b] \longmapsto \mathcal{H}_n(\mathbb{C})$ une application continue. Si pour tout t dans $[a,b]$ on note :*

$$\lambda_1(t) \leq \lambda_2(t) \leq \ldots \leq \lambda_n(t)$$

les valeurs propres de $A(t)$ rangées dans l'ordre croissant, alors les fonctions λ_k sont continues de $[a,b]$ dans $\mathbb{R}$.

Démonstration : Pour $t \in [a,b]$, soit $\{e_1(t), e_2(t), \ldots, e_n(t)\}$ une base orthonormale de vecteurs propres de $A(t)$, avec $A(t)\, e_k(t) = \lambda_k(t)\, e_k(t)$. On note $V_k(t)$ l'espace vectoriel engendré par $\{e_1(t), \ldots, e_k(t)\}$ (l'orthogonal de $\{e_{k+1}(t), \ldots, e_n(t)\}$). En utilisant le théorème de Courant-Fischer, on peut écrire que pour tous t, t_0 dans $[a,b]$ on a :

$$\begin{aligned}\lambda_k(t) &= \inf\left\{\mu_{A(t)}(V) \mid V \in E_k\right\} \\ &\leq \mu_{A(t)}\big(V_k(t_0)\big) = \sup\left\{R_{A(t)}(x) \mid x \in V_k(t_0) - \{0\}\right\}.\end{aligned}$$

Avec $R_{A(t)}(x) = R_{A(t_0)}(x) + R_{A(t)-A(t_0)}(x)$, on a alors :

$$\begin{aligned}\lambda_k(t) &\leq \sup\left\{R_{A(t_0)}(x) + R_{A(t)-A(t_0)}(x) \mid x \in V_k(t_0) - \{0\}\right\} \\ &\leq \sup\left\{R_{A(t_0)}(x) \mid x \in V_k(t_0) - \{0\}\right\} + \alpha_k(t_0),\end{aligned}$$

où on a noté :

$$\alpha_k(t_0) = \sup\left\{R_{A(t)-A(t_0)}(x) \mid x \in V_k(t_0) - \{0\}\right\}.$$

Mais on a aussi :

$$\sup\left\{R_{A(t_0)}(x) \mid x \in V_k(t_0) - \{0\}\right\} = \lambda_k(t_0),$$

donc :

$$\lambda_k(t) \leq \lambda_k(t_0) + \alpha_k(t_0),$$

avec :

$$\begin{aligned}\alpha_k(t_0) &\leq \sup\left\{R_{A(t)-A(t_0)}(x) \mid x \in \mathbb{C}^n - \{0\}\right\} \\ &= \sup\left\{\frac{\langle (A(t) - A(t_0))x | x\rangle}{\|x\|_2^2} \mid x \in \mathbb{C}^n - \{0\}\right\} \\ &\leq \|A(t) - A(t_0)\|_2.\end{aligned}$$

On a donc montré que pour tous t, t_0 dans $[a, b]$ on a :

$$\lambda_k(t) \leq \lambda_k(t_0) + \|A(t) - A(t_0)\|_2 .$$

En permutant les rôles de t et t_0, on déduit que :

$$|\lambda_k(t) - \lambda_k(t_0)| \leq \|A(t) - A(t_0)\|_2 ,$$

ce qui suffit à prouver la continuité de λ_k. ■

Remarque 4.14 *: Ce résultat peut s'interpréter en disant que de petites perturbations sur les coefficients d'une matrice hermitienne n'engendreront que de petites perturbations sur les valeurs propres. Ce qui revient à dire que le problème de valeurs propres est bien conditionné dans ce cas.*

Remarque 4.15 *: Si les valeurs propres d'une fonction continue A définie sur $[a, b]$ et à valeurs dans $\mathcal{H}_n(\mathbb{C})$ définissent des fonctions continues, en général il n'en est pas de même des vecteurs propres (exercice 4.24).*

Pour toute matrice A dans $\mathcal{H}_n(\mathbb{C})$, on note :

$$\lambda_1(A) \leq \lambda_2(A) \leq \ldots \leq \lambda_n(A)$$

ses valeurs propres rangées dans l'ordre croissant.

Corollaire 4.10 (Weyl) : *Soient A, B deux matrices hermitiennes. Pour tout entier k compris entre 1 et n on a :*

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B).$$

En particulier pour B positive, on a :

$$\lambda_k(A) \leq \lambda_k(A+B).$$

Démonstration : On a, pour tout entier k compris entre 1 et n :

$$\lambda_k(A+B) = \inf\{\mu_{A+B}(V) \mid V \in E_k\},$$

avec, pour tout $V \in E_k$:

$$\mu_{A+B}(V) = \sup\{R_{A+B}(x) \mid x \in V - \{0\}\}.$$

En écrivant que :

$$R_{A+B}(x) = R_A(x) + R_B(x) \geq R_A(x) + \lambda_1(B),$$

on déduit que :

$$\mu_{A+B}(V) \geq \mu_A(V) + \lambda_1(B) \geq \lambda_k(A) + \lambda_1(B)$$

et

$$\lambda_k(A) + \lambda_1(B) \leq \lambda_k(A+B).$$

En remarquant que les valeurs propres de la matrice $-B$ sont les $\lambda_k(-B) = -\lambda_{n-k+1}(B)$ et en appliquant l'inégalité précédente au couple $(A+B, -B)$, on obtient :

$$\lambda_k(A+B) + \lambda_1(-B) \leq \lambda_k(A+B-B),$$

soit :

$$\lambda_k(A+B) \leq \lambda_k(A) + \lambda_n(B).$$

Pour B positive, on a $\lambda_1(B) \geq 0$ et

$$\lambda_k(A) \leq \lambda_k(A+B) - \lambda_1(B) \leq \lambda_k(A+B).$$ ■

Corollaire 4.11 : *Pour tout entier k compris entre 1 et n, l'application :*

$$A \mapsto \lambda_k(A)$$

est continue de $\mathcal{H}_n(\mathbb{C})$ dans $\mathbb{R}$.

Démonstration : Soient A, B dans $\mathcal{H}_n(\mathbb{C})$. On a :

$$-\rho(B) \leq \lambda_1(B) \leq \lambda_k(A+B) - \lambda_k(A) \leq \lambda_n(B) \leq \rho(B),$$

c'est-à-dire :

$$|\lambda_k(A+B) - \lambda_k(A)| \leq \rho(B) = \|B\|_2$$

et la continuité de λ_k en découle. ■

En écrivant l'ensemble $\mathcal{H}_n^{++}(\mathbb{C})$ des matrices hermitiennes définies positives sous la forme :

$$\mathcal{H}_n^{++}(\mathbb{C}) = \bigcap_{k=1}^{n} \lambda_k^{-1}(\mathbb{R}^*)$$

et en utilisant la continuité des applications λ_k, on déduit que $\mathcal{H}_n^{++}(\mathbb{C})$ est ouvert dans $\mathcal{H}_n(\mathbb{C})$.

9. Exercices

Exercice 4.1 : *Montrer que pour toute norme $A \longmapsto \|A\|$ sur $\mathcal{M}_n(\mathbb{K})$ il existe une constante réelle $\lambda > 0$ telle que :*

$$\forall (A, B) \in \mathcal{M}_n(\mathbb{K}) \times \mathcal{M}_n(\mathbb{K}), \quad \|AB\| \leq \lambda \|A\| \|B\|.$$

Solution : L'application bilinéaire $(A, B) \longmapsto AB$ étant continue de $(\mathcal{M}_n(\mathbb{K}))^2$ dans $\mathcal{M}_n(\mathbb{K})$ (en dimension finie toute application bilinéaire est continue), il existe donc une constante réelle $\lambda > 0$ telle que :

$$\forall (A, B) \in M_n(\mathbb{K}) \times M_n(\mathbb{K}), \ \|AB\| \leq \lambda \|A\| \|B\|$$

(exercice 1.19).

On peut aussi travailler tout d'abord avec une norme sous-multiplicative (par exemple $\|\cdot\|_1$) puis utiliser le théorème d'équivalence des normes sur $\mathcal{M}_n(\mathbb{K})$.

Exercice 4.2 : *Montrer que la norme de* $\mathcal{M}_n(\mathbb{K})$ $A \mapsto \max_{1\leq i,j\leq n} |a_{ij}|$ *n'est pas sous-multiplicative.*

Solution : Par exemple, dans le cas $n = 2$, avec :

$$A = \begin{pmatrix} a & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b & 0 \\ 1 & 0 \end{pmatrix},$$

où $a > 1$, $b > 1$, on a $AB = \begin{pmatrix} ab+1 & 0 \\ 0 & 0 \end{pmatrix}$ et :

$$\|AB\| = ab + 1 > \|A\| \, \|B\| = ab.$$

Exercice 4.3 : *Montrer la densité de* $GL_n(\mathbb{K})$ *dans* $\mathcal{M}_n(\mathbb{K})$ *en utilisant le fait que toute matrice de rang* r *est équivalente à* $A_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$.

Solution : Soit $A \in M_n(\mathbb{K})$ de rang r. Si $r = 0$, alors $A = 0 = \lim_{k\to+\infty} \frac{1}{k} I_n$. Si $r > 0$, alors il existe deux matrices inversibles P et Q telles que $A = PA_rQ$, où $A_r = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$. On a alors $A = \lim_{k\to+\infty} PM_kQ$, avec $M_k = \begin{pmatrix} I_r & 0 \\ 0 & \frac{1}{k} I_{n-k} \end{pmatrix}$. Dans tous les cas on peut écrire A comme limite d'une suite de matrices inversibles. Donc $GL_n(\mathbb{K})$ est dense dans $\mathcal{M}_n(\mathbb{K})$.

Exercice 4.4 : *Déduire de la densité de* $GL_n(\mathbb{K})$ *dans* $\mathcal{M}_n(\mathbb{K})$ *que si* A *et* B *sont deux matrices dans* $\mathcal{M}_n(\mathbb{K})$ *alors* AB *et* BA *ont même polynôme caractéristique.*

Solution : Si A est inversible, alors AB et $BA = A^{-1}(AB)A$ sont semblables. Elles ont donc même polynôme caractéristique. Dans le cas général, on peut écrire que $A = \lim_{k\to+\infty} M_k$ où $(M_k)_{k\in\mathbb{N}}$ est une suite de matrices inversibles. Les matrices M_kB et BM_k ont donc même polynôme caractéristique et, avec la continuité du produit matriciel et du déterminant, on peut alors écrire que, pour tout λ dans $\mathbb{C}$, on a :

$$\begin{aligned} \det(AB - \lambda I_n) &= \lim_{k\to+\infty} \det(M_kB - \lambda I_n) \\ &= \lim_{k\to+\infty} \det(BM_k - \lambda I_n) = \det(BA - \lambda I_n), \end{aligned}$$

c'est-à-dire que AB et BA ont même polynôme caractéristique.

Exercice 4.5 : *Soient* $A \in \mathcal{M}_n(\mathbb{C})$ *et* $A = UH$ *une décomposition polaire de la matrice* A *avec* U *unitaire et* H *hermitienne positive. Montrer que la matrice* A *est normale si et seulement si* U *et* H *commutent.*

Solution : Dans tous les cas, on a :

$$A^*A = HU^*UH = H^2.$$

Si les matrices U et H commutent, on a alors :

$$AA^* = UHHU^* = H^2 = A^*A,$$

c'est-à-dire que la matrice A est normale.

Réciproquement, on suppose que la matrice A est normale.

La matrice $K = UHU^*$ est hermitienne positive ($\langle Kx \mid x\rangle = \langle H(U^*x) \mid U^*x\rangle \geq 0$ puisque H est positive) avec :

$$K^2 = UHHU^* = AA^* = A^*A = H^2.$$

On a donc $H^2 = K^2$ avec H et K hermitiennes positives, ce qui entraîne $H = K$ du fait de l'unicité de la racine carrée hermitienne positive (corollaire 3.14). On a donc $UHU^* = H$ et $UH = HU$.

Exercice 4.6 : *Déduire le théorème de Cayley-Hamilton de la densité de $\mathcal{D}_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$.*

Démonstration : Pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, on note P_A son polynôme caractéristique.

Si $A \in \mathcal{M}_n(\mathbb{C})$ est diagonalisable, il existe une matrice inversible P et une matrice diagonale :

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}$$

telles que $A = PDP^{-1}$. Ce qui entraîne :

$$P_A(X) = \prod_{k=1}^{n} (\lambda_k - X), \quad P_A(A) = PP_A(D)P^{-1} = 0.$$

Une matrice quelconque $A \in \mathcal{M}_n(\mathbb{C})$ peut s'écrire $A = \lim_{k\to+\infty} A_k$, où $(A_k)_{k\in\mathbb{N}}$ est une suite de matrices diagonalisables. Avec la continuité de l'application $M \longmapsto P_M(M)$ de $\mathcal{M}_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$ (les composantes de cette application sont des fonctions polynomiales des m_{ij}), on déduit alors que :

$$P_A(A) = \lim_{k\to+\infty} P_{A_k}(A_k) = 0. \qquad \blacksquare$$

Exercice 4.7 : *Soit φ une forme linéaire sur $\mathcal{M}_n(\mathbb{K})$ telle que $\varphi(AB) = \varphi(BA)$ pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$.*

1. En notant $\{E_{ij} \mid 1 \leq i, j \leq n\}$ la base canonique de $\mathcal{M}_n(\mathbb{K})$, montrer que $\varphi(E_{ii}) = \varphi(E_{jj})$ pour tous i, j compris entre 1 et n. On note λ cette valeur commune.

2. Montrer que $\varphi(A) = \lambda \operatorname{Tr}(A)$ pour toute matrice A dans $\mathcal{M}_n(\mathbb{K})$ (on peut d'abord supposer que la matrice A est diagonalisable).

3. Soit u un endomorphisme de $\mathcal{M}_n(\mathbb{K})$ tel que $u(I_n) = I_n$ et $u(AB) = u(BA)$ pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$. Montrer que u conserve la trace.

Solution : On rappelle que si $\{e_i \mid 1 \leq i \leq n\}$ désigne la base canonique de $\mathbb{K}^n$, alors la matrice E_{ij} est définie par :

$$\forall k \in \{1, \ldots, n\}, \quad E_{ij} e_k = \begin{cases} 0 & \text{si } k \neq j, \\ e_i & \text{si } k = j. \end{cases}$$

1. Pour $i \neq j$ dans $\{1, \ldots, n\}$ on a $E_{ij}E_{ji} = E_{ii}$ et $E_{ji}E_{ij} = E_{jj}$. On déduit alors que :

$$\varphi(E_{ii}) = \varphi(E_{ij}E_{ji}) = \varphi(E_{ji}E_{ij}) = \varphi(E_{jj}).$$

On peut donc poser $\lambda = \varphi(E_{ii})$ pour tout i compris entre 1 et n.

2. Si D est une matrice diagonale, elle s'écrit :

$$D = \sum_{i=1}^{n} \lambda_i E_{ii},$$

et

$$\varphi(D) = \sum_{i=1}^{n} \lambda_i \varphi(E_{ii}) = \lambda \sum_{i=1}^{n} \lambda_i = \lambda \operatorname{Tr}(D).$$

Si A est une matrice diagonalisable, elle s'écrit $A = PDP^{-1}$ avec P inversible, D diagonale et

$$\varphi(A) = \varphi\left(PDP^{-1}\right) = \varphi\left(DP^{-1}P\right) = \varphi(D) = \lambda \operatorname{Tr}(D) = \lambda \operatorname{Tr}(A).$$

Enfin, si A est quelconque dans $\mathcal{M}_n(\mathbb{K})$, elle peut s'écrire comme limite d'une suite $(A_k)_{k\in\mathbb{N}}$ de matrices diagonalisables et, avec la continuité des formes linéaires φ et Tr, on déduit que :

$$\varphi(A) = \lim_{k\to+\infty} \varphi(A_k) = \lambda \lim_{k\to+\infty} \operatorname{Tr}(A_k) = \lambda \operatorname{Tr}(A).$$

Remarque 4.16 : *Ce résultat est en fait valable pour tout corps commutatif de caractéristique nulle.*

3. On définit la forme linéaire φ par $\varphi(A) = \mathrm{Tr}(u(A))$ pour toute matrice A dans $\mathcal{M}_n(\mathbb{K})$. On a alors $\varphi(AB) = \varphi(BA)$ pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{K})$ et $\varphi(A) = \lambda \mathrm{Tr}(A)$, c'est-à-dire que $\mathrm{Tr}(u(A)) = \lambda \mathrm{Tr}(A)$ pour toute matrice A dans $\mathcal{M}_n(\mathbb{K})$. Enfin, avec $u(I_n) = I_n$ on déduit que $\lambda = 1$.

Exercice 4.8 : *Montrer que $GL_n(\mathbb{C})$ est connexe par arcs en utilisant le fait que toute matrice complexe est semblable à une matrice triangulaire.*

Solution : Pour toute matrice $A \in GL_n(\mathbb{C})$, il existe une matrice $P \in GL_n(\mathbb{C})$ et une matrice triangulaire supérieure $T = ((m_{ij}))_{1 \leq i,j \leq n}$ telles que $A = PTP^{-1}$. On note, pour tout $j \in \{1, 2, \ldots, \}$, $m_{jj} = \rho_j e^{i\theta_j}$ avec $\rho_j > 0$ et on définit un chemin continu $\varphi : [0,1] \to GL_n(\mathbb{C})$ par :

$$\forall t \in [0,1], \quad \varphi(t) = \begin{pmatrix} \varphi_{11}(t) & \varphi_{12}(t) & \ldots & \varphi_{1n}(t) \\ 0 & \varphi_{22}(t) & \ddots & \vdots \\ \vdots & \ddots & \ddots & \varphi_{n-1,n-1}(t) \\ 0 & \ldots & 0 & \varphi_{nn}(t) \end{pmatrix},$$

où

$$\varphi_{ij}(t) = \begin{cases} t\, m_{ij} & \text{si } 1 \leq i < j \leq n, \\ (1-t)\, e^{it\,\theta_j} + t\, m_{jj} & \text{si } i = j. \end{cases}$$

On a alors $\varphi(0) = I_n$, $\varphi(1) = T$ et $\gamma : t \longmapsto P\varphi(t) P^{-1}$ est un chemin continu qui relie la matrice identité à la matrice A dans $GL_n(\mathbb{C})$.

Exercice 4.9 : *Montrer que $GL_n(\mathbb{R})$ n'est pas connexe.*

Solution : Si $GL_n(\mathbb{R})$ est connexe, alors son image par l'application continue déterminant est un connexe de $\mathbb{R}$ (théorème des valeurs intermédiaires). Or cette image est $\mathbb{R}^*$ non connexe dans $\mathbb{R}$. Donc $GL_n(\mathbb{R})$ n'est pas connexe.

Exercice 4.10 : *Montrer que pour tout $r \in \{0, 1, \ldots, n\}$ l'ensemble $\mathcal{A}_r$ des matrices de $\mathcal{M}_n(\mathbb{C})$ de rang r est connexe par arcs.*

Solution : Deux matrices sont de même rang si et seulement si elles sont équivalentes. Donc pour toutes matrices A, B dans $\mathcal{A}_r$ il existe P et Q dans $GL_n(\mathbb{C})$ telles que $B = PAQ$. Si γ_1 et γ_2 sont deux fonctions continues de $[0,1]$ dans $GL_n(\mathbb{C})$ telles que $\gamma_1(0) = I_n$, $\gamma_2(0) = I_n$, $\gamma_1(1) = P$ et $\gamma_2(1) = Q$, alors $\gamma : t \longmapsto \gamma_1(t) A \gamma_2(t)$ est un chemin continu qui relie A et B dans $\mathcal{A}_r$. Ce qui prouve que $\mathcal{A}_r$ est connexe par arcs.

Exercice 4.11 : *Montrer que l'ensemble $\mathcal{D}_2(\mathbb{R})$ des matrices diagonalisables de $\mathcal{M}_2(\mathbb{R})$ n'est pas dense dans $\mathcal{M}_2(\mathbb{R})$.*

Solution : L'application $\varphi : \mathcal{M}_2(\mathbb{R}) \to \mathbb{R}$ qui associe à une matrice $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ le discriminant de son polynôme caractéristique :

$$\varphi(M) = (a-d)^2 + 4bc$$

(résultant de P_M et P'_M) est continue, donc :

$$A = \lim_{k\to+\infty} A_k \Longrightarrow \varphi(A) = \lim_{k\to+\infty} \varphi(A_k).$$

Mais pour A_k dans $\mathcal{D}'_2(\mathbb{R})$ ou dans $\mathcal{D}_2(\mathbb{R})$, on a $\varphi(A_k) \geq 0$ et pour A à valeurs propres complexes non réelles, on a $\varphi(A) < 0$. Une telle matrice A ne peut donc être limite d'une suite de matrices de $\mathcal{D}'_2(\mathbb{R})$ ou $\mathcal{D}_2(\mathbb{R})$.

Exercice 4.12 : *On désigne par $\mathcal{T}_n(\mathbb{R})$ l'ensemble des matrices trigonalisables de $\mathcal{M}_n(\mathbb{R})$ et par $\theta : \mathcal{T}_n(\mathbb{R}) \to \mathbb{R}^n$ l'application définie par $\theta(M) = (\lambda_1, \lambda_2, \dots, \lambda_n)$ où $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$ sont les valeurs propres de la matrice M.*

1. Montrer que $\mathcal{T}_n(\mathbb{R})$ est fermé dans $\mathcal{M}_n(\mathbb{R})$ (on peut montrer que si $(T_k)_{k\in\mathbb{N}}$ est une suite de $\mathcal{T}_n(\mathbb{R})$ qui converge vers $T \in \mathcal{M}_n(\mathbb{R})$, alors la suite $(\theta(T_k))_{k\in\mathbb{N}}$ est bornée dans $\mathbb{R}^n$ et en déduire que le polynôme caractéristique de T est scindé sur $\mathbb{R}$).

2. Montrer que $\mathcal{T}_n(\mathbb{R})$ est l'adhérence de l'ensemble $D_n(\mathbb{R})$ des matrices diagonalisables de $\mathcal{M}_n(\mathbb{R})$.

Solution : On munit $\mathbb{R}^n$ de la norme $\|\cdot\|_\infty$ et $\mathcal{M}_n(\mathbb{R})$ de la norme matricielle induite. On a alors :

$$\forall M \in T_n(\mathbb{R}), \quad \|\theta(M)\|_\infty = \rho(M) = \max_{1\leq k\leq n} |\lambda_k| \leq \|M\|_\infty .$$

1. Si $(T_k)_{k\in\mathbb{N}}$ est une suite dans $\mathcal{T}_n(\mathbb{R})$ qui converge vers $T \in \mathcal{M}_n(\mathbb{R})$, la suite $(\theta(T_k))_{k\in\mathbb{N}}$ est bornée dans $\mathbb{R}^n$ et on peut en extraire une sous-suite $\left(\theta\left(T_{\varphi(k)}\right)\right)_{k\in\mathbb{N}}$ qui converge vers $\lambda = (\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$. D'autre part, le polynôme caractéristique d'une matrice $M \in \mathcal{T}_n(\mathbb{R})$ s'écrit sous la forme $P_M(X) = \sum_{j=0}^{n} a_j X^j$, les coefficients a_j étant des fonctions continues de $\theta(M)$ (fonctions polynomiales symétriques élémentaires des racines). De $\lim_{k\to+\infty} \theta\left(T_{\varphi(k)}\right) = \lambda$, on déduit alors que $\lim_{k\to+\infty} P_{T_{\varphi(k)}}(X) = \prod_{j=1}^{n} (\lambda_j - X)$. L'application qui associe à une matrice M son polynôme caractéristique est continue (ses composantes sont des fonctions polynomiales des m_{ij}), donc $\lim_{k\to+\infty} P_{T_{\varphi(k)}} = P_T$ et le polynôme $P_T(X) = \prod_{j=1}^{n} (\lambda_j - X)$ est scindé dans $\mathbb{R}[X]$, ce qui entraîne que $T \in \mathcal{T}_n(\mathbb{R})$. On a donc ainsi prouvé que $T_n(\mathbb{R})$ est fermé dans $\mathcal{M}_n(\mathbb{R})$.

2. On a $\mathcal{D}_n(\mathbb{R}) \subset \mathcal{T}_n(\mathbb{R})$ avec $\mathcal{T}_n(\mathbb{R})$ fermé, donc $\overline{\mathcal{D}}_n(\mathbb{R}) \subset \mathcal{T}_n(\mathbb{R})$. Une démonstration identique à celle du théorème 4.12 permet de montrer l'inclusion $\mathcal{T}_n(\mathbb{R}) \subset \overline{\mathcal{D}}_n(\mathbb{R})$. D'où l'égalité $\overline{\mathcal{D}}_n(\mathbb{R}) = \mathcal{T}_n(\mathbb{R})$.

Exercice 4.13 : *Soient A et B deux matrices normales dans $\mathcal{M}_n(\mathbb{C})$. Montrer que :*

$$\rho(AB) \leq \rho(A)\,\rho(B).$$

Solution : On a $\rho(AB) \leq \|AB\|_2 \leq \|A\|_2\|B\|_2 = \rho(A)\,\rho(B)$ si A et B sont normales.

Exercice 4.14 : *Soient $\sum\limits_{k=0}^{+\infty} a_k z^k$ une série entière de rayon de convergence $R > 0$ et $A \in \mathcal{M}_n(\mathbb{K})$. Montrer que si $\rho(A) < R$ alors la série de terme général $a_k A^k$ est convergente et que si $\rho(A) > R$, elle est divergente.*

Solution : Supposons que $\rho(A) < R$. On sait qu'il existe une norme matricielle $\|\cdot\|$ subordonnée à une norme vectorielle telle que $\rho(A) \leq \|A\| < R$ (théorème 4.14). On a alors :

$$\sum_{k=0}^{+\infty} |a_k|\,\|A\|^k < +\infty$$

et, avec $\left\|A^k\right\| \leq \|A\|^k$ pour tout entier naturel k, on déduit que la série de terme général $a_k A^k$ est normalement convergente, donc convergente, dans l'espace complet $\mathcal{M}_n(\mathbb{K})$.

Si $\rho(A) > R$, alors la suite $\left(|a_k|\,\rho(A)^k\right)_{k\in\mathbb{N}}$ n'est pas bornée. Avec $\rho(A)^k = \rho\left(A^k\right)$ et $|a_k|\left\|A^k\right\| \geq |a_k|\,\rho\left(A^k\right)$, on en déduit que la suite $\left(|a_k|\left\|A^k\right\|\right)_{k\in\mathbb{N}}$ n'est pas bornée et la série de terme général $a_k A^k$ est divergente.

Exercice 4.15 : *On désigne par $\mathcal{F}$ l'ensemble des matrices non inversibles de $\mathcal{M}_n(\mathbb{C})$. On munit $\mathcal{M}_n(\mathbb{C})$ de la norme matricielle induite par la norme hermitienne canonique de $\mathbb{C}^n$ et pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, on note :*

$$d(A, \mathcal{F}) = \inf_{M\in\mathcal{F}} \|A - M\|_2.$$

1. Montrer que pour toute matrice $M \in \mathcal{F}$ on peut trouver un vecteur unitaire x tel que :

$$\frac{1}{\left\|A^{-1}\right\|_2} \leq \|(A - M)\,x\|_2.$$

2. Soient $A \in GL_n(\mathbb{C})$ et $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$ les valeurs propres de la matrice A^*A avec $0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. On rappelle qu'on peut écrire $A = UDV^*$ où U, V sont deux matrices unitaires et :

$$D = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}$$

(exercice 3.11).

(a) Montrer que :

$$\left\|A^{-1}\right\|_2 = \frac{1}{\lambda_1}.$$

(b) On pose $M_0 = UD_0V^*$, avec :

$$D_0 = \begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{pmatrix}.$$

Montrer que $\|A - M_0\|_2 = \lambda_1$.

(c) Déduire de ce qui précède que $d(A, \mathcal{F}) = \dfrac{1}{\|A^{-1}\|_2}$.

Solution :

1. Si $M \in \mathcal{F}$, alors 0 est valeur propre de M et il existe un vecteur unitaire x tel que $Mx = 0$. On a alors :

$$\|(A - M)x\|_2 = \|Ax\|_2 = \frac{\left\|A^{-1}\right\|_2 \|Ax\|_2}{\|A^{-1}\|_2} \geq \frac{\left\|A^{-1}Ax\right\|_2}{\|A^{-1}\|_2} = \frac{1}{\|A^{-1}\|_2}$$

et $\dfrac{1}{\|A^{-1}\|_2} \leq \|A - M\|_2$ pour toute $M \in \mathcal{F}$. Donc $\dfrac{1}{\|A^{-1}\|_2} \leq d(A, \mathcal{F})$.

2. La matrice A^*A est hermitienne définie positive. Ces valeurs propres sont donc toutes réelles strictement positives. On peut donc les noter $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$.

(a) On a :

$$\begin{aligned} \|A^{-1}\|_2^2 &= \left\|\left(A^{-1}\right)^* A^{-1}\right\|_2 = \left\|(AA^*)^{-1}\right\|_2 \\ &= \rho\left((AA^*)^{-1}\right) = \frac{1}{\min\limits_{1 \leq k \leq n} (\lambda_k^2)}. \end{aligned}$$

Ce qui donne $\left\|A^{-1}\right\|_2 = \dfrac{1}{\lambda_1}$.

(b) On a $M_0 \in F$ et :

$$\begin{aligned}\|A - M_0\|_2^2 &= \rho\left((A - M_0)(A - M_0)^*\right) = \rho\left(U(D - D_0)^2 U^*\right) \\ &= \rho\left((D - D_0)^2\right) = \lambda_1^2.\end{aligned}$$

C'est-à-dire que $\|A - M_0\|_2 = \lambda_1 = \dfrac{1}{\|A^{-1}\|_2}$.

(c) On peut donc conclure à l'égalité $d(A, \mathcal{F}) = \dfrac{1}{\|A^{-1}\|_2}$.

Exercice 4.16 : *On désigne par A la matrice réelle d'ordre n supérieur ou égal à 2 définie par :*

$$A = \begin{pmatrix} 1 & 0 & \ldots & 0 & -1 \\ -1 & 1 & 0 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & -1 & 1 & 0 \\ 0 & \ldots & 0 & -1 & 1 \end{pmatrix}.$$

1. Calculer les valeurs propres de ${}^tAA$.

2. Calculer $\|A\|_2$.

Solution :

1. On a :

$${}^tAA = \begin{pmatrix} 2 & -1 & 0 & \ldots & -1 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ -1 & \ldots & 0 & -1 & 2 \end{pmatrix}.$$

Les valeurs propres de la matrice symétrique réelle ${}^tAA$ sont réelles et le théorème de Gerschgörin-Hadamard nous dit que toute valeur propre λ de cette matrice est telle que $|\lambda - 2| \leq 2$. On peut donc écrire une telle valeur propre sous la forme :

$$\lambda = 2\big(1 - \cos(\theta)\big) = 4\sin^2\left(\frac{\theta}{2}\right)$$

avec $\theta \in [0, \pi]$.

De $\det(A) = 0$ (en développant suivant la première ligne) on déduit que 0 est valeur propre de ${}^tAA$, ce qui correspond à $\theta = 0$ ou $\theta = \pi$.

Si λ est une valeur propre non nulle, on a $\theta \in]0, \pi[$ et un vecteur propre associé x de coordonnées x_k $(1 \leq k \leq n)$ est défini par les relations de récurrence :

$$x_{k-1} + (\lambda - 2)\, x_k + x_{k+1} = 0 \quad (1 \leq k \leq n) \tag{4.8}$$

avec les conditions aux limites $x_0 = x_n$ et $x_{n+1} = x_1$. Tout revient à chercher les suites réelles $(x_k)_{k \in \mathbb{N}}$ périodiques de période n et vérifiant la récurrence (4.8). Le polynôme caractéristique de cette récurrence, $P(r) = r^2 - 2\cos(\theta)\, r + 1$, a pour racines $r_1 = e^{i\theta}$ et $r_2 = e^{-i\theta}$. On obtient donc les solutions définies par :

$$\forall k \in \mathbb{N}, \quad x_k = \alpha e^{ik\theta} + \beta e^{-ik\theta},$$

les coefficients α et β étant tels que $x_k \in \mathbb{R}$ et $x_{k+n} = x_n$ pour tout entier naturel k. Des conditions $x_k \in \mathbb{R}$ pour tout entier k on déduit que $(\alpha - \beta)\sin(k\theta) = 0$ et $\alpha = \beta$. La condition $x_0 = x_n$ donne $2\alpha\cos(n\theta) = 2\alpha$ et, si on s'intéresse à une solution non nulle, on a nécessairement $\alpha \neq 0$ et $\cos(n\theta) = 1$, ce qui donne $\theta = j\dfrac{2\pi}{n}$ avec $0 \leq j \leq n-1$. On a donc ainsi montré que la matrice ${}^tAA$ a n valeurs propres simples données par :

$$\lambda_j = 4\sin^2\left(j\frac{\pi}{n}\right) \quad (0 \leq j \leq n-1).$$

2. On a :

$$\|A\|_2 = \sqrt{\rho\left({}^tAA\right)} = \begin{cases} 2 & \text{si } n = 2p, \\ 2\sin\left(p\dfrac{2\pi}{2p+1}\right) & \text{si } n = 2p+1. \end{cases}$$

Exercice 4.17 : *Montrer les inégalités suivantes pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$:*

1. $\|A\|_2 \leq \sqrt{\|A\|_\infty \|A\|_1}$.

2. $\dfrac{1}{\sqrt{n}}\|A\|_\infty \leq \|A\|_2 \leq \sqrt{n}\|A\|_\infty$.

3. $\dfrac{1}{\sqrt{n}}\|A\|_1 \leq \|A\|_2 \leq \sqrt{n}\|A\|_1$.

Solution :

1. On note $\lambda_1^2, \lambda_2^2, \ldots, \lambda_n^2$ les valeurs propres de la matrice A^*A avec :

$$0 \leq \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$$

et on a $\|A\|_2 = \lambda_n$. En désignant par e_n un vecteur propre de A associé à la valeur propre λ_n et unitaire pour $\|\cdot\|_\infty$, on a :

$$\|A\|_2^2 = |\lambda_n|^2 = \left\|\lambda_n^2 e_n\right\|_\infty = \|A^*Ae_n\|_\infty \leq \|A^*\|_\infty \|A\|_\infty.$$

Puis, avec $\|A^*\|_\infty = \|A\|_1$, on déduit que $\|A\|_2 \leq \sqrt{\|A\|_\infty \|A\|_1}$, l'égalité étant réalisée pour $A = I_n$.

2. Pour tout x dans $\mathbb{C}^n$ on a :

$$\|Ax\|_2^2 = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij} x_j \right|^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}| \, |x_j| \right)^2$$

et, en utilisant l'inégalité de Cauchy-Schwarz, on déduit que :

$$\|Ax\|_2^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^2 \right) \left(\sum_{j=1}^n |x_j|^2 \right) = \|x\|_2^2 \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^2 \right).$$

Il en résulte que :

$$\|A\|_2^2 \leq \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}|^2 \right) \leq \sum_{i=1}^n \left(\sum_{j=1}^n |a_{ij}| \right)^2 \leq n \left(\max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \right)^2,$$

soit :

$$\|A\|_2 \leq \sqrt{n} \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| = \sqrt{n} \, \|A\|_\infty .$$

D'autre part il existe un indice i tel que $\|A\|_\infty = \sum\limits_{j=1}^n |a_{ij}|$ et en utilisant l'inégalité de Cauchy-Schwarz on obtient :

$$\|A\|_\infty = \sum_{j=1}^n |a_{ij}| \cdot 1 \leq \sqrt{n} \sqrt{\sum_{j=1}^n |a_{ij}|^2}.$$

En désignant par $\{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{C}^n$ on a :

$$\sqrt{\sum_{j=1}^n |a_{ij}|^2} = \|A^* e_i\|_2 .$$

Ce qui donne en définitive :

$$\|A\|_\infty \leq \sqrt{n} \, \|A^* e_i\|_2 \leq \sqrt{n} \, \|A^*\|_2 = \sqrt{n} \, \|A\|_2 .$$

3. Ces inégalités se déduisent des précédentes avec :

$$\|A^*\|_\infty = \|A\|_1 , \quad \|A^*\|_2 = \|A\|_2 .$$

Exercice 4.18 : *Soient α, β des réels et $A(\alpha,\beta) = ((a_{ij}))_{1\leq i,j\leq n}$ la matrice réelle d'ordre n supérieur ou égal à 3 définie par :*

$$\forall i \in \{1,2,\ldots,n\}, \begin{cases} a_{ii} = \beta, \\ a_{ij} = \alpha & \text{si } j \in \{1,2,\ldots,n\} - \{i\}. \end{cases}$$

1. Calculer le rayon spectral $\rho(\alpha,\beta)$ de $A(\alpha,\beta)$.

2. On suppose que $\alpha\beta \geq 0$. Comparer $\|A(\alpha,\beta)\|_\infty$, $\|A(\alpha,\beta)\|_1$ et $\|A(\alpha,\beta)\|_2$.

Solution :

1. On a vu (exercice 2.8) que les valeurs propres de la matrice $A(\alpha,\beta)$ sont données par $\lambda_1 = \beta + (n-1)\alpha$ et $\lambda_2 = \beta - \alpha$. Le rayon spectral de $A(\alpha,\beta)$ est donc :

$$\rho(\alpha,\beta) = \max\{|\beta + (n-1)\alpha|, |\beta - \alpha|\}.$$

Ce qui donne quatre possibilités :

(a) $\alpha \leq \beta$, $\beta > -(n-1)\alpha$. Alors :

$$\begin{aligned}\rho(\alpha,\beta) &= \max\{\beta + (n-1)\alpha, \beta - \alpha\} \\ &= \begin{cases} \beta + (n-1)\alpha & \text{si } \alpha > 0, \\ \beta - \alpha & \text{si } \alpha \leq 0. \end{cases}\end{aligned}$$

(b) $\alpha \leq \beta$, $\beta \leq -(n-1)\alpha$. Alors :

$$\begin{aligned}\rho(\alpha,\beta) &= \max\{-\beta - (n-1)\alpha, \beta - \alpha\} \\ &= \begin{cases} -\beta - (n-1)\alpha & \text{si } \beta < -\dfrac{n-2}{2}\alpha, \\ \beta - \alpha & \text{si } \beta \geq -\dfrac{n-2}{2}\alpha. \end{cases}\end{aligned}$$

(c) $\alpha > \beta$, $\beta > -(n-1)\alpha$. Alors :

$$\begin{aligned}\rho(\alpha,\beta) &= \max\{\beta + (n-1)\alpha, \alpha - \beta\} \\ &= \begin{cases} \beta + (n-1)\alpha & \text{si } \beta > -\dfrac{n-2}{2}\alpha, \\ \alpha - \beta & \text{si } \beta \leq -\dfrac{n-2}{2}\alpha. \end{cases}\end{aligned}$$

(d) $\alpha > \beta$, $\beta \leq -(n-1)\alpha$. Alors :

$$\begin{aligned}\rho(\alpha,\beta) &= \max\{-\beta - (n-1)\alpha, \alpha - \beta\} \\ &= \begin{cases} -\beta - (n-1)\alpha & \text{si } \alpha < 0, \\ \alpha - \beta & \text{si } \alpha \geq 0. \end{cases}\end{aligned}$$

Ce qui peut être résumé avec la figure 4.2, page ci-contre.

2. La matrice $A(\alpha,\beta)$ est symétrique réelle, donc $\rho(\alpha,\beta) = \|A(\alpha,\beta)\|_2$. Si $\alpha\beta \geq 0$, on a :

$$\rho(\alpha,\beta) = |\beta| + (n-1)|\alpha| = \|A(\alpha,\beta)\|_2 = \|A(\alpha,\beta)\|_\infty = \|A(\alpha,\beta)\|_1.$$

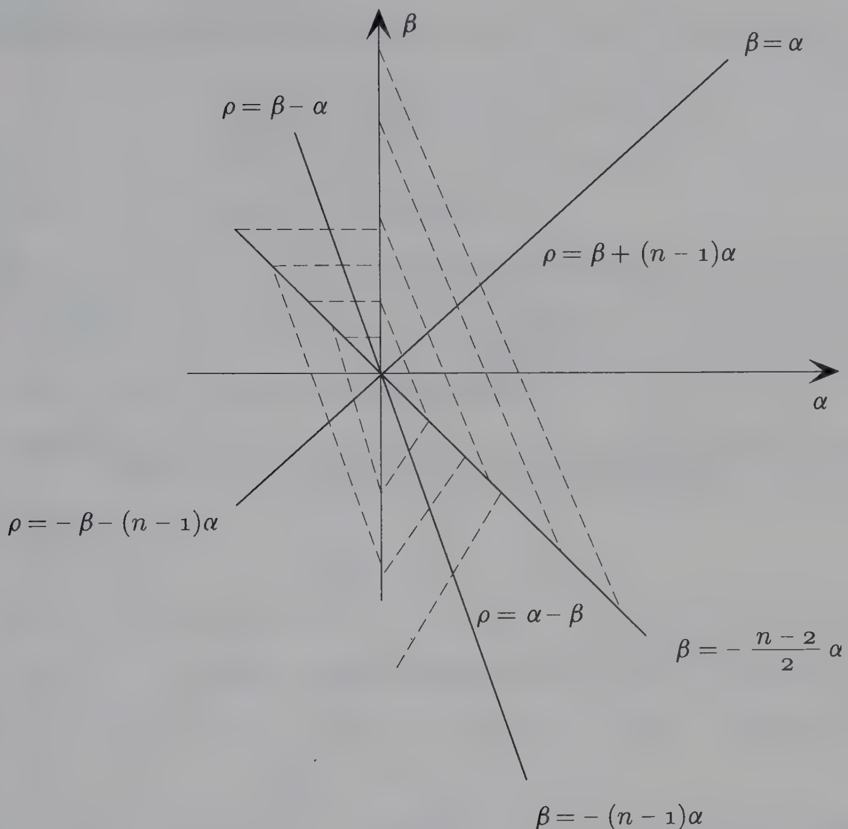

Figure 4.2.

Exercice 4.19 : *Calculer le rayon spectral et le conditionnement pour la norme matricielle induite par la norme euclidienne de la matrice :*

$$A = \begin{pmatrix} 2 & -1 & 0 & 0 & \dots & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 0 & \dots & 0 & -1 & 2 & -1 \\ 0 & \dots & 0 & 0 & -1 & 2 \end{pmatrix}.$$

Solution : On a vu (exercice 2.10) que les valeurs propres de A sont :

$$\lambda_k = 2 - 2\cos\left(\frac{k\pi}{n+1}\right) = 4\sin^2\left(\frac{k\pi}{2(n+1)}\right) \quad (1 \le k \le n).$$

Le rayon spectral de A est donné par :

$$\rho(A) = 4\sin^2\left(\frac{n\pi}{2(n+1)}\right)$$

et le conditionnement pour la norme matricielle induite par la norme euclidienne par :

$$\operatorname{cond}_2(A) = \frac{\lambda_1}{\lambda_n} = \frac{\sin^2\left(\dfrac{\pi}{2(n+1)}\right)}{\sin^2\left(\dfrac{n\pi}{2(n+1)}\right)}$$

(la matrice A est symétrique définie positive).

On peut remarquer que $\lim\limits_{n\to+\infty} \operatorname{cond}_2(A) = 0$, c'est-à-dire que la matrice A est mal conditionnée pour les grandes valeurs de n.

Exercice 4.20 : *Pour tout entier n supérieur ou égal à 2 l'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique avec le produit scalaire :*

$$(x, y) \mapsto \langle x \mid y\rangle = \sum_{i=1}^{n} x_i y_i$$

de norme associée notée $\|\cdot\|_2$.

On désigne par H_n la matrice de Hilbert d'ordre n définie par :

$$H_n = \left(\left(\frac{1}{i+j-1}\right)\right)_{1\leq i,j\leq n}.$$

1. Montrer que H_n est la matrice dans la base canonique de $\mathbb{R}_{n-1}[X]$ du produit scalaire défini par :

$$(P, Q) \mapsto \langle P \mid Q\rangle = \int_0^1 P(t)\, Q(t)\ \mathrm{dt}.$$

2. Qu'en déduit-on pour les valeurs propres de H_n ?

On désigne par R_n le quotient de Rayleigh-Ritz associé à la matrice H_n et par ρ_n son rayon spectral.

Dans ce qui suit on écrit la matrice H_{n+1} sous la forme :

$$H_{n+1} = \begin{pmatrix} H_n & C_n \\ {}^tC_n & \dfrac{1}{2n+1} \end{pmatrix},$$

avec :

$$C_n = \begin{pmatrix} \dfrac{1}{n+1} \\ \vdots \\ \dfrac{1}{2n} \end{pmatrix}.$$

Un vecteur x' de $\mathbb{R}^{n+1}$ est écrit sous la forme :

$$x' = \begin{pmatrix} x \\ x_{n+1} \end{pmatrix},$$

avec $x \in \mathbb{R}^n$ et $x_{n+1} \in \mathbb{R}$.

3. *Montrer que la suite $(\rho_n)_{n\geq 2}$ est croissante.*

4. *Soit $x' \in \mathbb{R}^{n+1}$ tel que $\|x'\|_2 = 1$ et $x \neq 0$. Montrer que :*

$$R_{n+1}(x') \leq \rho_n \|x\|_2^2 + 2|x_{n+1}| \|C_n\|_2 \|x\|_2 + \frac{1}{2n+1} x_{n+1}^2.$$

5. *On désigne par q la forme quadratique définie sur $\mathbb{R}^2$ par :*

$$\forall (u,v) \in \mathbb{R}^2, \quad q(u,v) = \rho_n u^2 + 2\|C_n\|_2 uv + \frac{1}{2n+1} v^2.$$

Calculer $\sup\limits_{\substack{(u,v)\in\mathbb{R}^2 \\ u^2+v^2=1}} q(u,v)$.

6. *Déduire de ce qui précède que :*

$$\rho_{n+1} \leq \rho_n + \|C_n\|_2 \leq \rho_n + \frac{1}{\sqrt{2n}}.$$

Solution : On désigne par $\{e_0, \ldots, e_{n-1}\}$ la base canonique de $\mathbb{R}_{n-1}[X]$ définie par :

$$\forall i \in \{0, \ldots, n-1\}, \quad e_i(X) = X^i.$$

1. La matrice du produit scalaire $(P, Q) \mapsto \langle P \mid Q \rangle$ dans la base canonique de $\mathbb{R}_{n-1}[X]$ est la matrice de coefficients :

$$\langle e_{i-1} \mid e_{j-1} \rangle = \int_0^1 t^{i+j-2}\, \mathrm{d}t = \frac{1}{i+j-1} \quad (1 \leq i,j \leq n).$$

C'est bien la matrice de Hilbert d'ordre n.

2. On en déduit alors que la matrice réelle H_n est symétrique définie positive et ses valeurs propres sont réelles strictement positives.

3. La matrice H_n étant symétrique réelle de plus grande valeur propre ρ_n, on a (théorème 4.21) :

$$\rho_n = \sup_{x\in\mathbb{R}^n-\{0\}} R_n(x).$$

Soit $x \in \mathbb{R}^n - \{0\}$ et $x' \in \mathbb{R}^{n+1}$ défini par $x'_j = x_j$ pour j compris entre 1 et n et $x_{n+1} = 0$. On a alors $x' \neq 0$ et :

$$R_n(x) = R_{n+1}(x') \leq \rho_{n+1}.$$

Le vecteur x étant quelconque dans $\mathbb{R}^n - \{0\}$ on en déduit que $\rho_n \leq \rho_{n+1}$. La suite $(\rho_n)_{n\geq 2}$ est donc croissante.

4. On a :

$$\langle H_{n+1}x' \mid x' \rangle = \langle H_n x + x_{n+1} C_n \mid x \rangle + x_{n+1} \left(\langle C_n \mid x \rangle + \frac{x_{n+1}}{2n+1} \right),$$

soit :

$$\langle H_{n+1}x' \mid x' \rangle = \langle H_n x \mid x \rangle + 2x_{n+1} \langle C_n \mid x \rangle + \frac{x_{n+1}^2}{2n+1} \cdot$$

Ce qui s'écrit, $x' \in \mathbb{R}^{n+1}$ tel que $\|x'\|_2 = 1$ et $x \neq 0$:

$$R_{n+1}(x') = R_n(x) \|x\|_2^2 + 2x_{n+1} \langle C_n \mid x \rangle + \frac{x_{n+1}^2}{2n+1} \cdot$$

En utilisant l'inégalité de Cauchy-Schwarz on a :

$$x_{n+1} \langle C_n \mid x \rangle \leq |x_{n+1}| \, |\langle C_n \mid x \rangle| \leq |x_{n+1}| \, \|C_n\|_2 \, \|x\|_2$$

et, avec $\rho_n = \sup\limits_{x \in \mathbb{R}^n - \{0\}} R_n(x)$, on aboutit à :

$$R_{n+1}(x') \leq \rho_n \|x\|_2^2 + 2 |x_{n+1}| \, \|C_n\|_2 \, \|x\|_2 + \frac{1}{2n+1} x_{n+1}^2.$$

5. La matrice de la forme quadratique q dans la base canonique de $\mathbb{R}^2$ est donnée par :

$$A = \begin{pmatrix} \rho_n & \|C_n\|_2 \\ \|C_n\|_2 & \dfrac{1}{2n+1} \end{pmatrix}.$$

Ces valeurs propres sont :

$$\frac{1}{2} \left(\rho_n + \frac{1}{2n+1} \pm \sqrt{\left(\rho_n - \frac{1}{2n+1} \right)^2 + 4 \|C_n\|_2^2} \right).$$

On en déduit que :

$$\sup_{\substack{(u,v) \in \mathbb{R}^2 \\ u^2+v^2=1}} q(u,v) = \frac{1}{2} \left(\rho_n + \frac{1}{2n+1} + \sqrt{\left(\rho_n - \frac{1}{2n+1} \right)^2 + 4 \|C_n\|_2^2} \right).$$

6. Soit $x' \in \mathbb{R}^{n+1}$ tel que $R_{n+1}(x') = \rho_{n+1}$ et $\|x'\|_2 = 1$. On a $x' = \begin{pmatrix} x \\ x_{n+1} \end{pmatrix}$ avec $x \neq 0$ et de ce qui précède on déduit que :

$$\rho_{n+1} \leq q(\|x\|_2, |x_{n+1}|) \leq \sup_{\substack{(u,v) \in \mathbb{R}^2 \\ u^2+v^2=1}} q(u,v).$$

En écrivant que :

$$\sqrt{\left(\rho_n - \frac{1}{2n+1}\right)^2 + 4\|C_n\|_2^2} \leq \left|\rho_n - \frac{1}{2n+1}\right| + 2\|C_n\|_2$$

et en remarquant que :

$$\rho_n \geq \rho_2 = \frac{4+\sqrt{13}}{6} \geq 1,$$

on déduit que :

$$\left|\rho_n - \frac{1}{2n+1}\right| = \rho_n - \frac{1}{2n+1}$$

et

$$\rho_{n+1} \leq \rho_n + \|C_n\|_2 .$$

Enfin, avec :

$$\|C_n\|_2^2 = \sum_{k=1}^{n} \frac{1}{(n+k)^2} \leq \int_n^{2n} \frac{\mathrm{d}t}{t^2} = \frac{1}{2n},$$

on déduit que :

$$\rho_{n+1} \leq \rho_n + \frac{1}{\sqrt{2n}} .$$

Exercice 4.21 : *On reprend les notations de l'exercice 4.20. On munit l'espace vectoriel $\mathbb{R}[X]$ du produit scalaire défini par :*

$$(P,Q) \mapsto \langle P \mid Q \rangle = \int_0^1 P(t)\,Q(t)\,\mathrm{d}t,$$

la norme associée étant notée $\|\cdot\|$.

On note $\{e_k \mid k \in \mathbb{N}\}$ la base canonique de $\mathbb{R}[X]$ définie par :

$$\forall k \in \mathbb{N}, \quad e_k(X) = X^k.$$

Pour tout entier n strictement positif on désigne par δ_n la distance de e_n à $\mathbb{R}_{n-1}[X]$. On rappelle qu'il existe un unique polynôme $P \in \mathbb{R}_{n-1}[X]$ tel que $\delta_n = \|e_n - P\|$ et que ce polynôme est aussi caractérisé par $e_n - P \in (\mathbb{R}_{n-1}[X])^{\perp}$. On note :

$$P(X) = \sum_{j=0}^{n-1} a_j X^j$$

et on associe à ce polynôme la fonction rationnelle f définie par :

$$f(t) = \frac{1}{t+n+1} - \sum_{j=0}^{n-1} \frac{a_j}{t+j+1} .$$

1. *Calculer $f(k)$ pour tout entier k compris entre 0 et n.*

2. *En écrivant la fonction f sous la forme :*

$$f(t) = \frac{Q(t)}{\prod_{j=1}^{n+1}(t+j)},$$

en déduire la valeur de δ_n.

3. *On désigne par μ_n la plus petite valeur propre de la matrice de Hilbert H_n et on note :*

$$x = \begin{pmatrix} -a_0 \\ \vdots \\ -a_{n-1} \\ 1 \end{pmatrix}.$$

(a) *Montrer que :*

$$\mu_{n+1} \leq \frac{\delta_n^2}{\|x\|_2^2}.$$

(b) *Montrer que :*

$$\mu_n \leq \frac{1}{2n-1}\frac{1}{\left(C_{2(n-1)}^{n-1}\right)^2} \leq \frac{1}{12}\frac{1}{15^{n-1}}.$$

(c) *Montrer que :*

$$\mathrm{cond}_2(H_n) \leq \frac{1}{2\left(4+\sqrt{13}\right)}\frac{1}{15^{n-1}}.$$

Solution :

1. Pour tout entier naturel k on a :

$$f(k) = \int_0^1 t^{n+k}\,\mathrm{d}t - \sum_{j=0}^{n-1} a_j \int_0^1 t^{j+k}\,\mathrm{d}t = \langle e_n - P \mid e_k \rangle.$$

De $e_n - P \in (\mathbb{R}_{n-1}[X])^{\perp}$ on déduit que $f(k) = 0$ pour tout entier k compris entre 1 et $n-1$ et :

$$f(n) = \langle e_n - P \mid e_n \rangle = \|e_n - P\|^2 = \delta_n^2.$$

2. En réduisant au même dénominateur, on peut écrire :

$$f(t) = \frac{Q(t)}{\prod_{j=1}^{n+1}(t+j)},$$

avec $Q \in \mathbb{R}_n[X]$. La fonction f s'annulant en $0, 1, \ldots, n-1$, il en est de même du polynôme Q, et ce dernier s'écrit :

$$Q(t) = \alpha_n \prod_{j=0}^{n-1} (t-j).$$

En utilisant le fait que :

$$\lim_{t \to -(n+1)} (t+n+1) f(t) = 1,$$

on déduit que :

$$\alpha_n = \frac{(n!)^2}{(2n)!}$$

et

$$\delta_n^2 = f(n) = \frac{(n!)^4}{(2n+1)\left((2n)!\right)^2}.$$

On a donc :

$$\delta_n = \frac{(n!)^2}{\sqrt{2n+1}\,(2n)!}.$$

3. On désigne par R_{n+1} le quotient de Rayleigh-Ritz associé à la matrice symétrique réelle H_{n+1}.

(a) On sait alors (théorème 4.21) que :

$$\mu_{n+1} = \inf_{y \in \mathbb{R}^{n+1} - \{0\}} R_{n+1}(y) \leq R_{n+1}(x),$$

avec :

$$R_{n+1}(x) = \frac{\langle H_{n+1}x \mid x\rangle}{\|x\|_2^2} = \frac{{}^t x H_{n+1} x}{\|x\|_2^2} = \frac{\|e_n - P\|^2}{\|x\|_2^2} = \frac{\delta_n^2}{\|x\|_2^2}$$

(H_{n+1} est la matrice dans la base canonique du produit scalaire de $\mathbb{R}_n[X]$). On a donc bien :

$$\mu_{n+1} \leq \frac{\delta_n^2}{\|x\|_2^2}.$$

(b) Du fait que $\|x\|_2 \geq 1$ on déduit de ce qui précède que :

$$\mu_{n+1} \leq \delta_n^2 = \frac{(n!)^4}{(2n+1)\left((2n)!\right)^2} = \frac{1}{2n+1}\frac{1}{\left(C_{2n}^n\right)^2}.$$

D'autre part avec :

$$\frac{\delta_{n+1}^2}{\delta_n^2} = \frac{(n+1)^2}{4(2n+1)(2n+3)}$$

et avec :

$$\begin{aligned} 4(2n+1)(2n+3) &= 16n^2 + 32n + 12 \\ &\geq 15\left(n^2 + 2n + 1\right) \end{aligned}$$

on déduit que :

$$\delta_{n+1}^2 \leq \frac{1}{15}\delta_n^2$$

et

$$\delta_n^2 \leq \left(\frac{1}{15}\right)^{n-1} \delta_1^2 = \frac{1}{12}\left(\frac{1}{15}\right)^{n-1}.$$

On a donc :

$$\mu_n \leq \frac{1}{12}\,\frac{1}{15^{n-1}}.$$

(c) La matrice H_n étant symétrique réelle et la suite $(\rho_n)_{n\geq 2}$ étant croissante (exercice 4.20), on déduit que :

$$\mathrm{cond}_2(H_n) = \frac{\mu_n}{\rho_n} \leq \frac{\mu_n}{\rho_2} \leq \frac{1}{2\left(4+\sqrt{13}\right)}\,\frac{1}{15^{n-1}}.$$

Exercice 4.22 : *En utilisant la décomposition de Dunford $A = D + N$ avec D diagonalisable, N nilpotente et $DN = ND$, montrer que pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ on a :*

$$\rho(A) = \lim_{k\to+\infty}\left(\left\|A^k\right\|^{\frac{1}{k}}\right).$$

Solution : On a la décomposition de Dunford $A = D + N$ avec D diagonalisable de mêmes valeurs propres que A, N nilpotente et $DN = ND$. Pour tout entier $k \geq n$, on a $N^k = 0$ et

$$A^k = (D+N)^k = \sum_{j=0}^{n} C_k^j D^{k-j} N^j = D^{k-n} \sum_{j=0}^{n} C_k^j D^{n-j} N^j.$$

Ce qui donne, en posant $\alpha = \max\limits_{0\leq j\leq n}\left\{\|D\|^{n-j}\|N\|^j\right\}$:

$$\left\|A^k\right\| \leq \left\|D^{k-n}\right\| \sum_{j=0}^{n} C_k^j \|D\|^{n-j}\|N\|^j \leq \alpha\left(\sum_{j=0}^{n} C_k^j\right)\left\|D^{k-n}\right\|.$$

Pour tout $j \in \{0, 1, \ldots, n\}$ on a :

$$C_k^j \leq k(k-1)\ldots(k-j+1) \leq k^j \leq k^n,$$

donc :

$$\left\| A^k \right\| \leq \alpha(n+1) k^n \left\| D^{k-n} \right\|$$

et

$$\forall k \geq n, \quad \rho(A) \leq \left\| A^k \right\|^{\frac{1}{k}} \leq \left(\alpha(n+1) \right)^{\frac{1}{k}} k^{\frac{n}{k}} \left\| D^{k-n} \right\|^{\frac{1}{k}},$$

avec $\lim\limits_{k \to +\infty} \left(\alpha(n+1) \right)^{\frac{1}{k}} k^{\frac{n}{k}} = 1$ et

$$\begin{aligned} \lim_{k \to +\infty} \left\| D^{k-n} \right\|^{\frac{1}{k}} &= \lim_{k \to +\infty} \left(\left\| D^{k-n} \right\| \frac{1}{k-n} \right)^{\frac{k-n}{k}} \\ &= \lim_{k \to +\infty} \left(\left\| D^{k-n} \right\|^{\frac{1}{k-n}} \right) = \rho(D) = \rho(A). \end{aligned}$$

Donc $\lim\limits_{k \to +\infty} \left(\left\| A^k \right\|^{\frac{1}{k}} \right) = \rho(A)$.

Exercice 4.23 : *Pour toute matrice A dans $\mathcal{M}_n(\mathbb{C})$ on désigne par $\mathcal{H}(A)$ le hausdorffien de A.*

1. Montrer que si A est diagonale alors $\mathcal{H}(A)$ est l'enveloppe convexe du spectre de A.

2. Montrer que pour toute matrice A dans $\mathcal{M}_n(\mathbb{C})$ et toute matrice unitaire U on a $\mathcal{H}(U^* A U) = \mathcal{H}(A)$.

3. Montrer que pour toute matrice normale A, $\mathcal{H}(A)$ est l'enveloppe convexe du spectre de A.

Solution :

1. Soit A une matrice diagonale de valeurs propres $\lambda_1, \ldots, \lambda_n$. Pour tout vecteur x dans la sphère unité de $(\mathbb{C}^n, \|\cdot\|_2)$, on a :

$$R_A(x) = \langle Ax \mid x \rangle = \sum_{i=1}^{n} |x_i|^2 \lambda_i,$$

avec $0 \leq |x_i|^2 \leq 1$ et $\sum_{i=1}^{n} |x_i|^2 = 1$. Il en résulte que $\mathcal{H}(A)$ est l'enveloppe convexe du spectre de A.

2. Pour tout vecteur x dans la sphère unité S_1 de $(\mathbb{C}^n, \|\cdot\|_2)$, on a :

$$\langle U^* A U x \mid x \rangle = \langle Ay \mid y \rangle$$

avec $y = Ux$ qui décrit tout S_1 quand x décrit S_1 pour U unitaire. On déduit alors que $\mathcal{H}(U^* A U) = \mathcal{H}(A)$.

3. Si A est une matrice normale elle se diagonalise alors dans une base orthonormée, c'est-à-dire qu'il existe une matrice unitaire U telle que la matrice $D = U^* AU$ soit diagonale. On a alors $\mathcal{H}(A) = \mathcal{H}(D)$ qui est égale à l'enveloppe convexe du spectre de D. Comme A et D ont même spectre on déduit que $\mathcal{H}(A)$ est l'enveloppe convexe du spectre de A.

Dans le cas où la matrice A est hermitienne on retrouve l'égalité $\mathcal{H}(A) = [\lambda_1, \lambda_n]$ où λ_1 est la plus petite valeur propre de A et λ_n la plus grande.

Exercice 4.24 : *Soit $A : [0,1] \to S_2(\mathbb{R})$ définie par :*

$$\forall t \in [0,1], \quad A(t) = \begin{pmatrix} a(t) & b(t) \\ b(t) & -a(t) \end{pmatrix},$$

avec :

$$a(t) = \begin{cases} 0 & si\ t = 0, \\ e^{-\frac{1}{t^2}} \cos\left(\frac{1}{t}\right) & si\ t \neq 0, \end{cases} \qquad b(t) = \begin{cases} 0 & si\ t = 0, \\ e^{-\frac{1}{t^2}} \sin\left(\frac{1}{t}\right) & si\ t \neq 0. \end{cases}$$

1. Montrer que A est continue et pour tout t dans $[0,1]$ calculer les valeurs et vecteurs propres de $A(t)$.

2. Montrer que les vecteurs propres de $A(t)$ ne définissent pas des fonctions continues en 0.

Solution :

1. La fonction A est continue sur $[0,1]$ et, pour tout t dans $[0,1]$, les valeurs propres de $A(t)$ sont données par :

$$\lambda_1(t) = \begin{cases} 0 & \text{si } t = 0, \\ e^{-\frac{1}{t^2}} & \text{si } t \neq 0, \end{cases} \qquad \lambda_2(t) = -\lambda_1(t).$$

Ce sont des fonctions continues sur $[0,1]$. On peut leur associer les fonctions vecteurs propres suivantes :

$$\forall t \in]0,1], \quad e_1(t) = \begin{pmatrix} \cos\left(\frac{1}{2t}\right) \\ \sin\left(\frac{1}{2t}\right) \end{pmatrix}, \quad e_2(t) = \begin{pmatrix} \sin\left(\frac{1}{2t}\right) \\ -\cos\left(\frac{1}{2t}\right) \end{pmatrix},$$

$$e_1(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad e_2(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Il est clair que ces fonctions vecteurs propres ne sont pas continues en 0.

Exercice 4.25 : *Soient A, B deux matrices hermitiennes et $[a,b]$ un intervalle réel. Pour tout $t \in [a,b]$ on pose $A(t) = A + tB$. Montrer que la fonction qui à tout réel t dans $[a,b]$ associe la plus petite valeur propre de $A(t)$ est concave.*

Solution : Pour tout $t \in [a,b]$, on désigne par $\{e_1(t), \ldots, e_n(t)\}$ une base orthonormée de vecteurs propres associée aux valeurs propres (réelles) $\lambda_1(t), \ldots, \lambda_n(t)$ de $A(t)$ avec $A(t) e_k(t) = \lambda_k(t) e_k(t)$ pour tout $k = 1, 2, \ldots, n$.

Pour tout vecteur $x = \sum_{j=1}^{n} x_j e_j(t)$ de norme 1 et tout $t \in [a,b]$, on a :

$$\langle A(t) x \mid x \rangle = \sum_{j=1}^{n} \lambda_j(t) x_j^2 \geq \lambda_1(t) \sum_{j=1}^{n} x_j^2 = \lambda_1(t).$$

C'est-à-dire que :

$$\forall t \in [a,b], \quad \langle Ax \mid x \rangle + t \langle Bx \mid x \rangle \geq \lambda_1(t).$$

En prenant $x = e_1(t_0)$ pour $t_0 \in [a,b]$, on a :

$$\begin{cases} \lambda_1(t_0) = \langle Ae_1(t_0) \mid e_1(t_0) \rangle + t_0 \langle Be_1(t_0) \mid e_1(t_0) \rangle \\ \forall t \in [a,b], \quad \lambda_1(t) \leq \langle Ae_1(t_0) \mid e_1(t_0) \rangle + t \langle Be_1(t_0) \mid e_1(t_0) \rangle. \end{cases}$$

Ce qui peut s'écrire, en posant $b_0 = \langle Be_1(t_0) \mid e_1(t_0) \rangle$:

$$\forall t \in [a,b], \quad \lambda_1(t) \leq \lambda_1(t_0) + b_0(t - t_0).$$

Pour u, v dans $[a,b]$ et $\theta \in [0,1]$, on a alors :

$$\begin{cases} \lambda_1(u) \leq \lambda_1\big(u + \theta(v-u)\big) - \theta(v-u) b_{\theta,u,v} \\ \lambda_1(v) \leq \lambda_1\big(u + \theta(v-u)\big) + (1-\theta)(v-u) b_{\theta,u,v} \end{cases}$$

et

$$(1-\theta)\lambda_1(u) + \theta\lambda_1(v) \leq \lambda_1\big(u + \theta(v-u)\big).$$

C'est-à-dire que la fonction λ_1 est concave.

5 | Systèmes linéaires

Dans ce chapitre, on désigne toujours par $\mathbb{K}$ le corps des réels ou des complexes, n est un entier supérieur ou égal à 2 et $\{e_1, e_2, \ldots, e_n\}$ désigne la base canonique de $\mathbb{K}^n$.

Pour i, j entiers compris entre 1 et n, on note E_{ij} la matrice dont tous les coefficients sont nuls sauf celui d'indice (i, j) qui vaut 1. La famille $\{E_{i,j} \mid 1 \leq i, j \leq n\}$ est alors une base de $\mathcal{M}_n(\mathbb{K})$.

Une matrice est notée $A = \left((a_{ij})\right)_{1 \leq i,j \leq n}$, où l'indice i est le numéro de la ligne et j celui de la colonne.

On note I_n la matrice identité d'ordre n.

1. Position des problèmes et notation

On s'intéresse dans ce chapitre à la résolution de systèmes linéaires, $Ax = b$, de n équations à n inconnues à coefficients dans $\mathbb{K}$ et au problème de l'inversion d'une matrice de $GL_n(\mathbb{K})$.

On distingue deux catégories de méthodes de résolution d'un système linéaire, les méthodes directes et les méthodes itératives.

Si on suppose une précision infinie, les méthodes directes conduisent à la solution exacte du système en un nombre fini d'étapes alors que les méthodes itératives donnent toujours une approximation de la solution, la solution exacte étant obtenue en un nombre infini d'étapes.

Les méthodes directes ont un caractère exact, mais pour les grands systèmes, la propagation des erreurs d'arrondis en diminue l'efficacité.

Les méthodes itératives sont bien adaptées au cas des matrices creuses (avec beaucoup de termes nuls), car ces méthodes ne transforment pas la matrice de départ, contrairement aux méthodes directes.

Pour l'aspect programmation on pourra consulter [16] ou [5].

Exemple 5.1 : ***Une méthode directe, la méthode de Cramer***

Si on note, pour $j = 1, 2, \ldots, n$, A_j la matrice déduite de A en remplaçant par b sa colonne numéro j, alors la solution du système $Ax = b$ est donnée par les *formules de Cramer :*

$$x_j = \frac{\det(A_j)}{\det(A)} \quad (j = 1, 2, \ldots, n).$$

En calculant un déterminant par la formule :

$$\det(A) = \sum_{\sigma \in \mathcal{S}_n} sign(\sigma) \prod a_{j,\sigma(j)},$$

où $\mathcal{S}_n$ désigne l'ensemble des permutations de $\{1, 2, \ldots, n\}$ et $sign(\sigma)$ la signature de σ, cela nécessite $n!(n-1)$ multiplications et $(n! - 1)$ additions, donc environ $nn!$ opérations élémentaires. Comme il y a $n+1$ déterminants à calculer puis n divisions à faire, on aura donc un total d'environ $n^2 n!$ opérations élémentaires à effectuer, ce qui peut être beaucoup trop important pour de grandes valeurs de n.

Les méthodes directes sont plutôt adaptées aux « petits systèmes » ($n < 50$) et les méthodes itératives aux « grands systèmes ».

Les méthodes décrites dans ce chapitre sont utilisées pour la résolution des systèmes non linéaires par la méthode de Newton-Raphson et pour résoudre des problèmes d'interpolation et d'approximation.

Les méthodes itératives sont utilisées en relation avec les méthodes de différences finies et d'éléments finis pour la résolution de certaines équations aux dérivées partielles linéaires.

Remarque 5.1 *: Si la matrice A et le vecteur b sont à coefficients complexes, en écrivant $A = C + iD$, $b = \alpha + i\beta$ et $x = \xi + i\eta$, avec $C, D, \alpha, \beta, \xi, \eta$ à coefficients réels, la résolution du système de n équations à n inconnues à coefficients complexes $Ax = b$ se ramène à un système de $2n$ équations à $2n$ inconnues à coefficients réels :*

$$\begin{pmatrix} C & -D \\ D & C \end{pmatrix} \begin{pmatrix} \xi \\ \eta \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix}.$$

On peut donc se contenter de ne considérer que les systèmes à coefficients réels.

Remarque 5.2 *: Si la matrice A est inversible, les colonnes de A^{-1} sont les solutions des n systèmes linéaires :*

$$Ax = e_j \quad (j = 1, 2, \ldots, n).$$

En effet, en notant, pour tout $j = 1, 2, \ldots, n$, C_j la colonne numéro j de A^{-1}, on a :

$$AA^{-1} = (AC_1, AC_2, \ldots, AC_n) = (e_1, e_2, \ldots, e_n) = I_n.$$

C'est-à-dire que le calcul de l'inverse d'une matrice se ramène à la résolution simultanée de n systèmes linéaires de même matrice. Le coût du calcul de l'inverse d'une matrice d'ordre n est donc n fois supérieur à celui de la résolution d'un système linéaire de n équations à n inconnues. Pour cette raison il est déconseillé, d'un point de vue numérique, d'inverser une matrice si cette opération ne s'avère pas nécessaire.

2. Problèmes numériques liés à la résolution des systèmes linéaires

La précision des ordinateurs n'étant pas infinie, un algorithme de résolution numérique d'un système linéaire ne donnera pas, en général, une solution exacte mais seulement une approximation de cette solution.

D'autre part il sera important de savoir évaluer le temps de calcul nécessaire à un tel algorithme.

Les erreurs numériques sont de trois types :

- les erreurs sur les données (pour des données expérimentales) ;
- les erreurs d'arrondis (calculs en virgule flottante) ;
- les erreurs de troncature.

Les erreurs d'arrondis vont s'accumuler au cours des calculs. Il est donc important de connaître le nombre d'opérations élémentaires que nécessite un algorithme.

Un système $Ax = b$ est dit (théoriquement) dégénéré si $\det(A) = 0$. Mais vérifier cette condition de manière exacte nécessite une précision infinie. Il faut donc définir un concept de dégénérescence numérique d'un système linéaire.

Un système sera dit numériquement dégénéré, si la valeur de son déterminant calculée en machine est non significative (c'est-à-dire trop petite ou trop grande).

Exemple 5.2 : *La matrice de Hilbert d'ordre n est définie par :*

$$H_n = \left(\left(\frac{1}{i+j-1}\right)\right)_{1 \le i,j \le n}.$$

C'est une matrice symétrique définie positive, donc non dégénérée. En fait, on a :

$$\det(H_n) = \frac{\Phi_{n-1}^4}{\Phi_{2n-1}}$$

où $\Phi_n = \prod_{k=1}^{n} k!$ *(exercice 5.2).*

Pour n grand, on a $\det(H_n)_{machine} = 0$ *et le système est numériquement dégénéré, mais non dégénéré.*

Des exemples de calculs de déterminants de matrices de Hilbert, en utilisant, d'une part la décomposition LR et d'autre part la formule de Hilbert sont donnés dans le tableau 5.1, page suivante.

Dans la pratique, les coefficients de la matrice et du second membre ne sont connus que de façon approximative. Un problème important est donc de savoir si de petites variations sur ces coefficients peuvent entraîner de grosses variations sur la solution. La notion de conditionnement d'une matrice (paragraphe 6, page 147) nous fournit un outil qui permet d'analyser ce problème.

Les matrices de Hilbert sont mal conditionnées (exercice 4.21).

Tableau 5.1.

n	Déterminant formule	Déterminant calculé avec LR
1	1.000 000 000 000 00 E+00	1.000 000 000 000 00 E+00
2	8.333 333 333 333 33 E−02	8.333 333 333 333 33 E−02
3	4.629 629 629 629 63 E−04	4.629 629 629 629 61 E−04
4	1.653 439 153 439 15 E−07	1.653 439 153 439 30 E−07
5	3.749 295 132 515 09 E−12	3.749 295 132 512 90 E−12
6	5.367 299 887 358 69 E−18	5.367 299 886 820 88 E−18
7	4.835 802 623 926 12 E−25	4.835 802 615 662 27 E−25
8	2.737 050 113 791 51 E−33	2.737 050 232 581 96 E−33
9	9.720 234 311 925 00 E−43	9.720 263 356 547 86 E−43
10	2.164 179 226 431 49 E−53	2.164 369 970 644 80 E−53
11	3.019 095 334 449 35 E−65	3.024 497 147 717 83 E−65
12	2.637 780 651 253 55 E−78	2.708 748 828 405 69 E−78
13	1.442 896 518 791 14 E−92	Pivot trop petit dans LR
14	4.940 314 914 590 83 E−108	Pivot trop petit dans LR
15	1.058 542 743 069 72 E−124	Pivot trop petit dans LR
16	1.419 139 211 443 17 E−142	Pivot trop petit dans LR
17	1.190 271 242 703 00 E−161	Pivot trop petit dans LR
18	6.244 860 614 123 11 E−182	Pivot trop petit dans LR
19	2.049 343 733 318 12 E−203	Pivot trop petit dans LR
20	4.206 178 956 624 73 E−226	Pivot trop petit dans LR
21	5.398 986 241 900 04 E−250	Pivot trop petit dans LR

3. Cas des matrices triangulaires

On suppose pour ce paragraphe que la matrice A est inversible, triangulaire supérieure et à coefficients réels ou complexes.

Le système $Ax = b$ s'écrit alors :

$$a_{ii}x_i + a_{i,i+1}x_{i+1} + \cdots + a_{i,n}x_n = b_i \quad (1 \leq i \leq n),$$

avec $a_{ii} \neq 0$ pour tout $i = 1, 2, \ldots, n$.

La résolution de ce système se fait alors « en remontant » :

$$\begin{cases} x_n = \dfrac{b_n}{a_{nn}}, \\ x_i = \dfrac{\left(b_i - \sum\limits_{j=i+1}^{n} a_{ij}x_j\right)}{a_{ii}} \quad (i = n-1, \ldots, 1). \end{cases}$$

Pour un système triangulaire inférieur, on procède de même « en descendant ».

Pour le calcul de l'inverse d'une matrice triangulaire on utilise le fait que les colonnes de A^{-1} sont les solutions des systèmes $Ax = e_i$ avec i compris entre 1 et n. Ce qui donne, si A est triangulaire inférieure, pour les coefficients de $U = A^{-1}$:

$$\begin{cases} u_{ii} = \dfrac{1}{a_{ii}}, \\ u_{ij} = 0 \quad (1 \leq i < j \leq n), \\ u_{ki} = -\dfrac{\sum\limits_{j=i}^{k-1} a_{kj}u_{ji}}{a_{kk}} \quad (1 \leq i < k \leq n). \end{cases}$$

Si la matrice A est triangulaire supérieure la matrice ${}^tA$ est alors triangulaire inférieure et on peut écrire :

$$A^{-1} = {}^t({}^tA)^{-1}.$$

4. Matrices de dilatation et de transvection. Opérations élémentaires

Définition 5.1 : *On appelle matrice de transvection toute matrice de la forme :*

$$T_{ij}(\lambda) = I_n + \lambda E_{ij},$$

avec $1 \leq i \neq j \leq n$ et $\lambda \in \mathbb{K}$.

Définition 5.2 : *On appelle matrice de dilatation toute matrice de la forme :*

$$D_i(\lambda) = I_n + (\lambda - 1) E_{ii},$$

avec $1 \leq i \leq n$ et $\lambda \in \mathbb{K}^*$.

Définition 5.3 : *On appelle matrice élémentaire une matrice de dilatation ou de transvection.*

Lemme 5.1 : *Une matrice élémentaire est inversible avec :*

$$T_{ij}(\lambda)^{-1} = T_{ij}(-\lambda)$$

pour une matrice de transvection, et :

$$D_i(\lambda)^{-1} = D_i\left(\frac{1}{\lambda}\right)$$

pour une matrice de dilatation.

Démonstration : Pour $1 \leq i \neq j \leq n$ et $\lambda \in \mathbb{K}$ on a :

$$T_{ij}(\lambda) T_{ij}(-\lambda) = I_n - \lambda^2 E_{ij}^2,$$

avec $E_{ij}^2 = 0$. En effet pour tout entier k compris entre 1 et n on a :

$$E_{ij} e_k = \begin{cases} 0 & \text{si } k \neq j, \\ e_i & \text{si } k = j, \end{cases} \quad \Rightarrow \quad E_{ij}^2 e_k = \begin{cases} 0 & \text{si } k \neq j, \\ E_{ij} e_i = 0 & \text{si } k = j. \end{cases}$$

Le deuxième résultat est évident. ■

Pour toute matrice $A \in \mathcal{M}_n(\mathbb{K})$ on note L_i $(1 \leq i \leq n)$ sa ligne numéro i (c'est une matrice à 1 ligne et n colonnes) et C_j $(1 \leq j \leq n)$ sa colonne numéro j (c'est une matrice à n lignes et 1 colonne). On écrira :

$$A = \begin{pmatrix} L_1 \\ \vdots \\ L_n \end{pmatrix} \text{ ou } A = \begin{pmatrix} C_1 & \ldots & C_n \end{pmatrix}.$$

On appelle matrice déduite de A par opération élémentaire sur les lignes de A toute matrice de la forme :

$$A_i(\lambda) = \begin{pmatrix} L_1 \\ \vdots \\ L_{i-1} \\ \lambda L_i \\ L_{i+1} \\ \vdots \\ L_n \end{pmatrix},$$

avec $1 \leq i \leq n$ et $\lambda \in \mathbb{K}^*$, c'est-à-dire que la matrice $A_i(\lambda)$ est déduite de la matrice A en multipliant sa ligne numéro i par λ, ou de la forme :

$$A_{ij}(\lambda) = \begin{pmatrix} L_1 \\ \vdots \\ L_{i-1} \\ L_i + \lambda L_j \\ L_{i+1} \\ \vdots \\ L_n \end{pmatrix},$$

avec $1 \leq i \neq j \leq n$ et $\lambda \in \mathbb{K}$, c'est-à-dire que la matrice $A_{ij}(\lambda)$ est déduite de la matrice A en ajoutant à la ligne numéro i la ligne numéro j multipliée par λ.

On appelle matrice déduite de A par opération élémentaire sur les colonnes de A toute matrice de la forme :

$$A_i'(\lambda) = \begin{pmatrix} C_1 & \ldots & C_{j-1} & \lambda C_j & C_{j+1} & \ldots & C_n \end{pmatrix},$$

avec $1 \leq j \leq n$ et $\lambda \in \mathbb{K}^*$, c'est-à-dire que la matrice $A'_i(\lambda)$ est déduite de la matrice A en multipliant sa colonne numéro j par λ, ou de la forme :

$$A'_{ij}(\lambda) = \begin{pmatrix} C_1 & \dots & C_{j-1} & C_j + \lambda C_i & C_{j+1} & \dots & C_n \end{pmatrix},$$

avec $1 \leq i \neq j \leq n$ et $\lambda \in \mathbb{K}$, c'est-à-dire que la matrice $A'_{ij}(\lambda)$ est déduite de la matrice A en ajoutant à la colonne numéro j la colonne numéro i multipliée par λ.

Lemme 5.2 : *Avec les notations qui précèdent on a :*

$$A_i(\lambda) = D_i(\lambda)A, \quad A_{ij}(\lambda) = T_{ij}(\lambda)A,$$
$$A'_j(\lambda) = AD_j(\lambda), \quad A'_{ij}(\lambda) = AT_{ij}(\lambda).$$

Démonstration : Le coefficient d'indice (p, q) du produit de matrices $D_i(\lambda)A$ est obtenu en faisant le produit de la ligne p de $D_i(\lambda)$ par la colonne q de A, ce qui donne, en notant $\alpha_{p,q}$ ce coefficient :

$$\alpha_{p,q} = \begin{cases} a_{p,q} & \text{si } 1 \leq p \neq i \leq n,\ 1 \leq q \leq n, \\ \lambda a_{iq} & \text{si } p = i,\ 1 \leq q \leq n. \end{cases}$$

On a donc bien $A_i(\lambda) = D_i(\lambda)A$.

Les autres égalités se montrent de façon analogue. ■

Remarque 5.3 *: On sera amené à utiliser deux autres types d'opérations élémentaires que sont les permutations de lignes ou de colonnes. En fait ces opérations se déduisent des précédentes. Par exemple, pour permuter les lignes i et j où $1 \leq i < j \leq n$ on effectue les opérations suivantes :*

$$\begin{pmatrix} L_1 \\ \vdots \\ L_n \end{pmatrix} \mapsto \begin{pmatrix} L_1 \\ \vdots \\ L_{i-1} \\ L_i + L_j \\ L_{i+1} \\ \vdots \\ L_n \end{pmatrix} \mapsto \begin{pmatrix} L_1 \\ \vdots \\ L_i + L_j \\ \vdots \\ -L_i \\ \vdots \\ L_n \end{pmatrix} \mapsto \begin{pmatrix} L_1 \\ \vdots \\ L_j \\ \vdots \\ -L_i \\ \vdots \\ L_n \end{pmatrix} \mapsto \begin{pmatrix} L_1 \\ \vdots \\ L_j \\ \vdots \\ L_i \\ \vdots \\ L_n \end{pmatrix},$$

ce qui revient à effectuer les produits :

$$D_j(-1)T_{ij}(1)T_{ji}(-1)T_{ij}(1)A.$$

La matrice $D_j(-1)T_{ij}(1)T_{ji}(-1)T_{ij}(1)$ est la matrice de permutation qui a pour action de permuter les vecteurs e_i et e_j. Elle s'écrit plus simplement :

$$P_{ij} = I_n - (E_{ii} + E_{jj}) + (E_{ij} + E_{ji}).$$

De même la permutation des colonnes i et j est obtenue avec :

$$AT_{ij}(1)T_{ji}(-1)T_{ij}(1)D_i(-1).$$

Le produit de matrices $T_{ij}(1)T_{ji}(-1)T_{ij}(1)D_i(-1)$ est la matrice de permutation ${}^tP_{ij} = P_{ji}$.

L'ensemble des matrices de dilatation ou de transvection forme un système générateur du groupe multiplicatif $GL_n(\mathbb{K})$.

Théorème 5.1 : *Pour toute matrice $A \in GL_n(\mathbb{K})$ il existe des matrices de transvection $P_1, \ldots, P_r$ et $Q_1, \ldots, Q_s$ telles que :*

$$A = P_1 \ldots P_r D_n\big(\det(A)\big) Q_1 \ldots Q_s.$$

Démonstration : On procède par récurrence sur $n > 0$. Pour $n = 1$ le résultat est évident. On le suppose vrai sur $GL_{n-1}(\mathbb{K})$.

On se ramène tout d'abord au cas où $a_{21} \neq 0$. Si $a_{21} = 0$, la matrice A étant inversible, il existe alors un indice $i \in \{1, 3, \ldots, n\}$ tel que $a_{i1} \neq 0$ et la matrice $T_{2i}(1)\, A$ (déduite de A en ajoutant la ligne i à la ligne 2) est telle que son coefficient d'indice $(2, 1)$ est non nul.

Une fois ramené à $a_{21} \neq 0$, on se ramène à $a_{11} = 1$ en remplaçant la première ligne L_1 par $L_1 + \lambda L_2$ (multiplication à gauche par $T_{12}(\lambda)$) où le scalaire λ est choisi tel que $a_{11} + \lambda a_{21} = 1$.

Ensuite, pour tout $i \in \{2, 3, \ldots, n\}$, en remplaçant la ligne L_i par $L_i - a_{i1} L_1$ (multiplication à gauche par $T_{i1}(-a_{i1})$), on annule le coefficient d'indice $(i, 1)$.

On peut donc trouver des matrices de transvection $P_1, \ldots, P_k$ telles que :

$$P_k \ldots P_1 A = \begin{pmatrix} 1 & \alpha_{12} & \ldots & \alpha_{1n} \\ 0 & \alpha_{22} & \ldots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \alpha_{n2} & \ldots & \alpha_{nn} \end{pmatrix}.$$

De manière analogue, en multipliant à droite par des matrices de transvection, $Q_1, \ldots, Q_m$, on obtient :

$$P_k \ldots P_1 A Q_1 \ldots Q_m = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \beta_{22} & \ldots & \beta_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \beta_{n2} & \ldots & \beta_{nn} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix},$$

avec $\det(B) = \det(A)$, puisque $\det(P) = 1$ pour toute matrice de transvection P. On peut alors conclure en appliquant l'hypothèse de récurrence à la matrice $B \in GL_{n-1}(\mathbb{K})$. ∎

En désignant par $GL_n^+(\mathbb{R})$ [resp. $GL_n^-(\mathbb{R})$] l'ensemble des matrices d'ordre n à coefficients réels de déterminant strictement positif [resp. négatif] on déduit du théorème précédent le résultat suivant.

Corollaire 5.1 : *Les ensembles $GL_n^+(\mathbb{R})$ et $GL_n^-(\mathbb{R})$ sont connexes par arcs et ce sont les composantes connexes de $GL_n(\mathbb{R})$.*

Démonstration : Soit $A \in GL_n^+(\mathbb{R})$. Elle s'écrit :

$$A = P_1 \dots P_r D_n(\det(A)) Q_1 \dots Q_s,$$

P_k et Q_j étant des matrices de transvection.

Pour toute matrice de transvection $T = T_{ij}(\lambda)$ et tout $t \in [0,1]$, on note $T(t) = T_{ij}(t\lambda)$ et on définit l'application $\gamma : [0,1] \to \mathcal{M}_n(\mathbb{R})$ par :

$$\forall t \in [0,1], \quad \gamma(t) = \prod_{k=1}^{r} P_k(t) \Delta(t) \prod_{k=1}^{s} Q_k(t),$$

où

$$\Delta(t) = D_n(t \det(A) + (1-t)).$$

Pour tout $t \in [0,1]$, on a $\det(\gamma(t)) = t \det(A) + (1-t) > 0$ du fait que $A \in GL_n^+(\mathbb{R})$. De plus γ est continue avec $\gamma(0) = I_n$ et $\gamma(1) = A$. On a donc ainsi prouvé que $GL_n^+(\mathbb{R})$ est connexe par arcs.

Pour toute matrice $A \in GL_n^-(\mathbb{R})$ (par exemple $A = D_n(-1)$) l'application $M \longmapsto AM$ réalise un homéomorphisme de $GL_n^+(\mathbb{R})$ sur $GL_n^-(\mathbb{R})$. On en déduit alors que $GL_n^-(\mathbb{R})$ est connexe par arcs.

Avec $GL_n(\mathbb{R}) = GL_n^+(\mathbb{R}) \cup GL_n^-(\mathbb{R})$, les ensembles $GL_n^+(\mathbb{R})$ et $GL_n^-(\mathbb{R})$ étant des ouverts connexes disjoints, on déduit que ce sont les composantes connexes de $GL_n(\mathbb{R})$. ∎

Ce résultat permet de définir une orientation sur $\mathbb{R}^n$. On dit que deux bases $\mathcal{B}$ et $\mathcal{B}'$ de $\mathbb{R}^n$ définissent la même orientation si la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$ est dans $GL_n^+(\mathbb{R})$.

5. Méthode des pivots de Gauss

En effectuant des opérations élémentaires sur les lignes du système linéaire $Ax = b$ (permutations et combinaisons linéaires de lignes, c'est-à-dire des opérations qui ne vont pas changer l'ensemble de ses solutions), on le transforme en un système triangulaire supérieur $Rx = c$. Du fait qu'une permutation de lignes change le déterminant de signe et que d'ajouter un multiple d'une ligne à une autre ne change pas ce dernier, on a :

$$\det(A) = \det(R) = \pm \prod_{i=1}^{n} r_{ii}.$$

L'intérêt de cette méthode est surtout pédagogique. Dans la pratique il est préférable d'utiliser la méthode LR (paragraphe 7, page 192) quand cette dernière peut s'appliquer, ou la méthode de Gauss-Jordan (paragraphe 10, page 198).

On note ici L_i la ligne numéro i du système linéaire $Ax = b$.

Étape 0 — On se ramène à un système tel que a_{11} soit non nul.

Si pour tout $i = 1, 2, \dots, n$, on a $a_{i1} = 0$, alors $\det(A) = 0$ et c'est fini. Sinon, il existe $i > 1$ tel que a_{i1} soit non nul, et en permutant les lignes 1 et i (si $i = 1$, on ne fait rien) on se ramène à un système $A^{(1)}x = b^{(1)}$, avec $a_{11}^{(1)}$ non nul.

Le coefficient $a_{11}^{(1)}$ est le premier pivot.

On a alors $\det(A) = \pm \det\left(A^{(1)}\right)$, avec le signe moins si et seulement si il y a une permutation des lignes 1 et i avec $i > 1$.

Étape 1 — Élimination de x_1 dans les équations $2, \ldots, n$.

On effectue pour cela les transformations élémentaires suivantes :

$$L_i^{(1)} \mapsto L_i^{(1)} - \frac{a_{i1}^{(1)}}{a_{11}^{(1)}} L_1^{(1)} \quad (i = 2, 3, \ldots, n),$$

et, après une éventuelle permutation des lignes 2 et $i > 3$, on obtient le système $A^{(2)}x = b^{(2)}$, avec $a_{22}^{(2)}$ non nul où :

$$A^{(2)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \ldots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ldots & a_{2n}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(2)} & \ldots & a_{nn}^{(2)} \end{pmatrix}.$$

Le coefficient $a_{22}^{(2)}$ est le deuxième pivot.

Étape k — Élimination de x_k dans les équations $k+1, \ldots, n$.

À la fin de l'étape $k-1$, on a obtenu le système $A^{(k)}x = b^{(k)}$, avec :

$$A^{(k)} = \begin{pmatrix} a_{11}^{(1)} & a_{12}^{(1)} & \ldots & a_{1k}^{(1)} & \ldots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & \ldots & a_{2k}^{(2)} & \ldots & a_{2n}^{(2)} \\ \vdots & \ddots & \ddots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & a_{kk}^{(k)} & \ldots & a_{kn}^{(k)} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & a_{nk}^{(k)} & \ldots & a_{nn}^{(k)} \end{pmatrix}$$

et $a_{kk}^{(k)}$ non nul.

Le coefficient $a_{kk}^{(k)}$ est le pivot numéro k.

On effectue alors les transformations élémentaires suivantes :

$$L_i^{(k)} \mapsto L_i^{(k)} - \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} L_k^{(k)} \, (i = k+1, \ldots, n),$$

puis une éventuelle permutation des lignes $k+1$ et $j > k+1$ pour se ramener à $a_{k+1,k+1}^{(k+1)}$ non nul.

Au bout de $n-1$ étapes, on est donc ramené à un système triangulaire supérieur $A^{(n)}x = b^{(n)}$.

De plus, on a $\det(A) = (-1)^p \det\left(A^{(n)}\right)$, où p est le nombre de permutations qui ont été nécessaires pour avoir des pivots non nuls et $\det\left(A^{(n)}\right)$ est le produit des pivots.

***Remarque 5.4** : Pour éviter de faire une division par un nombre trop petit, dans le choix du pivot, on aura intérêt, à l'étape $k-1$, à permuter la ligne k avec la ligne $j \geq k$ telle que :*

$$|a_{jk}| = \max\{|a_{ik}| \mid i = k, \ldots, n\},$$

de manière à avoir le pivot le plus grand possible en valeur absolue.

Si ce maximum est trop petit, alors le système est numériquement dégénéré.

On peut donner un ordre de grandeur du nombre d'opérations élémentaires que nécessite la méthode de Gauss.

Les opérations $L_i^{(k)} \mapsto L_i^{(k)} - \dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}} L_k^{(k)}$ $(i = k+1, \ldots, n)$, à l'étape numéro k, demandent $(n-k)$ divisions, $(n-k)^2$ multiplications et $(n-k)^2$ additions.

Ce qui donne un total de $\sum\limits_{k=1}^{n-1} n-k = \dfrac{n(n-1)}{2}$ divisions et un nombre d'additions et de multiplications égal à $2\sum\limits_{k=1}^{n-1}(n-k)^2 = \dfrac{2n(n-1)(2n-1)}{6}$.

Il nous faut donc $\dfrac{n(n-1)(4n+1)}{6} \backsim \dfrac{2n^3}{3}$ opérations élémentaires pour aboutir à un système triangulaire supérieur. Puis la résolution de ce dernier système nécessite n divisions et $n(n-1)$ additions et multiplications, soit n^2 opérations élémentaires.

En résumé, la résolution d'un système de n équations à n inconnues par la méthode de Gauss va nécessiter un nombre d'opérations qui est un $O(n^3)$.

6. Résolution des systèmes linéaires à coefficients entiers

Si la matrice A et le vecteur b sont à coefficients entiers relatifs, en modifiant la méthode de Gauss, on peut calculer la solution x de façon exacte comme vecteur à coefficients dans $\mathbb{Q}$.

On procède comme suit.

Étape 0 — Après une éventuelle permutation de lignes, on a obtenu le système $A^{(1)}x = b^{(1)}$ à coefficients dans $\mathbb{Z}$ avec un premier pivot non nul.

Étape 1 — Élimination de x_1 dans les équations $2, \ldots, n$.

En effectuant la division de la ligne 1 par le premier pivot, on perd le caractère entier du système. Il est préférable de procéder comme suit :

- garder la ligne 1 ;
- $L_i^{(1)} \mapsto a_{11}^{(1)} L_i^{(1)} - a_{i1}^{(1)} L_1^{(1)}$, pour $i = 2, \ldots, n$.

Ce qui donne le système $A^{(2)}x = b^{(2)}$, avec un deuxième pivot non nul (au prix d'une éventuelle permutation).

Étape 2 — Élimination de x_2 dans les équations $3, \ldots, n$.

Si on procède de la même façon pour cette étape, on constate que les coefficients de la matrice $A^{(3)}$ et du vecteur $b^{(3)}$ sont divisibles par le premier pivot. On peut donc s'autoriser une division par ce premier pivot pour cette étape 2, ce qui aura l'avantage de diminuer les coefficients obtenus.

Étape k — Élimination de x_k dans les équations $k+1, \ldots, n$.

Cette étape est décrite par les formules :

$$a_{ij}^{(k+1)} = \frac{\left(a_{kk}^{(k)} a_{ij}^{(k)} - a_{ik}^{(k)} a_{kj}^{(k)}\right)}{a_{k-1,k-1}^{(k-1)}},$$

pour $i = k+1, \ldots, n$ et $j = k+1, \ldots, n+1$ (en prenant b comme colonne numéro $n+1$).

Ce qui donne au bout de $n-1$ étapes un système triangulaire supérieur à coefficients entiers qui peut se résoudre de façon exacte dans $\mathbb{Q}^n$.

Calcul du déterminant — Au signe près, on a :

$$\det\left(A^{(2)}\right) = \left(a_{11}^{(1)}\right)^{n-1} \det(A),$$

$$\det\left(A^{(3)}\right) = \left(\frac{a_{22}^{(2)}}{a_{11}^{(1)}}\right)^{n-2} \det\left(A^{(2)}\right) = \left(a_{22}^{(2)}\right)^{n-2} a_{11}^{(1)} \det\left(A^{(1)}\right).$$

De proche en proche, on a alors :

$$\det\left(A^{(n)}\right) = (-1)^p \left(\prod_{k=1}^{n} a_{kk}^{(k)}\right) \det(A),$$

ce qui donne, en tenant compte de l'expression de $\det\left(A^{(n)}\right)$ comme produit des pivots :

$$\det(A) = (-1)^p\, a_{nn}^{(n)}.$$

7. Décomposition LR (méthode de Crout)

Définition 5.4 : *On appelle sous-matrices principales d'une matrice $A \in \mathcal{M}_n(\mathbb{K})$ les matrices :*

$$A_k = \left((a_{ij})\right)_{1 \le i,j \le k}, \quad (k = 1, \ldots, n).$$

Les déterminants principaux sont les $\Delta_k = \det(A_k)$.

Si tous les déterminants principaux de A sont non nuls alors, dans la méthode de Gauss, tous les pivots seront non nuls et il ne sera pas nécessaire de faire des permutations de lignes (on ne s'occupe pas de la taille des pivots). Réciproquement, si tous les pivots sont non nuls, alors tous les déterminants principaux sont non nuls.

Exemple 5.3 : *Si A est à diagonale strictement dominante alors toutes les sous-matrices principales de A sont aussi à diagonale strictement dominante et donc inversibles. On en déduit donc que pour A à diagonale strictement dominante, la méthode de Gauss ne nécessite pas de permutations de lignes.*

Exemple 5.4 : *Si A est symétrique définie positive, il en est de même de toutes les sous-matrices principales de A. Donc, pour A symétrique définie positive, la méthode de Gauss ne nécessite pas de permutations de lignes.*

La méthode des pivots de Gauss est basée sur les résultats suivants.

Lemme 5.3 : *Soit $A \in \mathcal{M}_n(\mathbb{K})$ de coefficient a_{11} non nul. Il existe des matrices de transvection $P_1, \ldots, P_r$ telles que :*

$$P_r \ldots P_1 A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22}^{(1)} & \ldots & a_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \ldots & a_{nn}^{(1)} \end{pmatrix}.$$

Démonstration : En reprenant les notations du paragraphe 5, page 189, le passage de la matrice A, de premier terme a_{11} non nul, à la matrice unité $A^{(2)}$ se fait en multipliant à gauche la matrice A par les matrices de transvection :

$$P_i = T_{i1}(\lambda_{i1}) \quad (2 \leq i \leq n)$$

où on a noté :

$$\lambda_{i1} = -\frac{a_{i1}^{(1)}}{a_{11}}.$$

C'est-à-dire que :

$$P_{n-1} \ldots P_1 A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ 0 & a_{22}^{(1)} & \ldots & a_{2n}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & a_{n2}^{(1)} & \ldots & a_{nn}^{(1)} \end{pmatrix}.$$ ■

Remarque 5.5 *: On a $\operatorname{rg}(A) = 1 + \operatorname{rg}\left(B^{(1)}\right)$, en notant :*

$$B^{(1)} = \begin{pmatrix} a_{22}^{(1)} & \ldots & a_{2n}^{(1)} \\ \vdots & \ddots & \vdots \\ a_{n2}^{(1)} & \ldots & a_{nn}^{(1)} \end{pmatrix}.$$

On a $P_{n-1} \ldots P_1 = F_1$ avec :

$$F_1 = \begin{pmatrix} 1 & 0 & \ldots & \ldots & 0 \\ \lambda_{21} & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \lambda_{n-1,1} & 0 & \ldots & 1 & 0 \\ \lambda_{n1} & 0 & \ldots & 0 & 1 \end{pmatrix}.$$

La matrice F_1 est une matrice de Frobenius.

Définition 5.5 : *On appelle matrice de Frobenius une matrice carrée d'ordre n qui ne diffère de l'identité que par une colonne (ou une ligne).*

Théorème 5.2 : *Toute matrice $A \in GL_n(\mathbb{K})$ peut être réduite à la forme triangulaire supérieure en la multipliant à gauche par des matrices de transvection ou de dilatation de la forme $D_i(-1)$.*

Démonstration : C'est la transcription matricielle de l'algorithme de Gauss.

On peut aussi raisonner par récurrence sur $n \geq 1$ en utilisant le lemme précédent.

En multipliant au besoin la matrice A par une matrice de permutation (produit d'une matrice de dilatation par trois matrices de transvection) on se ramène au cas où le coefficient a_{11} est non nul. On conclut alors facilement en utilisant le lemme 5.3 et l'hypothèse de récurrence. ∎

Dans l'hypothèse où tous les déterminants principaux de la matrice A sont non nuls (on n'effectue pas de permutations de lignes dans la méthode de Gauss), le résultat précédent s'exprime en disant qu'il existe des matrices de Frobenius de la forme :

$$F_k = \prod_{i=k+1}^{n} T_{ik}(\lambda_{ik})$$

($\lambda_{ik} = -\dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$ avec les notations du paragraphe 5, page 189) telles que :

$$F_{n-1} \dots F_1 A = R,$$

où R est une matrice triangulaire supérieure.

La matrice produit $F_{n-1} \dots F_1$ est triangulaire inférieure à diagonale unité. En notant $L = (F_{n-1} \dots F_1)^{-1}$, on a alors la décomposition $A = LR$ avec L triangulaire inférieure à diagonale unité et R triangulaire supérieure.

Théorème 5.3 : *Une matrice inversible A possède une décomposition $A = LR$, avec L triangulaire inférieure à diagonale unité et R triangulaire supérieure, si et seulement si tous les déterminants principaux de A sont non nuls. Une telle décomposition est unique et les coefficients diagonaux de R sont donnés par :*

$$\begin{cases} r_{11} = a_{11}, \\ r_{kk} = \dfrac{\det(A_k)}{\det(A_{k-1})} \quad (k = 2, \dots, n), \end{cases}$$

où les A_k désignent les sous-matrices principales de A.

Démonstration : Si tous les déterminants principaux de A sont non nuls on a vu que la matrice A admet une décomposition LR.

Pour $k = 1, \dots, n$, on peut décomposer les matrices A, L et R par blocs de la façon suivante :

$$A = \begin{pmatrix} A_k & B_k \\ C_k & D_k \end{pmatrix}, \quad L = \begin{pmatrix} L_k & 0 \\ E_k & G_k \end{pmatrix}, \quad R = \begin{pmatrix} R_k & H_k \\ 0 & M_k \end{pmatrix},$$

où A_k, L_k et R_k désignent respectivement les sous-matrices principales de A, L et R. Le produit $A = LR$ donne alors $A_k = L_k R_k$ et :

$$\det(A_k) = \det(L_k)\det(R_k) = \det(R_k) = \prod_{i=1}^{k} r_{ii}.$$

L'unicité de la décomposition LR provient du fait que si $A = LR$ et $A = L'R'$, avec L, L' triangulaires inférieures de diagonale unité et R, R' triangulaires supérieures, alors $L'^{-1}L = R'R^{-1}$ est à la fois triangulaire inférieure et triangulaire supérieure de diagonale unité et c'est alors nécessairement l'identité.

Inversement, si la matrice inversible A admet une décomposition LR, alors R est aussi inversible et la décomposition par blocs faite précédemment nous montre que tous les déterminants principaux de A sont non nuls. ■

Pour obtenir pratiquement la décomposition LR on peut procéder par coefficients indéterminés, c'est-à-dire qu'on écrit $A = LR$, qu'on effectue le produit et qu'on identifie, ce qui donne, en effectuant les calculs dans l'ordre indiqué :

$$\begin{cases} r_{1j} = a_{1j} & (j = 1, \ldots, n), \\ L_{i1} = \dfrac{a_{i1}}{a_{11}} & (i = 2, \ldots, n), \end{cases}$$

puis, pour $k = 2, \ldots n$:

$$r_{kj} = a_{kj} - \sum_{i=1}^{k-1} L_{ki} r_{ij} \quad (j = k, \ldots, n)$$

et

$$L_{ik} = \frac{\left(a_{ik} - \sum_{j=1}^{k-1} L_{ij} r_{jk}\right)}{r_{kk}} \quad (i = k+1, \ldots n),$$

avec $L_{ii} = 1$.

En utilisant la décomposition $A = LR$, le système $Ax = b$ équivaut aux deux systèmes triangulaires :

$$\begin{cases} Ly = b & \text{(triangulaire inférieur)}, \\ Rx = y & \text{(triangulaire supérieur)}. \end{cases}$$

La méthode obtenue est appelée méthode de Crout. Cette méthode est intéressante pour résoudre en parallèle plusieurs systèmes linéaires de même matrice A, car la décomposition LR ne fait pas intervenir le second membre.

En utilisant la décomposition $A = LR$, on a $A^{-1} = R^{-1}L^{-1}$ et il suffit alors d'utiliser les procédures d'inversion des matrices triangulaires du paragraphe 3, page 184.

8. Décomposition $LD\,{}^tL$ des matrices symétriques réelles

On suppose dans ce paragraphe que la matrice A est inversible symétrique et qu'il n'y a pas de permutations dans la méthode de Gauss.

Dans la décomposition LR de A, on a $\det(R) = \det(A) \neq 0$, donc tous les termes diagonaux de R sont non nuls et on peut écrire R sous la forme $R = DR'$, où D est diagonale et R' est triangulaire supérieure à diagonale unité (il suffit de diviser chaque ligne de R par son terme diagonal). On a donc $A = LDR'$, puis, en écrivant que ${}^tA = A$ et en utilisant le fait que la décomposition LR est unique, on déduit que $R' = {}^tL$.

On a donc pour toute matrice symétrique A, dont tous les déterminants principaux sont non nuls, la décomposition unique $A = LD\,{}^tL$, la matrice L étant triangulaire inférieure et à diagonale unité et la matrice D diagonale.

Remarque 5.6 : *Cette décomposition nous donne un moyen de calculer la signature de la matrice symétrique réelle A.*

On rappelle qu'une matrice symétrique réelle a toutes ses valeurs propres réelles et que, pour A inversible, le couple d'entiers (p, q) formé du nombre p de valeurs propres strictement positives et du nombre $q = n - p$ de valeurs propres strictement négatives est uniquement déterminé par A. Ce couple d'entiers est la signature de A.

L'égalité $A = LD\,{}^tL$ avec L inversible se traduit en disant que les matrices A et D sont congruentes et le théorème de Sylvester nous dit que deux matrices congruentes ont même signature. On a donc :

$$\operatorname{sign}(A) = \operatorname{sign}(D) = (p, n-p),$$

où p est le nombre de termes strictement positifs de la diagonale de D.

Remarque 5.7 : *La matrice A est définie positive si et seulement si tous les coefficients de D sont strictement positifs.*

Comme pour la décomposition LR, on trouve les coefficients de L et D par identification, ce qui donne $d_1 = a_{11}$ et, pour $i = 2, \ldots, n$:

$$\begin{cases} L_{ij} = \dfrac{\left(a_{ij} - \sum\limits_{k=1}^{j-1} L_{ik} d_k L_{jk}\right)}{d_j} \quad (j = 1, \ldots, i-1), \\ d_i = \left(a_{ii} - \sum\limits_{k=1}^{i-1} L_{ik}^2 d_k\right). \end{cases}$$

9. Décomposition de Cholesky des matrices symétriques réelles définies positives

L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique.

Dans le cas où la matrice A est symétrique définie positive, la décomposition $LD\,{}^tL$ permet de montrer le résultat suivant.

Théorème 5.4 : *Une matrice réelle A est symétrique définie positive si et seulement si il existe une matrice B triangulaire inférieure et inversible telle que $A = B\,{}^tB$. De plus une telle décomposition est unique si on impose la positivité des coefficients diagonaux de la matrice B.*

Démonstration : Si A est symétrique définie positive, dans la décomposition $LD\,{}^tL$, on a alors $d_i > 0$ pour tout $i = 1, \ldots, n$. On peut donc écrire $D = D'^2$ et, en posant $B = LD'$, on a $A = B\,{}^tB$ avec B triangulaire inférieure, la diagonale de B étant formée des $\pm\sqrt{d_i}$.

Inversement, si $A = B\,{}^tB$, alors A est symétrique et pour tout vecteur x non nul, on a :

$$\langle Ax \mid x\rangle = \left\langle B\,{}^tBx \mid x\right\rangle = \left\langle {}^tBx \mid {}^tBx\right\rangle > 0,$$

puisque B est inversible. ∎

Le calcul effectif des coefficients de B se fait par identification, ce qui donne :

$$\begin{cases} b_{ij} = \dfrac{\left(a_{ij} - \sum\limits_{k=1}^{j-1} b_{ik}b_{jk}\right)}{b_{jj}} \quad (j = 1, \ldots, i-1)\,, \\ b_{ii}^2 = a_{ii} - \sum\limits_{k=1}^{i-1} b_{ik}^2, \end{cases} \qquad (i = 1, \ldots, n)\,.$$

Le nombre d'opérations dans la décomposition de Cholesky est un $O\left(n^3\right)$. En effet, il y a n racines carrées à calculer,

$$\sum_{i=1}^{n} i - 1 = \frac{n\,(n-1)}{2} \text{ divisions,}$$

$$\begin{aligned} \sum_{i=1}^{n}\left(i - 1 + \sum_{j=1}^{i-1} j - 1\right) &= \sum_{i=1}^{n}\left(i - 1 + \frac{(i-1)\,(i-2)}{2}\right) \\ &= \frac{1}{2}\sum_{i=1}^{n} i\,(i-1) \text{ additions,} \end{aligned}$$

et autant de multiplications. Ce qui donne un total de $\dfrac{n\left(2n^2 + 3n + 1\right)}{6}$ opérations.

Le déterminant de A se calcule avec :

$$\det(A) = \prod_{i=1}^{b} b_{ii}^2.$$

La résolution du système $Ax = e$ se ramène à la résolution de deux systèmes triangulaires.

Pour calculer l'inverse de A il suffit d'inverser la matrice triangulaire B.

10. Méthode d'élimination de Gauss-Jordan

La méthode de Gauss-Jordan est une variante de la méthode de Gauss. Elle consiste à se ramener à un système diagonal à l'aide de transformations élémentaires.

À l'étape k, on ne se contente pas seulement d'éliminer le coefficient de x_k dans les lignes $k+1$ à n, mais on l'élimine aussi dans les lignes du dessus soit 1 à $k-1$. Ce qui donnera au bout de n étapes un système diagonal.

Description de la k-ième étape du calcul $(1 \leq k \leq n)$ — À la k-ième étape, on est au départ dans la situation suivante :

$$\begin{cases} x_1 + \quad + a_{1k}^{(k)} x_k + \cdots + a_{1n}^{(k)} x_n = b_1^{(k)} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_{k-1} + a_{k-1,k}^{(k)} x_k + \cdots + a_{k-1,n}^{(k)} x_n = b_{k-1}^{(k)} \\ \quad + a_{kk}^{(k)} x_k + \cdots + a_{kn}^{(k)} x_n = b_k^{(k)} \\ \quad + a_{k+1,k}^{(k)} x_k + \cdots + a_{k+1,n}^{(k)} x_n = b_{k+1}^{(k)} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \quad + a_{nk}^{(k)} x_k + \cdots + a_{nn}^{(k)} x_n = b_n^{(k)} \end{cases}$$

On peut supposer que $a_{kk}^{(k)}$ est le pivot maximal, sinon, il suffit de permuter avec une des lignes suivantes.

On commence par extraire x_k de la k-ième équation :

$$x_k = \frac{\left(b_k^{(k)} - a_{k,k+1}^{(k)} x_{k+1} - \cdots - a_{kn}^{(k)} x_n\right)}{a_{k,k}^{(k)}},$$

que l'on reporte ensuite dans les autres équations, ce qui donne :

$$\begin{cases} x_1 + \quad + a_{1k+1}^{(k+1)} x_{k+1} + \cdots + a_{1n}^{(k+1)} x_n = b_1^{(k+1)} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ x_{k-1} + a_{k-1,k+1}^{(k+1)} x_{k+1} + \cdots + a_{k-1,n}^{(k+1)} x_n = b_{k-1}^{(k+1)} \\ x_k \quad + a_{k,k+1}^{(k+1)} x_{k+1} + \cdots + a_{kn}^{(k+1)} x_n = b_k^{(k+1)} \\ \quad + a_{k+1,k+1}^{(k+1)} x_{k+1} + \cdots + a_{k+1,n}^{(k+1)} x_n = b_{k+1}^{(k+1)} \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ \quad + a_{nk+1}^{(k+1)} x_{k+1} + \cdots + a_{nn}^{(k+1)} x_n = b_n^{(k+1)} \end{cases}$$

avec, pour $j = k+1, \dots, n$:

$$a_{kj}^{(k+1)} = \frac{a_{kj}^{(k)}}{a_{kk}^{(k)}}, \quad b_k^{(k+1)} = \frac{b_k^{(k)}}{a_{kk}^{(k)}},$$

et, pour $i \in \{1, \ldots, n\} - \{k\}$, $j \in \{k+1, \ldots, n\}$:

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - a_{ik}^{(k)} a_{kj}^{(k+1)}, \quad b_i^{(k+1)} = b_i^{(k)} - a_{ik}^{(k)} b_k^{(k+1)}.$$

On obtient alors directement la solution, après la n-ième étape :

$$\begin{cases} x_1 = b_1^{(n)} \\ \vdots \\ x_n = b_n^{(n)}. \end{cases}$$

11. Méthodes itératives de résolution des systèmes linéaires

Étant donné un système linéaire $Ax = b$ de n équations à n inconnues, la matrice A étant supposée inversible, on veut construire une suite $\left(x^{(k)}\right)_{k \in \mathbb{N}}$ d'éléments de $\mathbb{R}^n$ qui va converger vers la solution x de ce système, le calcul de chaque $x^{(k)}$ étant plus simple que la résolution directe du système.

Pour ce faire, on écrit la matrice A sous la forme $A = M - N$, où M est facilement inversible, et la résolution de $Ax = b$ est ramenée au problème de point fixe qui consiste à trouver x dans $\mathbb{R}^n$ solution de $x = M^{-1}Nx + M^{-1}b$.

Pour résoudre ce problème, on utilise la méthode des approximations successives, c'est-à-dire qu'on considère la suite $\left(x^{(k)}\right)_{k \in \mathbb{N}}$ définie par :

$$\begin{cases} x^{(0)} \in \mathbb{R}^n, \\ x^{(k+1)} = M^{-1}Nx^{(k)} + M^{-1}b. \end{cases} \tag{5.1}$$

Si cette suite converge, c'est nécessairement vers la solution de $Ax = b$.

Une méthode itérative de résolution du système linéaire $Ax = b$ est donc définie par le choix d'une décomposition $A = M - N$.

Définition 5.6 : *Si A est une matrice réelle inversible, on dit que la méthode itérative associée à la décomposition $A = M - N$, avec M inversible, est convergente si pour tout $x^{(0)} \in \mathbb{R}^n$, la suite $\left(x^{(k)}\right)_{k \in \mathbb{N}}$ définie par (5.1) est convergente.*

Les principaux problèmes posés par les méthodes itératives sont les suivants :

- À quelles conditions portant sur la matrice M, la suite définie par (5.1) converge-t-elle ?
- Comment choisir M facilement inversible ?
- Quelle est la rapidité de la convergence ?

Le théorème qui suit nous donne une condition nécessaire et suffisante de convergence de la suite définie par (5.1).

Théorème 5.5 : *Soient $A = M - N$ une matrice inversible avec M inversible et $\rho\left(M^{-1}N\right)$ le rayon spectral de la matrice $M^{-1}N$. La méthode itérative associée à la décomposition $A = M - N$ converge si et seulement si $\rho\left(M^{-1}N\right) < 1$.*

Démonstration : On pose $B = M^{-1}N$ et on suppose que $\rho(B) < 1$. En écrivant que l'équation $Ax = b$ équivaut à $x = Bx + M^{-1}b$, on déduit que :

$$\forall k \in \mathbb{N}, \quad x^{(k+1)} - x = B\left(x^{(k)} - x\right) = B^{k+1}\left(x^{(0)} - x\right).$$

Avec $\rho(B) < 1$ on déduit que $\lim\limits_{k \to +\infty} B^k = 0$ et donc que la suite $\left(x^{(k)}\right)_{k \in \mathbb{N}}$ converge vers x, quelle que soit la valeur initiale $x^{(0)}$ (corollaire 4.16).

Réciproquement, si la méthode associée à la décomposition $A = M - N$ est convergente, alors pour tout vecteur $y^{(0)} \in \mathbb{R}^n$ la suite $\left(y^{(k)}\right)_{k \in \mathbb{N}}$ définie par $y^{(k+1)} = By^{(k)}$ pour tout entier k est convergente vers le vecteur nul, ce qui équivaut à $\rho(B) < 1$ (corollaire 4.16). ■

Remarque 5.8 : *S'il existe une norme matricielle telle que $\|B\| < 1$, alors la méthode itérative correspondante est convergente puisque $\rho(B) \leq \|B\|$.*

Pour toute norme matricielle induite par une norme vectorielle, on a :

$$\left\|x^{(k)} - x\right\| = \left\|B^k\left(x^{(0)} - x\right)\right\| \leq \left\|B^k\right\| \left\|x^{(0)} - x\right\|.$$

On peut donc utiliser $\left\|B^k\right\|$ pour avoir une idée de la vitesse de convergence d'une méthode itérative. De manière plus précise, on peut remarquer que :

$$\forall k \in \mathbb{N}, \quad \left\|B^k\right\| = \sup_{y^{(0)} \in \mathbb{R}^n - \{0\}} \frac{\left\|B^k y^{(0)}\right\|}{\left\|y^{(0)}\right\|} = \sup_{y^{(0)} \in \mathbb{R}^n - \{0\}} \frac{\left\|y^{(k)}\right\|}{\left\|y^{(0)}\right\|},$$

la suite $\left(y^{(k)}\right)_{k \in \mathbb{N}}$ étant définie par $y^{(k+1)} = By^{(k)}$. Ce qui peut aussi s'écrire, en notant x la solution de $Ax = b$:

$$\forall k \in \mathbb{N}, \quad \left\|B^k\right\| = \sup_{x^{(0)} \in \mathbb{R}^n - \{x\}} \frac{\left\|x^{(k)} - x\right\|}{\left\|x^{(0)} - x\right\|}.$$

D'autre part, on sait que $\rho(B) = \lim\limits_{k \to +\infty} \left(\left\|B^k\right\|^{\frac{1}{k}}\right)$ (corollaire 4.6).

Ces considérations nous amènent à donner les définitions suivantes.

Définition 5.7 : *Soit $A = M - N$ une matrice inversible avec M inversible telle que $\rho\left(M^{-1}N\right) < 1$. La quantité :*

$$R_\infty\left(M^{-1}N\right) = -\ln\left(\rho\left(M^{-1}N\right)\right)$$

est appelée le taux asymptotique de convergence de la méthode itérative associée à la décomposition $A = M - N$.

Pour toute norme matricielle $\|\cdot\|$ induite par une norme vectorielle et pour tout entier k strictement positif tel que $\left\|B^k\right\| < 1$, la quantité :

$$R_k\left(M^{-1}N\right) = -\frac{1}{k} \ln\left(\left\|\left(M^{-1}N\right)^k\right\|\right)$$

est appelée le taux moyen de convergence pour k itérations et pour la norme $\|\cdot\|$ de la méthode itérative associée à la décomposition $A = M - N$.

On a bien sûr :

$$R_\infty\left(M^{-1}N\right) = \lim_{k\to+\infty} R_k\left(M^{-1}N\right).$$

La méthode itérative associée à la décomposition $A = M - N$ sera d'autant plus performante que $R_\infty\left(M^{-1}N\right)$ est grand.

12. Méthode de Jacobi

Si $A = \left((a_{ij})\right)_{1\le i,j\le n}$, on choisit $M = D$ matrice diagonale définie par $d_{ii} = a_{ii}$ pour tout $i = 1,\dots,n$. Il faut donc supposer que tous les coefficients diagonaux de A sont non nuls.

L'algorithme de construction des $x^{(k)}$ est alors le suivant :

$$x_i^{(k+1)} = -\frac{1}{a_{ii}}\sum_{\substack{j=1\\ j\neq i}}^{n} a_{ij}x_j^{(k)} + b_i \quad (i = 1,\dots,n).$$

Remarque 5.9 *: Le calcul des composantes de $x^{(k+1)}$ nécessite de garder en mémoire le vecteur $x^{(k)}$. Une itération va donc immobiliser $2n$ cases mémoires.*

Remarque 5.10 *: On peut décider d'arrêter les itérations à un rang $MaxIter$ donné ou lorsque*

$$\left\|x^{(k+1)} - x^{(k)}\right\| < \epsilon\left\|x^{(k+1)}\right\|.$$

(où ϵ est une précision donnée).

Remarque 5.11 *: Comme valeur initiale, on peut prendre $x^{(0)} = \left(\left(\dfrac{b_i}{a_{ii}}\right)\right)_{1\le i\le n}$.*

Dans le cas particulier des matrices à diagonale strictement dominante, on a le résultat suivant.

Théorème 5.6 : *Si la matrice A est à diagonale strictement dominante, alors la méthode de Jacobi converge.*

Démonstration : Une matrice à diagonale strictement dominante ayant tous ses termes diagonaux non nuls, on peut donc utiliser la méthode de Jacobi.

Si $J = D^{-1}N$, avec $A = D - N$, on a :

$$\|J\|_\infty = \max_{1\le i\le n}\sum_{\substack{j=1\\ j\neq i}}^{n}\left|\frac{a_{ij}}{a_{ii}}\right| < 1,$$

et la méthode de Jacobi est convergente. ∎

13. Méthode de Gauss-Seidel

On choisit pour matrice M le triangle inférieur de A. C'est-à-dire que M est définie par :

$$\begin{cases} m_{ij} = 0, & \text{pour } 1 \leq i < j \leq n, \\ m_{ij} = a_{ij}, & \text{pour } 1 \leq j \leq i \leq n. \end{cases}$$

En notant $A = M - N$, le vecteur $x^{(k+1)}$ est solution du système triangulaire inférieur $Mx^{(k+1)} = Nx^{(k)} + b$. D'où l'algorithme de Gauss-Seidel :

$$a_{ii}x_i^{(k+1)} = -\sum_{j=1}^{i-1} a_{ij}x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{ij}x_j^{(k)} + b_i \quad (i = 1, \ldots, n).$$

Remarque 5.12 : *Cet algorithme est en fait une amélioration de l'algorithme de Jacobi. En effet, dans le calcul de $x_i^{(k+1)}$, on utilise les composantes 1 à $i-1$ de $x^{(k+1)}$ (alors que dans la méthode de Jacobi ce sont celles de $x^{(k)}$ qui sont utilisées) et les composantes $i+1$ à n de $x^{(k)}$ (comme dans la méthode de Jacobi). Cet algorithme sera donc en général plus performant que celui de Jacobi.*

Remarque 5.13 : *Pour une itération on garde seulement n termes en mémoire.*

Les théorèmes qui suivent nous donnent des conditions suffisantes de convergence de la méthode de Gauss-Seidel.

Théorème 5.7 : *Si la matrice A est à diagonale strictement dominante, alors la méthode de Gauss-Seidel converge.*

Démonstration : Une matrice à diagonale strictement dominante ayant tous ses termes diagonaux non nuls, on peut donc utiliser la méthode de Gauss-Seidel.

Soient A à diagonale strictement dominante, $G = M^{-1}N$, $\lambda \in \mathbb{C}$ une valeur propre de G et $x \in \mathbb{C}^n$ un vecteur propre non nul associé. On a alors $Nx = \lambda Mx$, c'est-à-dire :

$$\begin{cases} -\sum\limits_{j=i+1}^{n} a_{ij}x_j = \lambda \sum\limits_{j=1}^{i} a_{ij}x_j \quad (i = 1, \ldots, n-1), \\ 0 = \lambda \sum\limits_{j=1}^{n} a_{nj}x_j. \end{cases}$$

Si on suppose que $|\lambda| \geq 1$, en prenant i tel que $\|x\|_\infty = |x_i|$, on déduit que :

$$|\lambda|\,|a_{ii}| \leq \sum_{j \neq i} |a_{ij}|\,|\lambda|,$$

ce qui contredit le fait que la matrice A est à diagonale strictement dominante. Le rayon spectral de G est donc strictement inférieur à 1 et la méthode de Gauss-Seidel converge. ■

Théorème 5.8 : *Si la matrice A est symétrique définie positive, alors la méthode de Gauss-Seidel converge.*

Démonstration : Une matrice symétrique définie positive ayant tous ses termes diagonaux strictement positifs, on peut donc utiliser la méthode de Gauss-Seidel.

Soient A symétrique définie positive, $\lambda \in \mathbb{C}$ une valeur propre de $G = M^{-1}N$ et $x \in \mathbb{C}^n$ un vecteur propre non nul associé. En écrivant $A = D + E + F$, où D est la diagonale de A, E le triangle inférieur strict, F le triangle supérieur strict et en notant $\langle \cdot \mid \cdot \rangle$ le produit scalaire hermitien canonique de $\mathbb{C}^n$, on a :

$$\langle x \mid Ax \rangle = \langle x \mid Dx \rangle + \langle x \mid Ex \rangle + \langle x \mid Fx \rangle . \tag{5.2}$$

Puis, avec $Fx = -\lambda (Dx + Ex)$ $(M = D + E,\ N = -F)$, on déduit que :

$$\langle x \mid Ax \rangle = \left(1 - \overline{\lambda}\right) \left(\langle x \mid Dx \rangle + \langle x \mid Ex \rangle\right),$$

donc $\lambda \neq 1$ (puisque $\langle x \mid Ax \rangle > 0$) et

$$\frac{1}{1 - \overline{\lambda}} \langle x \mid Ax \rangle = \langle x \mid Dx \rangle + \langle x \mid Ex \rangle . \tag{5.3}$$

Par conjugaison complexe de (5.2), en considérant que A et D sont hermitiennes et que ${}^t\overline{E} = F$ on déduit que :

$$\frac{1}{1 - \lambda} \langle x \mid Ax \rangle = \langle x \mid Dx \rangle + \langle x \mid Fx \rangle . \tag{5.4}$$

En faisant (5.3) + (5.4) − (5.2), on en conclut que :

$$\frac{1 - |\lambda|^2}{|1 - \lambda|^2} \langle x \mid Ax \rangle = \langle x \mid Dx \rangle .$$

Avec $\langle x \mid Ax \rangle > 0$ et $\langle x \mid Dx \rangle > 0$ (A et D sont définies positives) on déduit que $|\lambda| < 1$. Le rayon spectral de G est donc strictement inférieur à 1 et la méthode de Gauss-Seidel converge. ∎

Remarque 5.14 : *De manière générale, quand les méthodes de Jacobi et de Gauss-Seidel sont convergentes il vaut mieux choisir celle de Gauss-Seidel. Mais il se peut que la méthode de Gauss-Seidel diverge alors que celle de Jacobi converge comme le montre l'exemple de la matrice* $A = \begin{pmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}$.

Pour la méthode de Jacobi toutes les valeurs propres de J sont nulles et pour la méthode de Gauss-Seidel, on a $G = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & -2 & 4 \end{pmatrix}$ *avec les valeurs propres* 0, 0,35 *et* 5,64.

Remarque 5.15 : *Dans le cas des matrices tridiagonales les méthodes de Jacobi et Gauss-Seidel convergent ou divergent simultanément et dans le cas de la convergence c'est la méthode de Gauss-Seidel qui est la plus rapide (exercice 5.9).*

14. Méthode de relaxation

On pose $A = D+E+F$, où D est la partie diagonale de A, E sa partie triangulaire inférieure stricte et F sa partie triangulaire supérieure stricte. On suppose toujours que la matrice A est inversible et que tous ses coefficients diagonaux sont non nuls.

En vue d'accélérer la convergence de la méthode de Gauss-Seidel, on introduit un paramètre dans la matrice $G = -(D+E)^{-1} F$. Précisément, on pose :

$$M_\omega = \frac{1}{\omega} D + E, \quad A = M_\omega - N_\omega, \quad L_\omega = M_\omega^{-1} N_\omega,$$

où ω est un paramètre réel non nul à préciser. On a alors :

$$L_\omega = (D+\omega E)^{-1} \big((1-\omega) D - \omega F\big).$$

Le vecteur $x^{(k+1)}$ est alors la solution du système triangulaire :

$$(D+\omega E)\, x^{(k+1)} = \big((1-\omega) D - \omega F\big) x^{(k)} + \omega b.$$

Ce qui donne l'algorithme de calcul :

$$x_i^{(k+1)} = (1-\omega)\, x_i^{(k)} + \omega \hat{x}_i^{(k+1)} \quad (i = 1, \ldots, n),$$

où $\widehat{x}_i^{(k+1)}$ est donné par les formules de Gauss-Seidel en fonction des composantes 1 à $i-1$ de $x^{(k+1)}$ et des composantes $i+1$ à n de $x^{(k)}$, soit :

$$a_{ii} \widehat{x}_i^{(k+1)} = -\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} - \sum_{j=i+1}^{n} a_{i,j} x_j^{(k)} + b_i \quad (i = 1, \ldots, n).$$

Ce qui peut encore s'écrire :

$$x^{(k+1)} = x^{(k)} - \omega \xi^{(k+1)} \quad (k \geq 0),$$

où $\xi^{(k+1)}$ est le vecteur résidu défini par :

$$\xi_i^{(k+1)} = \frac{1}{a_{ii}} \left(\sum_{j=1}^{i-1} a_{ij} x_j^{(k+1)} + \sum_{j=i}^{n} a_{i,j} x_j^{(k)} - b_i \right) \quad (i = 1, \ldots, n).$$

Comme test de convergence on peut prendre $\left\| \xi^{(k)} \right\| < \epsilon$, où $\epsilon > 0$ est une précision donnée, une norme matricielle étant choisie.

Une condition nécessaire sur ω pour que la méthode converge est donnée par le résultat suivant.

Théorème 5.9 : *Avec les notations de ce paragraphe, la méthode de relaxation correspondante ne peut converger que si $\omega \in \left]0, 2\right[$.*

Démonstration : Si $\lambda_1, \ldots, \lambda_n$ sont les valeurs propres de L_ω, on a alors :

$$\prod_{i=1}^{n} \lambda_i = \det(L_\omega) = \frac{\det\big((1-\omega) D - \omega F\big)}{\det(D + \omega E)} = (1-\omega)^n$$

et $\rho(L_\omega) \geq |\omega - 1|$. ∎

Dans le cas où la matrice A est symétrique définie positive, la condition nécessaire du théorème précédent est aussi suffisante.

Théorème 5.10 (Ostrowski, Reich) : *Soit A une matrice symétrique définie positive. Avec les notations de ce paragraphe, la méthode de relaxation correspondante est convergente si et seulement si $\omega \in]0,2[$.*

Démonstration : Si la matrice A est symétrique définie positive alors $a_{ii} > 0$ pour tout i compris entre 1 et n. On peut donc définir la suite $\left(x^{(k)}\right)_{k\in\mathbb{N}}$ de la méthode de relaxation.

Avec le théorème précédent, il nous suffit de montrer que la méthode converge si $\omega \in]0,2[$.

En reprenant les notations du début de ce paragraphe, on a $A = M_\omega - N_\omega$, avec $M_\omega = \dfrac{1}{\omega} D + E$ et $N_\omega = \left(\dfrac{1}{\omega} - 1\right) D - F$.

On note $\langle \cdot \mid \cdot \rangle$ le produit scalaire hermitien canonique de $\mathbb{C}^n$.

Soient $\lambda \in \mathbb{C}$ une valeur propre de $L_\omega = M_\omega^{-1} N_\omega$ et $x \in \mathbb{C}^n - \{0\}$ un vecteur propre associé.

On a $N_\omega x = \lambda M_\omega x$ et

$$\langle x \mid Ax \rangle = \langle x \mid M_\omega x \rangle - \langle x \mid N_\omega x \rangle = \left(1 - \bar{\lambda}\right) \langle x \mid M_\omega x \rangle .$$

La matrice A étant symétrique définie positive, on en déduit que $\lambda \neq 1$ et

$$\frac{1}{1-\bar{\lambda}} \langle x \mid Ax \rangle = \frac{1}{\omega} \langle x \mid Dx \rangle + \langle x \mid Ex \rangle . \tag{5.5}$$

Par conjugaison complexe, en considérant que A est hermitienne et que ${}^tE = F$, on déduit que :

$$\frac{1}{1-\lambda} \langle x \mid Ax \rangle = \frac{1}{\omega} \langle x \mid Dx \rangle + \langle x \mid Fx \rangle . \tag{5.6}$$

Enfin, avec $A = D + E + F$, on peut aussi écrire :

$$\langle x \mid Ax \rangle = \langle x \mid Dx \rangle + \langle x \mid Ex \rangle + \langle x \mid Fx \rangle . \tag{5.7}$$

En faisant (5.5) + (5.6) − (5.7), on en conclut que :

$$\frac{\left(1 - |\lambda|^2\right)}{|1-\lambda|^2} \langle x \mid Ax \rangle = \left(\frac{2}{\omega} - 1\right) \langle x \mid Dx \rangle .$$

En tenant compte de la positivité de A et D et du fait que $\omega \in]0,2[$, on déduit que $|\lambda| < 1$.

On a donc ainsi montré que $\rho(L_\omega) < 1$ et que la méthode de relaxation associée converge. ∎

Remarque 5.16 : *Pour $\omega = 1$, on retrouve le fait que la méthode de Gauss-Seidel converge pour les matrices symétriques définies positives.*

Dans le cas des matrices tridiagonales symétriques définies positives on peut faire une étude plus précise.

Lemme 5.4 : *Pour toute matrice $M = ((m_{ij}))_{1 \leq i,j \leq n}$ dans $\mathcal{M}_n(\mathbb{K})$ et pour tout scalaire t la matrice $M(t) = \left((m_{ij}(t))\right)_{1 \leq i,j \leq n}$ définie par :*

$$m_{ij}(t) = t^{i-j} m_{ij} \quad (1 \leq i, j \leq n),$$

est semblable à M.

Démonstration : Il suffit de remarquer que $M(t) = P(t) M P(t)^{-1}$, avec :

$$P(t) = \begin{pmatrix} t & 0 & \dots & 0 \\ 0 & t^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & t^n \end{pmatrix} \in GL_n(\mathbb{K})$$

pour t non nul. ∎

On suppose dans ce qui suit que la matrice A est tridiagonale symétrique définie positive.

Les coefficients diagonaux de A sont alors tous non nuls et on peut définir les matrices $J = -D^{-1}(E + F)$ (méthode de Jacobi) et :

$$L_\omega = \left(\frac{1}{\omega} D + E\right)^{-1} \left(\left(\frac{1}{\omega} - 1\right) D - F\right)$$

(méthode de relaxation).

On désigne par P_J le polynôme caractéristique de J et par P_{L_ω} celui de L_ω.

Lemme 5.5 : *Soit A tridiagonale symétrique définie positive. Avec les notations de ce paragraphe, pour tout $\omega \in \mathbb{R}^*$ et tout $\lambda \in \mathbb{C}^*$ on a :*

$$P_{L_\omega}(\lambda^2) = \omega^n \lambda^n P_J \left(\frac{\lambda^2 + \omega - 1}{\omega \lambda}\right).$$

Démonstration : On a :

$$P_{L_\omega}(\lambda) = \frac{1}{\det\left(\frac{1}{\omega} D + E\right)} \det\left(\left(\frac{1}{\omega} - 1\right) D - F - \lambda \left(\frac{1}{\omega} D + E\right)\right).$$

Soit :

$$P_{L_\omega}(\lambda^2) = \frac{\omega^n}{\det(D)} \det(H_{\omega,\lambda})$$

où $H_{\omega,\lambda}$ est la matrice tridiagonale :

$$H_{\omega,\lambda} = \frac{1-\omega-\lambda^2}{\omega} D - \lambda^2 E - F.$$

En utilisant les notations du lemme 5.4, on peut écrire, pour $\lambda \in \mathbb{C}^*$, que :

$$\det(H_{\omega,\lambda}) = \det\left(H_{\omega,\lambda}\left(\frac{1}{\lambda}\right)\right),$$

$$\text{avec : } H_{\omega,\lambda}\left(\frac{1}{\lambda}\right) = \lambda\left(\frac{1-\omega-\lambda^2}{\omega\lambda} D - E - F\right).$$

Ce qui donne, en tenant compte de :

$$P_J(\mu) = \frac{(-1)^n}{\det(D)} \det(E+F+\mu D),$$

$$P_{L_\omega}(\lambda^2) = \omega^n \lambda^n P_J\left(\frac{\lambda^2+\omega-1}{\omega\lambda}\right). \qquad \blacksquare$$

Lemme 5.6 : *Soit A tridiagonale symétrique définie positive. Avec les notations de ce paragraphe, on a :*

$$P_J(\lambda) = (-1)^n \lambda^q \prod_{k=1}^{p} (\lambda^2 - \mu_k^2),$$

avec $0 < \mu_1 \leq \mu_2 \leq \ldots \leq \mu_p < 1$ dans $\mathbb{R}$, $2p+q=n$ et :

$$P_{L_\omega}(\lambda) = (-1)^n (\lambda+\omega-1)^q \prod_{k=1}^{p} \left((\lambda+\omega-1)^2 - \omega^2\mu_k^2\lambda\right).$$

C'est-à-dire que les valeurs propres de L_ω sont $1-\omega$ d'ordre supérieur ou égal à q et les $\lambda_1(\mu_k,\omega)$, $\lambda_2(\mu_k,\omega)$ pour $k=1,\ldots,p$, où les $\lambda_i(\mu_k,\omega)$ ($i=1,2$) sont les racines de :

$$(\lambda+\omega-1)^2 - \omega^2\mu_k^2\lambda = 0. \tag{5.8}$$

Démonstration : Pour la forme du polynôme caractéristique de J, voir l'exercice 5.9. Il reste à montrer que toutes les valeurs propres de J sont réelles. Si $\mu \in \mathbb{C}$ est une valeur propre de J et $x \in \mathbb{C}^n - \{0\}$ est un vecteur propre associé, on a alors $(1-\mu) Dx = Ax$ et $(1-\mu)\langle Dx \mid x\rangle = \langle Ax \mid x\rangle$. Les matrices A et D étant symétriques définies positives, on déduit que μ est réelle et $\mu < 1$. Comme $-\mu$ est aussi valeur propre de J on déduit que $|\mu| < 1$.

En utilisant le lemme 5.5, on a pour $\lambda \in \mathbb{C}^*$:

$$\begin{aligned} P_{L_\omega}(\lambda^2) &= \omega^n\lambda^n(-1)^n\left(\frac{\lambda^2+\omega-1}{\omega\lambda}\right)^q \prod_{k=1}^{p}\left(\left(\frac{\lambda^2+\omega-1}{\omega\lambda}\right)^2 - \mu_k^2\right) \\ &= (-1)^n(\lambda^2+\omega-1)^q \prod_{k=1}^{p}\left((\lambda^2+\omega-1)^2 - \mu_k^2\omega^2\lambda^2\right), \end{aligned}$$

ce qui équivaut à :

$$P_{L_\omega}(\lambda) = (-1)^n (\lambda+\omega-1)^q \prod_{k=1}^{p} \left((\lambda+\omega-1)^2 - \omega^2\mu_k^2\lambda\right). \qquad \blacksquare$$

Lemme 5.7 : *Pour $\mu \in \left]0,1\right[$, $\omega \in \left]0,2\right[$, on considère l'équation en λ :*

$$(\lambda+\omega-1)^2-\omega^2\mu^2\lambda=0 \tag{5.9}$$

et on pose $\omega_0(\mu)=\dfrac{2}{1+\sqrt{1-\mu^2}}$.

Pour $\omega \in \left]0,\omega_0(\mu)\right[$ l'équation (5.9) admet deux racines réelles distinctes :

$$\lambda_1(\mu,\omega)=\frac{\omega^2\mu^2-2(\omega-1)-\omega\mu\sqrt{\omega^2\mu^2-4\omega+4}}{2},$$
$$\lambda_2(\mu,\omega)=\frac{\omega^2\mu^2-2(\omega-1)+\omega\mu\sqrt{\omega^2\mu^2-4\omega+4}}{2}=\frac{(\omega-1)^2}{\lambda_1(\mu,\omega)},$$

pour $\omega=\omega_0(\mu)$ elle admet une racine double :

$$\lambda_1(\mu,\omega)=\lambda_2(\mu,\omega)=\frac{\omega^2\mu^2}{2}-(\omega-1)=\omega_0(\mu)-1$$

et pour $\omega \in \left]\omega_0(\mu),2\right[$ elle admet deux racines complexes conjuguées :

$$\lambda_1(\mu,\omega)=\frac{\omega^2\mu^2-2(\omega-1)-i\omega\mu\sqrt{-\omega^2\mu^2+4\omega-4}}{2},$$
$$\lambda_2(\mu,\omega)=\frac{\omega^2\mu^2-2(\omega-1)+i\omega\mu\sqrt{-\omega^2\mu^2+4\omega-4}}{2}=\frac{(\omega-1)^2}{\lambda_1(\mu,\omega)},$$

avec :

$$|\lambda_1(\mu,\omega)|=|\lambda_2(\mu,\omega)|=\omega-1.$$

Démonstration : À $\mu \in \left]0,1\right[$ fixé, le discriminant de (5.9) :

$$\Delta(\mu,\omega)=\omega^2\mu^2\left(\omega^2\mu^2-4\omega+4\right)$$

a une seule racine dans $\left]0,2\right[$:

$$\omega_0(\mu)=\frac{2}{1+\sqrt{1-\mu^2}}.$$

Le calcul des racines de (5.9) est alors immédiat. ■

Théorème 5.11 (Young, Varga) : *Si A est symétrique définie positive et tridiagonale, alors :*

1. *les méthodes de Jacobi et de relaxation pour $\omega \in \left]0,2\right[$ sont convergentes ;*
2. *en posant $r=\rho(J)$ et $\omega_0(r)=\dfrac{2}{1+\sqrt{1-r^2}}$, on a :*

$$\rho(L_\omega)=\begin{cases}\dfrac{\omega^2r^2-2(\omega-1)+\omega r\sqrt{\omega^2r^2-4\omega+4}}{2} & \text{si } 0<\omega<\omega_0(r),\\ \omega-1 & \text{si } \omega_0(r)\leq\omega<2.\end{cases}$$

C'est-à-dire que la fonction $\omega \longmapsto \rho(L_\omega)$ a l'allure indiquée par la figure 5.1, page ci-contre, et que la valeur optimale du paramètre de relaxation est $\omega_0(r)$ avec $\rho\left(L_{\omega_0(r)}\right)=\omega_0(r)-1$.

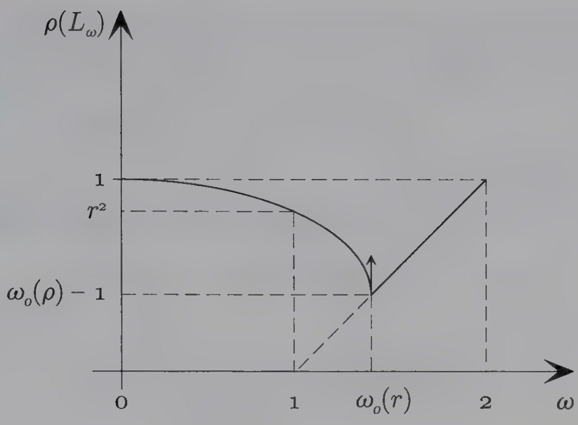

Figure 5.1.

Démonstration :

1. La matrice A étant symétrique définie positive, on sait alors que $r = \rho(J) \in [0, 1[$ et que la méthode de Jacobi converge. Pour $\omega \in]0, 2[$ la méthode de relaxation converge d'après le théorème 5.10.

2. Si $r = 0$, alors toutes les valeurs propres de J sont nulles et $1 - \omega$ est la seule valeur propre de L_ω d'après le lemme 5.6. Et dans ce cas on a $\rho(L_\omega) = |1 - \omega|$ (Fig. 5.2).

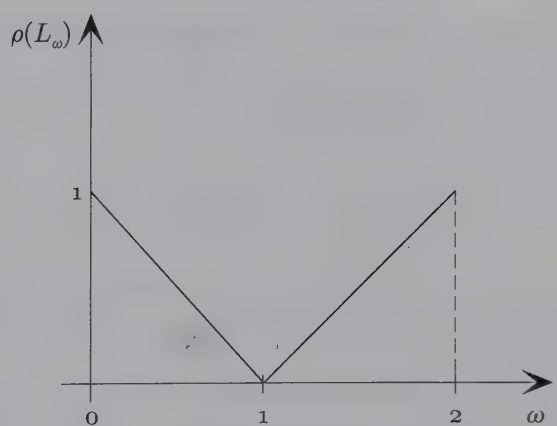

Figure 5.2.

Si $r \in]0, 1[$, d'après les lemmes 5.6 et 5.7 on a pour $\omega \in]\omega_0(\mu), 2[\subset]1, 2[$, $\rho(L_\omega) = |1 - \omega| = 1 - \omega$. Si $\omega \in]0, \omega_0(\mu)[$, en remarquant que la fonction $\mu \longmapsto \lambda_1(\mu, \omega)$ est décroissante et $\mu \longmapsto \lambda_2(\mu, \omega)$ croissante sur $\mathbb{R}^+$, on déduit avec les notations du lemme 5.6 que les valeurs propres de L_ω vérifient :

$$\lambda_1(\mu_p, \omega) \leq \ldots \leq \lambda_1(\mu_1, \omega) < 1 - \omega < \lambda_2(\mu_1, \omega) \leq \ldots \leq \lambda_2(\mu_p, \omega)$$

et $\rho(L_\omega) = \lambda_2(\mu_p, \omega) = \lambda_2(r, \omega)$. ■

***Remarque 5.17** : Du point de vue numérique il est préférable de surestimer la valeur optimale de ω du fait que la pente de la demi-tangente à droite en $\omega_0(r)$ est égale à 1 alors qu'elle est infinie à gauche.*

***Remarque 5.18** : Si on sait calculer la valeur optimale de ω, ce choix n'est pas judicieux dès la première itération. Le procédé d'accélération de convergence de Tchebychev consiste à changer de paramètre ω à chaque itération de la manière suivante. Pour première valeur de ω, on prend $\omega_1 = 1$, puis à l'étape suivante on prend $\omega_2 = \dfrac{1}{1-\dfrac{\rho(J)^2}{2}}$ et aux étapes suivantes on prendra :*

$$\omega_{k+1} = \frac{1}{1-\dfrac{\rho(J)^2\omega_k}{4}}.$$

La suite $(\omega_k)_{k\in\mathbb{N}}$ converge en fait vers $\omega_0(r)$ et le procédé de Tchebychev va diminuer, en général, le nombre d'itérations.

***Remarque 5.19 (méthode de relaxation par blocs)** : Lorsque la dimension du système est très grande, on aura intérêt à partitionner la matrice A en blocs, ce découpage étant adapté à la forme particulière de A. On a donc :*

$$A = A = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1m} \\ A_{21} & A_{22} & \dots & A_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1} & A_{m2} & \dots & A_{mm} \end{pmatrix},$$

où A_{ij} est une matrice à n_i lignes et n_j colonnes, avec $\sum_{i=1}^{n} n_i = n$, les matrices A_{ii} étant supposées inversibles. Le second membre b et l'inconnue x sont alors partitionnés de manière analogue, soit :

$$b = \begin{pmatrix} B_1 \\ B_2 \\ . \\ . \\ B_m \end{pmatrix}, \quad x = \begin{pmatrix} X_1 \\ X_2 \\ . \\ . \\ X_m \end{pmatrix}.$$

Le système à résoudre peut alors s'écrire par blocs :

$$\sum_{j=1}^{m} A_{ij}X_j = B_i \quad (i = 1, \dots, m).$$

Et les formules pour la méthode de relaxation s'écrivent :

$$X^{(k+1)} = X^{(k)} - \omega\Xi^{(k+1)} \quad (k \geq 0),$$

où $\Xi^{(k+1)}$ est le vecteur résidu défini par :

$$\Xi^{(k+1)} = A_{ii}^{-1}\left(\sum_{j=1}^{i-1} A_{ij}X_j^{(k+1)} + \sum_{j=i}^{n} A_{ij}X_j^{(k)} - B_i\right) \quad (i = 1, \dots, m).$$

Cette méthode sera intéressante si les A_{ii} sont facilement inversibles.

La méthode de relaxation est utilisée pour la résolution d'une équation aux dérivées partielles de type elliptique par la méthode des différences finies. On aura à résoudre des systèmes tridiagonaux par blocs, les matrices blocs de la diagonale étant tridiagonales et les autres diagonales.

15. Méthodes de descente et de gradient

L'espace vectoriel $\mathbb{R}^n$ est muni de sa structure euclidienne canonique. On note $\langle \cdot \mid \cdot \rangle$ le produit scalaire euclidien, $\|\cdot\|_2$ la norme associée et l'espace vectoriel $\mathcal{M}_n(\mathbb{R})$ est muni de la norme matricielle induite par $\|\cdot\|_2$.

On désigne par $\{e_1, e_2, \ldots, e_n\}$ la base canonique de $\mathbb{R}^n$.

On rappelle qu'une fonction φ définie sur un ouvert non vide $\mathcal{O}$ de $\mathbb{R}^n$ et à valeurs réelles est dite différentiable en $x \in \mathcal{O}$ s'il existe une forme linéaire L sur $\mathbb{R}^n$ telle que :

$$\varphi(x+h) = \varphi(x) + L(h) + o(\|h\|)$$

pour tout h dans un voisinage de 0.

Le vecteur gradient de φ en x est le vecteur noté $d\varphi(x)$ défini par :

$$\forall h \in \mathbb{R}^n, \quad L(h) = \langle d\varphi(x) \mid h \rangle \cdot$$

Dans la base canonique de $\mathbb{R}^n$ ce vecteur s'écrit :

$$d\varphi(x) = \sum_{k=1}^{n} \frac{\partial \varphi}{\partial x_k}(x)\, e_k.$$

On dit que la fonction φ est différentiable sur $\mathcal{O}$ si elle est différentiable en tout point de $\mathcal{O}$.

On dit la fonction φ admet un minimum [resp. maximum] local en $x_0 \in \mathcal{O}$ s'il existe un voisinage ouvert $\mathcal{V}$ de x_0 dans $\mathcal{O}$ tel que :

$$\forall x \in \mathcal{V}, \quad \varphi(x) \geq \varphi(x_0) \quad [\text{resp. } \forall x \in \mathcal{V}, \quad \varphi(x) \leq \varphi(x_0).]$$

On dit la fonction φ admet un extremum local en $x_0 \in \mathcal{O}$ si elle admet un minimum ou un maximum local en x_0.

On rappelle enfin un résultat donnant une condition nécessaire pour qu'une fonction différentiable en un point d'un ouvert de $\mathbb{R}^n$ admette un extremum local en ce point.

Théorème 5.12 : *Soient $\mathcal{O}$ un ouvert non vide de $\mathbb{R}^n$, φ une fonction définie sur $\mathcal{O}$ à valeurs réelles et différentiable en un point $x_0 \in \mathcal{O}$. Si φ admet un extremum local en x_0 alors le vecteur gradient $d\varphi(x_0)$ est nul.*

Démonstration : On suppose que φ admet un maximum local en x_0.

L'ensemble $\mathcal{O}$ étant ouvert, pour tout vecteur $h \in \mathbb{R}^n$ il existe un réel α strictement positif tel que la fonction :

$$f : t \mapsto \varphi(x_0 + th)$$

soit définie sur $]-\alpha, \alpha[$. Cette fonction est dérivable en 0 avec $f'(0) = \langle d\varphi(x_0) \mid h \rangle$.

En écrivant que :

$$\begin{cases} f'(0) = \lim\limits_{\substack{t \to 0 \\ t > 0}} \dfrac{\varphi(x_0 + th) - \varphi(x_0)}{t} \leq 0, \\ f'(0) = \lim\limits_{\substack{t \to 0 \\ t < 0}} \dfrac{\varphi(x_0 + th) - \varphi(x_0)}{t} \geq 0, \end{cases}$$

on déduit que $f'(0) = 0$.

On a donc ainsi montré que $\langle d\varphi(x_0) \mid h \rangle = 0$ pour tout $h \in \mathbb{R}^n$, ce qui équivaut à $d\varphi(x_0) = 0$. ■

Définition 5.8 : *On appelle fonctionnelle quadratique sur $\mathbb{R}^n$ toute fonction φ définie sur $\mathbb{R}^n$ à valeurs réelles telle que :*

$$\forall x \in \mathbb{R}^n, \quad \varphi(x) = \frac{1}{2} \langle Ax \mid x \rangle - \langle b \mid x \rangle,$$

où A est une matrice symétrique réelle d'ordre n et b un vecteur de $\mathbb{R}^n$.

Dans la suite de ce paragraphe, on se donne une matrice symétrique réelle définie positive d'ordre n, un vecteur non nul b dans $\mathbb{R}^n$ et on note φ la fonctionnelle quadratique associée.

Pour tout réel λ on désigne par $\mathcal{H}_\lambda$ l'hyper-quadrique d'équation $\varphi(x) = \lambda$.

La matrice A étant inversible, le système linéaire $Ax = b$ admet une unique solution que nous noterons u.

On a écarté le cas $b = 0$ qui donne la solution triviale $u = 0$.

Du fait que la matrice A est symétrique, on déduit que pour tout vecteur $\delta \in \mathbb{R}^n$ et tout réel t, on a :

$$\varphi(u + t\delta) = \frac{1}{2} \langle A\delta \mid \delta \rangle t^2 + \langle Au - b \mid \delta \rangle t + \varphi(u)$$

et avec $Au = b$, $\varphi(u) = -\frac{1}{2} \langle Au \mid u \rangle$, on obtient :

$$\varphi(u + t\delta) = \frac{1}{2} \langle A\delta \mid \delta \rangle t^2 - \frac{1}{2} \langle Au \mid u \rangle. \tag{5.10}$$

Le vecteur u étant non nul (on a pris b non nul), on a $\langle Au \mid u \rangle > 0$ (la matrice A est définie positive) et pour δ non nul le vecteur $u + t\delta$ appartient à l'hyper-quadrique $\mathcal{H}_\lambda$ si et seulement si t est solution de l'équation polynomiale de degré 2 :

$$\langle A\delta \mid \delta \rangle t^2 = 2\lambda + \langle Au \mid u \rangle.$$

Pour λ positif ou nul on a deux solutions réelles :

$$t = \pm\sqrt{\frac{2\lambda + \langle Au \mid u \rangle}{\langle A\delta \mid \delta \rangle}}.$$

Le vecteur δ étant quelconque dans $\mathbb{R}^n - \{0\}$, on déduit que pour tout réel λ positif ou nul l'ensemble $\mathcal{H}_\lambda$ est un hyper-ellipsoïde de centre u.

Lemme 5.8 : *La fonctionnelle quadratique φ associée à la matrice A et au vecteur b est différentiable sur $\mathbb{R}^n$ avec :*

$$\forall x \in \mathbb{R}^n, \quad d\varphi(x) = Ax - b.$$

Démonstration : En utilisant la symétrie de la matrice A, on a pour tous vecteurs x, h dans $\mathbb{R}^n$:

$$\varphi(x+h) = \varphi(x) + \langle Ax - b \mid h \rangle + \frac{1}{2} \langle Ah \mid h \rangle$$

et avec $|\langle Ah \mid h \rangle| \leq \|A\|_2 \|h\|_2^2$, on déduit le résultat. ∎

Remarque 5.20 *: Le vecteur gradient $d\varphi(x) = Ax - b$ est aussi appelé vecteur résidu et souvent noté $r(x)$.*

Remarque 5.21 *: Pour tout réel λ positif ou nul le vecteur $d\varphi(x)$ est dirigé suivant la normale à l'hyper-ellipsoïde $\mathcal{H}_\lambda$ passant par le centre u.*

Théorème 5.13 : *La solution u du système linéaire $Ax = b$ est le vecteur qui réalise le minimum global de la fonctionnelle quadratique φ.*

Démonstration : En utilisant (5.10) avec $t = 1$ et $\delta = x - u$ on a pour tout vecteur x dans $\mathbb{R}^n$:

$$\varphi(x) = \frac{1}{2} \langle A(x-u) \mid x - u \rangle - \frac{1}{2} \langle Au \mid u \rangle \cdot$$

La matrice A étant définie positive, on en déduit que :

$$\forall x \in \mathbb{R}^n - \{0\}, \quad \varphi(x) > -\frac{1}{2} \langle Au \mid u \rangle = \varphi(u),$$

ce qui entraîne que φ atteint son minimum en u et que ce minimum est unique sur $\mathbb{R}^n$.

Réciproquement si φ admet un extremum global en $x \in \mathbb{R}^n$ alors $d\varphi(x) = 0$, c'est-à-dire que x est solution de $Ax = b$ et nécessairement $x = u$. Cet extremum est un minimum et il est unique. ∎

La résolution du système linéaire $Ax = b$ est donc équivalente à la résolution du problème d'optimisation sans contraintes qui consiste à déterminer le vecteur u de $\mathbb{R}^n$ solution de :

$$\varphi(u) = \inf_{x \in \mathbb{R}^n} \varphi(x).$$

Le principe des méthodes de descente pour résoudre ce type de problème consiste à construire une suite de vecteurs $(x_k)_{k\in\mathbb{N}}$ destinée à converger vers la solution cherchée u telle que le passage de x_k à x_{k+1} se fait en résolvant un problème de minimisation plus simple à une variable.

À chaque itération on se donne une direction de descente $\delta_k \in \mathbb{R}^n - \{0\}$ et on cherche à minimiser la restriction de φ à la droite passant par x_k et dirigée par δ_k, c'est-à-dire qu'on veut minimiser la fonction d'une variable réelle :

$$t \mapsto \varphi_k(t) = \varphi(x_k + t\delta_k).$$

En remarquant (grâce à la symétrie de la matrice A) que :

$$\varphi_k(t) = \frac{1}{2}\langle A\delta_k \mid \delta_k\rangle t^2 + \langle d\varphi(x_k) \mid \delta_k\rangle t + \varphi(x_k)$$

est un polynôme de degré 2 à coefficient dominant strictement positif (la matrice A est définie positive et le vecteur δ_k est non nul), on calcule facilement la valeur de t qui réalise le minimum de la fonction φ_k sur $\mathbb{R}$. Cette valeur est donnée par :

$$t_k = -\frac{\langle d\varphi(x_k) \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}.$$

En notant :

$$r_k = d\varphi(x_k) = Ax_k - b = A(x_k - u)$$

le vecteur résidu d'ordre k, le vecteur x_{k+1} s'écrit :

$$x_{k+1} = x_k - \frac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\delta_k.$$

Lemme 5.9 : *Avec les notations qui précèdent, le vecteur r_{k+1} est orthogonal au vecteur δ_k.*

Démonstration : On a :

$$r_{k+1} = Ax_{k+1} - b = Ax_k - b + t_k A\delta_k = r_k + t_k A\delta_k$$

et

$$\langle r_{k+1} \mid \delta_k\rangle = \langle r_k \mid \delta_k\rangle - \frac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\langle A\delta_k \mid \delta_k\rangle = 0. \qquad \blacksquare$$

Remarque 5.22 : *Pour $n = 2$, le vecteur $r_{k+1} = d\varphi(x_{k+1})$ est orthogonal à l'ellipse $\mathcal{H}_{\varphi(x_{k+1})}$ et δ_k orthogonal à r_{k+1} est donc tangent à cette ellipse.*

C'est le choix des vecteurs de descente qui va définir une méthode.

En prenant $\delta_k = e_1$, on obtient, en notant $x_j^{(k)}$ les composantes du vecteur x_k dans la base canonique de $\mathbb{R}^n$:

$$x_{k+1} = x_k - \frac{1}{a_{11}} \left(\sum_{j=1}^{n} a_{1j} x_j^{(k)} - b_1 \right) e_1,$$

ce qui donne les formules :

$$\begin{cases} a_{11} x_1^{(k+1)} = - \sum\limits_{j=2}^{n} a_{1j} x_j^{(k)} + b_1, \\ x_j^{(k+1)} = x_j^{(k)} \quad (j = 2, \ldots, n). \end{cases}$$

En remplaçant le vecteur x_k par le vecteur, encore noté x_k, de composantes $x_1^{(k+1)}, x_2^{(k)}, \ldots, x_n^{(k)}$ et en prenant $\delta_k = e_2$, on obtient pour la deuxième composante de x_{k+1} :

$$a_{22} x_2^{(k+1)} = -a_{21} x_1^{(k+1)} - \sum_{j=3}^{n} a_{2j} x_j^{(k)} + b_2,$$

les autres composantes de x_k étant inchangées.

En continuant ainsi avec $e_3, \ldots, e_n$ on retrouve les formules de Gauss-Seidel (paragraphe 13, page 202). On sait que pour A symétrique définie positive cette méthode est convergente.

Le théorème qui suit nous indique comment choisir des directions de descente qui assurent la convergence de la méthode.

On note :

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

les valeurs propres de la matrice A rangées dans l'ordre croissant.

Lemme 5.10 : *Avec les notations qui précèdent, pour tout vecteur x dans $\mathbb{R}^n$ on a :*

$$\langle Ax \mid x \rangle \geq \lambda_1 \|x\|_2^2 .$$

Démonstration : La matrice A, symétrique, se diagonalise dans une base orthonormée $\{f_1, f_2, \ldots, f_n\}$ avec $Af_k = \lambda_k f_k$ pour tout entier k compris entre 1 et n. Dans cette base on a :

$$\begin{cases} x = \sum\limits_{k=1}^{n} \alpha_k f_k, \\ Ax = \sum\limits_{k=1}^{n} \alpha_k \lambda_k f_k, \end{cases}$$

de sorte que :

$$\langle Ax \mid x \rangle = \sum_{k=1}^{n} \lambda_k \alpha_k^2 \geq \lambda_1 \sum_{k=1}^{n} \alpha_k^2 = \lambda_1 \|x\|_2^2 . \qquad \blacksquare$$

Théorème 5.14 : *Soit $(x_k)_{k\in\mathbb{N}}$ la suite de vecteurs définie par :*

$$\begin{cases} x_0 \in \mathbb{R}^n, \\ \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \dfrac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\delta_k, \end{cases}$$

où $r_k = Ax_k - b$ est le vecteur résidu d'ordre k et δ_k un vecteur unitaire dans $\mathbb{R}^n$.
S'il existe un réel α strictement positif tel que :

$$\forall k \in \mathbb{N}, \quad \langle r_k \mid \delta_k\rangle \geq \alpha \left\|r_k\right\|_2, \tag{5.11}$$

alors la suite $(x_k)_{k\in\mathbb{N}}$ converge vers la solution u du système linéaire $Ax = b$.

Démonstration : Si il existe un entier naturel k_0 tel que r_{k_0} soit nul alors la suite $(x_k)_{k\in\mathbb{N}}$ est stationnaire sur u à partir du rang k_0 et c'est terminé.

On suppose donc r_k non nul pour tout entier naturel k.

On remarque tout d'abord que les inégalités (5.11) entraînent $0 < \alpha \leq 1$. En effet avec l'inégalité de Cauchy-Schwarz on a $\langle r_k \mid \delta_k\rangle \leq \left\|\delta_k\right\|_2 \left\|r_k\right\|_2 = \left\|r_k\right\|_2$ ce qui entraîne $\alpha \leq 1$ puisque r_k est non nul.

Avec les inégalités :

$$\forall k \in \mathbb{N}, \quad \left\|x_k - u\right\|_2^2 \leq \frac{1}{\lambda_1}\left\langle A\left(x_k - u\right) \mid x_k - u\right\rangle = \frac{1}{\lambda_1}e_k,$$

il suffit de montrer que la suite $(e_k)_{k\in\mathbb{N}}$ converge vers 0 sur $\mathbb{R}$.

Pour tout entier naturel k on a :

$$e_k = \langle Ax_k - b \mid x_k - u\rangle = \langle r_k \mid x_k - u\rangle$$

et avec $\langle r_{k+1} \mid \delta_k\rangle = 0$ on déduit que :

$$e_{k+1} = \left\langle r_{k+1} \mid x_k - u - \frac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\delta_k\right\rangle = \langle r_{k+1} \mid x_k - u\rangle.$$

Avec $r_{k+1} = r_k - \dfrac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}A\delta_k$ (démonstration du lemme 5.9) on déduit que :

$$e_{k+1} = e_k - \frac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\langle A\delta_k \mid x_k - u\rangle.$$

Enfin avec la symétrie de la matrice A et $Au = b$ on peut écrire que :

$$\langle A\delta_k \mid x_k - u\rangle = \langle \delta_k \mid Ax_k - b\rangle = \langle \delta_k \mid r_k\rangle$$

et

$$e_{k+1} = e_k - \frac{\langle r_k \mid \delta_k\rangle^2}{\langle A\delta_k \mid \delta_k\rangle}.$$

Avec l'hypothèse r_k non nul pour tout entier naturel k on déduit que e_k est non nul pour tout entier naturel k et l'égalité précédente s'écrit :

$$e_{k+1} = e_k \left(1 - \frac{1}{e_k} \frac{\langle r_k \mid \delta_k \rangle^2}{\langle A\delta_k \mid \delta_k \rangle}\right). \tag{5.12}$$

Avec les inégalités :

$$\begin{cases} \langle r_k \mid \delta_k \rangle \geq \alpha \left\| r_k \right\|_2, \\ \langle A\delta_k \mid \delta_k \rangle \leq \left\| A \right\|_2 \left\| \delta_k \right\|_2^2 = \left\| A \right\|_2 = \lambda_n, \\ e_k = \langle r_k \mid A^{-1} r_k \rangle \leq \left\| A^{-1} \right\|_2 \left\| r_k \right\|_2^2 = \dfrac{1}{\lambda_1} \left\| r_k \right\|_2^2, \end{cases}$$

on déduit que :

$$e_{k+1} \leq \left(1 - \frac{\lambda_1}{\lambda_n} \alpha^2\right) e_k,$$

ce qui s'écrit aussi :

$$e_{k+1} \leq \left(1 - \operatorname{cond}_2(A)\, \alpha^2\right) e_k.$$

Avec α et $\operatorname{cond}_2(A)$ dans l'intervalle $]0,1]$ on déduit que :

$$\forall k \in \mathbb{N}, \quad 0 < e_k \leq \beta^k e_0$$

avec $\beta = 1 - \operatorname{cond}_2(A)\, \alpha^2 \in [0, 1[$. Il en résulte que $\lim\limits_{k \to +\infty} e_k = 0$ et $\lim\limits_{k \to +\infty} x_k = u$. ■

En désignant par θ_k la mesure dans $[0, \pi]$ de l'angle des vecteurs r_k et δ_k, la condition (5.11) s'écrit :

$$\forall k \in \mathbb{N}, \quad \cos(\theta_k) \geq \alpha$$

avec $\alpha \in]0,1]$. En particulier on a $\theta_k \in \left[0, \dfrac{\pi}{2}\right[$ et les vecteurs r_k et δ_k ne sont pas orthogonaux.

Le cas $\theta_k = 0$ est réalisé pour $\delta_k = \dfrac{1}{\left\| r_k \right\|_2} r_k$ (et donc $\alpha = 1$) en supposant toujours r_k non nul. La méthode obtenue pour ces choix de vecteurs de descente est la méthode de plus profonde descente ou méthode du gradient $(r_k = d\varphi(x_k))$ à paramètre optimal.

Le théorème précédent nous dit que cette méthode est convergente et nous donne une majoration de l'erreur, ce qui est résumé par le résultat suivant.

Théorème 5.15 : *Soient A une matrice réelle symétrique définie positive d'ordre n, b un vecteur non nul dans $\mathbb{R}^n$ et $(x_k)_{k\in\mathbb{N}}$ la suite de vecteurs définie par :*

$$\begin{cases} x_0 \in \mathbb{R}^n, \\ \forall k \in \mathbb{N}, \quad x_{k+1} = \begin{cases} x_k & \text{si } r_k = Ax_k - b = 0, \\ x_k - \dfrac{\|r_k\|_2^2}{\langle Ar_k \mid r_k\rangle} r_k & \text{si } r_k \neq 0. \end{cases} \end{cases}$$

La suite $(x_k)_{k\in\mathbb{N}}$ converge vers la solution u du système linéaire $Ax = b$ et une majoration de l'erreur est donnée par :

$$\forall k \in \mathbb{N}, \quad \|x_k - u\|_2 \leq \gamma\big(1 - \mathrm{cond}_2(A)\big)^{\frac{k}{2}}$$

où γ est une constante réelle.

La méthode du gradient conjugué de Hestenes et Stiefel consiste à construire la suite $(x_k)_{k\in\mathbb{N}}$ de la manière suivante

$$\begin{cases} x_0 \in \mathbb{R}^n, \\ \forall k \in \mathbb{N}, \quad x_{k+1} = \begin{cases} x_k & \text{si } r_k = Ax_k - b = 0, \\ x_k - \dfrac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle} \delta_k & \text{si } r_k \neq 0, \end{cases} \end{cases}$$

avec les vecteurs de descente δ_k, pour $k \geq 1$ (quand ils existent), choisis dans le plan engendré par les vecteurs orthogonaux r_k et δ_{k-1} de la forme :

$$\delta_k = r_k + t_k \delta_{k-1},$$

le coefficient t_k étant déterminé de sorte que le coefficient $1 - \dfrac{1}{e_k} \dfrac{\langle r_k \mid \delta_k\rangle^2}{\langle A\delta_k \mid \delta_k\rangle}$ qui intervient dans la formule (5.12) soit minimum.

Si à l'étape $k \geq 1$ le vecteur r_k est non nul alors le vecteur $e_k = \langle r_k \mid x_k - u\rangle$ est déjà déterminé et avec l'orthogonalité des vecteurs r_k et δ_{k-1} (lemme 5.9) on a :

$$\langle r_k \mid \delta_k\rangle = \langle r_k \mid r_k + t_k\delta_{k-1}\rangle = \|r_k\|_2^2$$

qui est également déterminé. Il s'agit donc de trouver $t_k \in \mathbb{R}$ qui minimise la fonction polynomiale de degré 2 :

$$P_k(t) = \langle A\delta_k \mid \delta_k\rangle = \langle A\delta_{k-1} \mid \delta_{k-1}\rangle t^2 + 2\langle A\delta_{k-1} \mid r_k\rangle t + \langle Ar_k \mid r_k\rangle .$$

Cette fonction atteint son minimum en :

$$t_k = -\frac{\langle A\delta_{k-1} \mid r_k\rangle}{\langle A\delta_{k-1} \mid \delta_{k-1}\rangle} .$$

On préfère écrire que le réel t_k est déterminé par :

$$\langle A\delta_{k-1} \mid r_k + t_k\delta_{k-1}\rangle = 0,$$

ce qui équivaut à :

$$\langle A\delta_{k-1} \mid \delta_k\rangle = 0. \tag{5.13}$$

La matrice A étant symétrique définie positive, l'application $(x, y) \mapsto \langle Ax \mid y\rangle$ définit un produit scalaire sur $\mathbb{R}^n$ et la condition (5.13) signifie que les vecteurs δ_k et δ_{k-1} sont orthogonaux pour ce produit scalaire.

Définition 5.9 : *Si A est une matrice réelle d'ordre n symétrique définie positive, on dit que deux vecteurs x, y dans $\mathbb{R}^n$ sont A-conjugués si $\langle Ax \mid y \rangle = 0$.*

En définitive le vecteur de descente δ_k est défini par :

$$\delta_k = r_k - \frac{\langle A\delta_{k-1} \mid r_k \rangle}{\langle A\delta_{k-1} \mid \delta_{k-1} \rangle}\delta_{k-1}$$

et on a $\delta_k \neq 0$ avec δ_k et δ_{k-1} qui sont A-conjugués.

Lemme 5.11 : *Avec les hypothèses et notations qui précèdent on a :*

$$\begin{cases} \langle r_{k+1} \mid \delta_{k-1} \rangle = 0, \\ \langle r_{k+1} \mid r_k \rangle = 0, \\ t_k = \dfrac{\|r_k\|_2^2}{\|r_{k-1}\|_2^2}. \end{cases}$$

Démonstration : On a, en utilisant l'expression de r_{k+1} obtenue dans la démonstration du lemme 5.9 :

$$\langle r_{k+1} \mid \delta_{k-1} \rangle = \langle r_k \mid \delta_{k-1} \rangle - \frac{\langle r_k \mid \delta_k \rangle}{\langle A\delta_k \mid \delta_k \rangle} \langle A\delta_k \mid \delta_{k-1} \rangle,$$

avec $\langle r_k \mid \delta_{k-1} \rangle = 0$ et :

$$\langle A\delta_k \mid \delta_{k-1} \rangle = \langle A\delta_{k-1} \mid \delta_k \rangle = 0,$$

ce qui donne bien $\langle r_{k+1} \mid \delta_{k-1} \rangle = 0$.

Le vecteur r_{k+1} est donc orthogonal aux vecteurs δ_{k-1} et δ_k, il est donc orthogonal au vecteur r_k qui appartient au plan engendré par les vecteurs δ_{k-1} et δ_k.

Avec :

$$A\delta_{k-1} = \frac{\langle A\delta_{k-1} \mid \delta_{k-1} \rangle}{\langle r_{k-1} \mid \delta_{k-1} \rangle} (r_k - r_{k-1}),$$

$\langle r_k \mid r_{k-1} \rangle = 0$ et $\langle r_k \mid \delta_{k-1} \rangle = 0$, on déduit que :

$$\begin{cases} \langle A\delta_{k-1} \mid r_k \rangle = \dfrac{\langle A\delta_{k-1} \mid \delta_{k-1} \rangle}{\langle r_{k-1} \mid \delta_{k-1} \rangle} \|r_k\|_2^2, \\ \langle A\delta_{k-1} \mid \delta_{k-1} \rangle = -\dfrac{\langle A\delta_{k-1} \mid \delta_{k-1} \rangle}{\langle r_{k-1} \mid \delta_{k-1} \rangle} \langle r_{k-1} \mid \delta_{k-1} \rangle \end{cases}$$

et

$$t_k = \frac{\|r_k\|_2^2}{\langle r_{k-1} \mid \delta_{k-1} \rangle}.$$

Enfin avec $\delta_k = r_k + t_k\delta_{k-1}$ et $\langle r_k \mid \delta_{k-1} \rangle = 0$, on déduit que $\langle r_k \mid \delta_k \rangle = \|r_k\|_2^2$ et

$$t_k = \frac{\|r_k\|_2^2}{\|r_{k-1}\|_2^2}. \qquad \blacksquare$$

En résumé, à l'étape $k \geq 1$, en supposant toujours que r_k est non nul, le calcul de x_{k+1} se fait de la manière suivante :

$$\left\{ \begin{aligned} \delta_k &= r_k + \frac{\|r_k\|_2^2}{\|r_{k-1}\|_2^2}\delta_{k-1}, \\ x_{k+1} &= x_k - \frac{\langle r_k \mid \delta_k\rangle}{\langle A\delta_k \mid \delta_k\rangle}\delta_k. \end{aligned} \right. \tag{5.14}$$

Pour l'initialisation on se donne un vecteur x_0 dans $\mathbb{R}^n$ et on prend $\delta_0 = r_0$.

Lemme 5.12 : *Avec les hypothèses et notations qui précèdent, pour j compris entre 0 et k, les vecteurs r_j sont deux à deux orthogonaux et les vecteurs δ_j sont deux à deux A-conjugués.*

Démonstration : Si $r_0 = Ax_0 - b$ est non nul, on pose alors $\delta_0 = r_0$ et on peut construire les vecteurs x_1, r_1 et δ_1, le vecteur r_1 étant orthogonal à $\delta_0 = r_0$ et le vecteur δ_1 A-conjugué avec δ_0.

Supposons le résultat acquis jusqu'à l'ordre $j-1$ avec j compris entre 2 et k. Il s'agit alors de montrer que $\langle r_j \mid \delta_i\rangle = 0$ et $\langle A\delta_j \mid \delta_i\rangle = 0$ pour i compris entre 0 et $j-1$.

On sait déjà que ces égalités sont vérifiées pour $i = j-1$.

On a :

$$r_j = r_{j-1} + \frac{\langle r_{j-1} \mid \delta_{j-1}\rangle}{\langle A\delta_{j-1} \mid \delta_{j-1}\rangle} A\delta_{j-1}$$

et avec $\langle r_{j-1} \mid r_{j-2}\rangle = 0$ il reste :

$$\langle r_j \mid r_{j-2}\rangle = \frac{\langle r_{j-1} \mid \delta_{j-1}\rangle}{\langle A\delta_{j-1} \mid \delta_{j-1}\rangle} \langle A\delta_{j-1} \mid r_{j-2}\rangle .$$

Pour $j = 2$ on a $r_0 = \delta_0$ et :

$$\langle A\delta_1 \mid r_0\rangle = \langle A\delta_1 \mid \delta_0\rangle = 0.$$

Pour $j > 2$, on a :

$$r_{j-2} = \delta_{j-2} - \frac{\|r_{j-2}\|_2^2}{\|r_{j-3}\|_2^2}\delta_{j-3}$$

et avec l'hypothèse de récurrence on aboutit à :

$$\langle A\delta_{j-1} \mid r_{j-2}\rangle = \langle A\delta_{j-1} \mid \delta_{j-2}\rangle - \frac{\|r_{j-2}\|_2^2}{\|r_{j-3}\|_2^2} \langle A\delta_{j-1} \mid \delta_{j-3}\rangle = 0.$$

On a donc $\langle r_j \mid r_{j-2}\rangle = 0$.

D'autre part, on a :

$$\langle A\delta_j \mid \delta_{j-2}\rangle = \langle \delta_j \mid A\delta_{j-2}\rangle = \left\langle r_j + \frac{\|r_j\|_2^2}{\|r_{j-1}\|_2^2}\delta_{j-1} \mid A\delta_{j-2} \right\rangle = \langle r_j \mid A\delta_{j-2}\rangle$$

et avec :

$$r_{j-1} = r_{j-2} + \frac{\langle r_{j-2} \mid \delta_{j-2} \rangle}{\langle A\delta_{j-2} \mid \delta_{j-2} \rangle} A\delta_{j-2}$$

on a :

$$\langle r_j \mid A\delta_{j-2} \rangle = \frac{\langle A\delta_{j-2} \mid \delta_{j-2} \rangle}{\langle r_{j-2} \mid \delta_{j-2} \rangle} \left(\langle r_j \mid r_{j-1} \rangle - \langle r_j \mid r_{j-2} \rangle \right) = 0$$

$(\langle r_{j-2} \mid \delta_{j-2} \rangle = \|r_{j-2}\|_2^2 \neq 0)$.

On a donc $\langle A\delta_j \mid \delta_{j-2} \rangle = 0$.

De manière analogue on montre que $\langle r_j \mid r_i \rangle = 0$ et $\langle A\delta_j \mid \delta_i \rangle = 0$ pour $i = j-3, \ldots, 0$. ∎

Les vecteurs r_j étant deux à deux orthogonaux et non nuls pour j compris entre 0 et k, on a nécessairement $k \leq n-1$ (dans $\mathbb{R}^n$ il n'est pas possible d'avoir plus de n vecteurs orthogonaux non nuls, puisqu'un tel système est nécessairement libre), c'est-à-dire que le vecteur r_n est nécessairement nul. Précisément il existe un entier p compris entre 0 et $n-1$ tel que $r_p \neq 0$ et $r_{p+1} = 0$, ce qui équivaut à $Ax_{p+1} = b$ et $Ax_k = b$ pour tout $k \geq p+1$. En d'autres termes on a obtenu le résultat suivant.

Théorème 5.16 (Stiefel) : *La méthode du gradient conjugué converge en n itérations au plus, c'est-à-dire que la suite de vecteurs $(x_k)_{k\in\mathbb{N}}$ définie par (5.14) avec $x_0 \in \mathbb{R}^n$ et $\delta_0 = r_0$ est stationnaire à partir d'un rang $p \leq n$ sur la solution du système linéaire $Ax = b$.*

16. Exercices

Exercice 5.1 : *Soient α, β dans $\mathbb{K}$ et $A(\alpha,\beta) = \left((a_{ij})\right)_{1\leq i,j\leq n}$ la matrice d'ordre n supérieur ou égal à 3 définie par :*

$$\forall i \in \{1, 2, \ldots, n\}, \quad \begin{cases} a_{ii} = \beta, \\ a_{ij} = \alpha & \text{si } j \in \{1, 2, \ldots, n\} - \{i\}. \end{cases}$$

1. Donner une condition nécessaire et suffisante sur α et β pour que la matrice $A(\alpha,\beta)$ soit inversible.

2. On suppose que $A(\alpha,\beta)$ est inversible. Résoudre le système $A(\alpha,\beta)\, x = e$ dans les deux cas suivants :

(a) Toutes les composantes de e valent 1.

(b) e est un vecteur quelconque.

Solution :

1. Le déterminant de $A(\alpha,\beta)$ est $\Delta(\alpha,\beta) = \left(\beta + (n-1)\alpha\right)(\beta-\alpha)^{n-1}$ (exercice 2.8). On en déduit que la matrice $A(\alpha,\beta)$ est inversible si et seulement si $\beta \neq \alpha$ et $\beta \neq -(n-1)\alpha$.

2. (a) En ajoutant les lignes 2 à n à la première équation, on obtient :

$$\alpha \sum_{j=1}^{n} x_j = \frac{n\alpha}{\beta + (n-1)\alpha}.$$

Puis, en retranchant cette équation aux équations 1 à n, on obtient :

$$\forall i \in \{1,2,\dots,n\}, \quad x_i = \frac{1}{\beta + (n-1)\alpha}.$$

(b) En effectuant les mêmes opérations qu'en 2(a), on aboutit à :

$$\alpha \sum_{j=1}^{n} x_j = \frac{\alpha S(e)}{\beta + (n-1)\alpha},$$

où $S(e) = \sum\limits_{i=1}^{n} e_i$. Puis :

$$\forall i \in \{1,2,\dots,n\}, \quad x_i = \frac{1}{\beta-\alpha}\left\{e_i - \frac{\alpha S(e)}{\beta + (n-1)\alpha}\right\}.$$

Exercice 5.2 : *Pour tout entier $n \geq 1$, on note H_n la matrice de Hilbert d'ordre n définie par :*

$$H_n = \left(\left(\frac{1}{i+j-1}\right)\right)_{1\leq i,j\leq n}.$$

***1.** Calculer le coefficient de $\dfrac{1}{x+n}$ dans la décomposition en éléments simples de la fraction rationnelle :*

$$R_n(x) = \frac{\prod\limits_{k=1}^{n}(x-k)}{\prod\limits_{k=0}^{n}(x+k)}.$$

***2.** On note $A_n = \left((a_{ij})\right)_{1\leq i,j\leq n}$ la matrice d'ordre n définie par :*

$$a_{ij} = \begin{cases} \dfrac{1}{i+j-1} & si\ j = 1,2,\dots,n-1, \quad i = 1,2,\dots,n, \\ R_{n-1}(i) & si\ j = n, \quad i = 1,2,\dots,n. \end{cases}$$

(a) *Montrer que* $\det(A_n) = C_{2n}^n \det(H_n)$.

(b) *Montrer que* $\det(A_n) = R_{n-1}(n) \det(H_{n-1})$.

3. *En posant, pour tout entier* $n \geq 1$, $\Phi_n = \prod_{k=1}^{n} k!$, *montrer que :*

$$\det(H_n) = \frac{\Phi_{n-1}^4}{\Phi_{2n-1}}.$$

4. *Montrer que pour tout entier* $n \geq 1$, $\det(H_n) \leq \frac{1}{n^n}$.

Solution :

1. La décomposition en éléments simples de R_n s'écrit :

$$R_n(x) = \sum_{k=0}^{n} \frac{\lambda_{k,n}}{(x+k)},$$

avec

$$\lambda_{k,n} = \lim_{x \to -k} \big((x+k) R_n(x)\big).$$

Et en particulier le coefficient de $\frac{1}{x+n}$ est donné par :

$$\lambda_{n,n} = \frac{(2n)!}{(n!)^2} = C_{2n}^n.$$

2. On a, en notant $h_{ij} = \frac{1}{i+j-1}$ pour i, j compris entre 1 et n :

$$A_n = \begin{pmatrix} h_{11} & \cdots & h_{1,n-1} & R_{n-1}(1) \\ h_{21} & \cdots & h_{2,n-1} & R_{n-1}(2) \\ \vdots & \ddots & \vdots & \vdots \\ h_{n1} & \cdots & h_{n,n-1} & R_{n-1}(n) \end{pmatrix},$$

avec :

$$R_{n-1}(i) = \sum_{k=0}^{n-1} \frac{\lambda_{k,n-1}}{(i+k)} = \sum_{k=0}^{n-1} \lambda_{k,n-1} h_{i,k+1},$$

ou encore $R_{n-1}(i) = \sum_{j=1}^{n} \lambda_{j-1,n-1} h_{i,j}$ pour tout $i = 1, 2, \ldots, n$.

(a) Si on note C_j la colonne numéro j de H_n, alors la colonne numéro n de A_n est combinaison linéaire des $C_1, C_2, \ldots, C_n$. Le déterminant étant une forme n-linéaire alternée, on en déduit que :

$$\det(A_n) = \lambda_{n-1,n-1} \det(H_n) = C_{2(n-1)}^{n-1} \det(H_n).$$

(b) D'autre part, on a $R_{n-1}(i) = 0$ pour $i = 1, 2, \ldots, n-1$. En développant le déterminant suivant la dernière colonne on en déduit alors que :

$$\det(A_n) = R_{n-1}(n) \det(H_{n-1}).$$

3. On déduit de ce qui précède que

$$\det(H_n) = \frac{R_{n-1}(n)}{\lambda_{n-1,n-1}} \det(H_{n-1}),$$

avec $\lambda_{n-1,n-1} = \dfrac{(2n-2)!}{\left((n-1)!\right)^2}$ et $R_{n-1}(n) = \dfrac{\left((n-1)!\right)^2}{(2n-1)!}$. Ce qui donne :

$$\det(H_n) = \frac{\left((n-1)!\right)^4}{(2n-1)!\,(2n-2)!} \det(H_{n-1}).$$

Par récurrence, on déduit alors que :

$$\det(H_n) = \frac{\left((n-1)!\,(n-2)!\ldots 2!\right)^4}{(2n-1)!\,(2n-2)!\ldots 3!2!} = \frac{\Phi_{n-1}^4}{\Phi_{2n-1}}.$$

4. On a :

$$\det(H_n) = \frac{\left((n-1)!\,(n-2)!\ldots 2!\right)^3}{(2n-1)!\,(2n-2)!\ldots n!},$$

soit, avec $C_{2k}^k = \dfrac{(2k)!}{(k!)^2}$:

$$\det(H_n) = \frac{(n-1)!}{(2n-1)!} \frac{1}{C_{2(n-1)}^{n-1}} \frac{(n-2)!}{(2n-3)!} \frac{1}{C_{2(n-2)}^{n-2}} \cdots \frac{2!}{5!} \frac{1}{C_4^2} \frac{1}{3!} \frac{1}{C_2^1}.$$

En posant $\psi_k = \dfrac{k!}{(2k+1)!} \dfrac{1}{C_{2k}^k}$, on a $\det(H_n) = \prod_{k=1}^{n-1} \psi_k$.

Avec $C_{2k}^k \geq 1$ on déduit que $\psi_k \leq 1$ et :

$$\begin{aligned} \det(H_n) \leq \psi_{n-1} &= \frac{(n-1)!}{(2n-1)!} \frac{1}{C_{2(n-1)}^{n-1}} \\ &\leq \frac{(n-1)!}{(2n-1)!} = \frac{1}{(2n-1)\ldots n} \leq \frac{1}{n^n}. \end{aligned}$$

C'est-à-dire que $\det(H_n)$ tend vers 0 très vite quand n tend vers l'infini.

Exercice 5.3 : *Montrer que pour tout $r \in \{0, 1, \ldots, n-1\}$ l'ensemble $\mathcal{A}_r$ des matrices de $\mathcal{M}_n(\mathbb{R})$ de rang r est connexe par arcs.*

Solution : Pour $r = 0$, le résultat est clair. Pour A de rang $r \geq 1$, il existe P et Q dans $GL_n(\mathbb{R})$ telles que $A = P \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix} Q = PA_rQ$. En fait, pour $r < n$, on peut prendre P et Q dans $GL_n^+(\mathbb{R})$. En effet, si P est dans $GL_n^-(\mathbb{R})$, alors la matrice $P \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & 1 & 0 \\ 0 & \dots & 0 & -1 \end{pmatrix} = PJ$ est dans $GL_n^+(\mathbb{R})$ et en remarquant que, pour $r < n$, $JA_r = A_r$, on a $A = PJA_rQ$. De même pour Q.

On déduit alors la connexité de $\mathcal{A}_r$ de celle de $GL_n^+(\mathbb{R})$ comme dans l'exercice 4.10.

Exercice 5.4 : *Soit*

$$A = \begin{pmatrix} a_1 & c_1 & 0 & \dots & 0 \\ b_2 & a_2 & c_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{n-1} & a_{n-1} & c_{n-1} \\ 0 & \dots & 0 & b_n & a_n \end{pmatrix},$$

une matrice tridiagonale à coefficients réels ou complexes dont tous les déterminants principaux sont non nuls.

Donner un algorithme de calcul du déterminant de la matrice A.

Décrire un algorithme de résolution du système linéaire $Ax = e$ qui utilise la méthode de Gauss sans échange de lignes.

Solution : En notant D_k le déterminant principal d'ordre k d'une matrice tridiagonale d'ordre n et en développant D_k suivant la dernière ligne on a :

$$D_k = a_k D_{k-1} - b_k c_{k-1} D_{k-2}.$$

Ce qui donne, avec les valeurs initiales $D_0 = 1$ et $D_1 = a_1$, un algorithme de calcul très simple.

Si tous les déterminants principaux de la matrice A sont non nuls on sait alors que la méthode de Gauss peut s'effectuer sans permutation de lignes (une telle permutation ferait perdre le caractère tridiagonal), et qu'on n'aura à chaque étape qu'une ligne à traiter et pour chaque ligne seulement deux opérations.

À l'étape k de la méthode, la ligne L_k du système devient :

$$L_k - m_k L_{k-1} \quad (k = 2, \dots, n),$$

avec $m_k = \dfrac{b_k}{a_{k-1}}\cdot$ Ce qui donne les formules de transformations :

$$\begin{cases} b_k = 0, \\ a_k = a_k - m_k c_{k-1}, \\ c_k \text{ inchangé}, \\ e_k = e_k - m_k e_{k-1}. \end{cases}$$

Le système triangulaire supérieur obtenu sera bidiagonal et aura pour solution :

$$\begin{cases} x_n = \dfrac{e_n}{a_n}, \\ x_i = \dfrac{(e_i - c_i x_{i+1})}{a_i} \quad (i = n-1, \dots, 1). \end{cases}$$

Cette méthode de résolution d'un système tridiagonal est la méthode de double balayage de Cholesky.

Le produit des a_i donnera, en fin d'opération, le déterminant de la matrice A.

Exercice 5.5 : *On reprend les notations de l'exercice 5.4.*

Décrire un algorithme de résolution du système $Ax = e$ qui utilise la décomposition LR.

Décrire un algorithme de calcul de l'inverse de A qui utilise la décomposition LR.

Solution : Si tous les déterminants principaux de la matrice A sont non nuls on sait alors que cette matrice admet une décomposition LR avec L triangulaire inférieure à diagonale unité et R triangulaire supérieure.

On cherche les matrices L et R sous la forme bidiagonale, soit :

$$L = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ L_2 & 1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & L_{n-1} & 1 & 0 \\ 0 & \dots & 0 & L_n & 1 \end{pmatrix}, \quad R = \begin{pmatrix} d_1 & r_1 & 0 & \dots & 0 \\ 0 & d_2 & r_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & d_{n-1} & r_{n-1} \\ 0 & \dots & 0 & 0 & d_n \end{pmatrix}.$$

On effectuant le produit $A = LR$ et en identifiant avec les coefficients de A, on obtient alors :

$$\begin{cases} d_1 = a_1, \\ r_j = c_j \quad (j = 1, \dots, n-1), \end{cases}$$

puis :

$$\begin{cases} L_{j+1} = \dfrac{b_{j+1}}{d_j}, \\ d_{j+1} = a_{j+1} - r_j L_{j+1}, \end{cases} \qquad (j = 1, \dots, n-1).$$

On en déduit alors une méthode de résolution d'un système tridiagonal utilisant la décomposition LR. Soit donc à résoudre le système linéaire $Ax = e$.

Tout d'abord, on résout le système linéaire $Ly = e$, ce qui donne :

$$\begin{cases} y_1 = e_1, \\ y_j = e_j - L_j y_{j-1} \quad (j = 2, \ldots, n) . \end{cases}$$

Puis on résout le système linéaire $Rx = y$, ce qui donne :

$$\begin{cases} x_n = \dfrac{y_n}{d_n}, \\ x_j = \dfrac{(y_j - r_j x_{j+1})}{d_j} \quad (j = n-1, \ldots, 1) . \end{cases}$$

On a $A^{-1} = R^{-1}L^{-1}$. En utilisant les résultats du paragraphe 3, page 184, on sait que les coefficients de $U = R^{-1}$ et $V = L^{-1}$ sont donnés, pour j compris entre 1 et n, par :

$$\begin{cases} u_{jj} = \dfrac{1}{d_j}, \\ u_{ij} = -\dfrac{r_i u_{i+1,j}}{d_i} \quad (i = j-1, \ldots, 1), \\ u_{ij} = 0 \quad (i = j+1, \ldots, n), \end{cases} \qquad \begin{cases} v_{ij} = 0 \quad (i = 1, \ldots, j-1), \\ u_{jj} = 1, \\ v_{ij} = -L_i u_{i-1,j} \quad (i = j+1, \ldots, n) . \end{cases}$$

Les coefficients de $W = A^{-1}$ sont alors donnés par :

$$w_{ij} = \sum_{k=\max(i,j)}^{n} u_{ik} v_{kj} \quad (1 \leq i, j \leq n) .$$

Exercice 5.6 : *On note A_n la matrice symétrique réelle d'ordre $n \geq 2$ définie par $A_n = \big((a_{ij})\big)_{1 \leq i,j \leq n}$, avec :*

$$\begin{cases} a_{ij} = n - j + 1 & \textit{pour } 1 \leq i \leq j \leq n, \\ a_{ij} = a_{ji} & \textit{pour } 1 \leq j < i \leq n. \end{cases}$$

1. *Calculer* $\det(A_n)$.

2. *Calculer l'inverse B_n de la matrice A_n.*

3. *Donner la décomposition LR de A_n.*

Solution :

1. On a :

$$A_n = \begin{pmatrix} n & n-1 & \dots & 2 & 1 \\ n-1 & n-1 & \dots & 2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 2 & 2 & \dots & 2 & 1 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix}.$$

En retranchant la deuxième ligne à la première, on déduit que :

$$\forall n \geq 2, \quad \det(A_n) = \det(A_{n-1})$$

et par récurrence :

$$\forall n \geq 2, \quad \det(A_n) = 1.$$

2. La matrice A_n est inversible dans $M_n(\mathbb{Z})$ (son déterminant vaut 1), c'est-à-dire que les coefficients de A_n^{-1} sont entiers. Après quelques expériences numériques on fait l'hypothèse que :

$$A_n^{-1} = B_n = \begin{pmatrix} 1 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \dots & 0 & -1 & 2 \end{pmatrix}.$$

Pour vérifier cette hypothèse, il suffit de vérifier que la colonne numéro j de B_n est solution du système linéaire $A_n x = e_j$ où e_j est le j-ième vecteur de base canonique de $\mathbb{R}^n$.

3. Si $A_n = LR$, alors $A_n^{-1} = R^{-1}L^{-1}$ avec $L' = L^{-1}$ triangulaire inférieure à diagonale unité et $R' = R^{-1}$ triangulaire supérieure. La matrice B_n étant tridiagonale, on cherche les matrices L' et R' sous la forme :

$$L' = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ L'_2 & 1 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & L'_{n-1} & 1 & 0 \\ 0 & \dots & 0 & L'_n & 1 \end{pmatrix}, \quad R' = \begin{pmatrix} d'_1 & r'_1 & 0 & \dots & 0 \\ 0 & d'_2 & r'_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 0 & d'_{n-1} & r'_{n-1} \\ 0 & \dots & 0 & 0 & d'_n \end{pmatrix}.$$

Par identification des coefficients dans le produit $A_n^{-1} = R^{-1}L^{-1}$, on obtient :

$$\begin{cases} r'_i = -1, \quad (i = 1, \dots, n-1), \\ d'_1 = \dfrac{1}{n}, \\ d'_i = -\dfrac{1}{L'_i} = \dfrac{n+2-i}{n+1-i}, \quad (i = 2, \dots, n). \end{cases}$$

On vérifie alors que l'inverse de la matrice R' est la matrice :

$$R = \begin{pmatrix} n & n-1 & \dots & 2 & 1 \\ 0 & \dfrac{n-1}{n} & \dots & \dfrac{2}{n} & \dfrac{1}{n} \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & 0 & \dfrac{2}{3} & \dfrac{1}{3} \\ 0 & \dots & 0 & 0 & \dfrac{1}{2} \end{pmatrix}$$

et que l'inverse de la matrice L' est la matrice :

$$L = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \dfrac{n-1}{n} & 1 & 0 & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \dfrac{2}{n} & \dfrac{2}{n-1} & \dots & 1 & 0 \\ \dfrac{1}{n} & \dfrac{1}{n-1} & \dots & \dfrac{1}{2} & 1 \end{pmatrix}.$$

Exercice 5.7 : *Soient α, β dans $\mathbb{K}$ avec $\beta \neq 0$ et $A(\alpha,\beta) = \left((a_{ij})\right)_{1 \le i,j \le n}$ la matrice d'ordre n supérieur ou égal à 3 définie par :*

$$\forall i \in \{1, 2, \dots, n\}, \quad \begin{cases} a_{ii} = \beta, \\ a_{ij} = \alpha \quad \text{si } j \in \{1, 2, \dots, n\} - \{i\}. \end{cases}$$

1. Calculer le rayon spectral de la matrice J intervenant dans la méthode de Jacobi (on peut utiliser les résultats de l'exercice 4.18).

2. Donner une condition nécessaire et suffisante sur la matrice $A(\alpha,\beta)$ pour que la méthode de Jacobi converge.

3. Montrer que si la méthode de Jacobi converge alors celle de Gauss-Seidel converge aussi.

Solution : On pose $D = \beta I_n$ et $N = D - A(\alpha,\beta)$. La matrice intervenant dans la méthode de Jacobi est alors donnée par :

$$J = D^{-1}N = A\left(-\frac{\alpha}{\beta}, 0\right).$$

1. Le rayon spectral de J est donné par :

$$\rho(J) = \rho\left(A\left(-\frac{\alpha}{\beta}, 0\right)\right) = (n-1)\frac{|\alpha|}{|\beta|}$$

(exercice 4.18).

2. La méthode de Jacobi est convergente si et seulement si le rayon spectral de J est strictement inférieur à 1, ce qui équivaut à $|\beta| > (n-1)|\alpha|$ encore équivalent à dire que la matrice $A(\alpha,\beta)$ est à diagonale strictement dominante.

3. Si la méthode de Jacobi converge alors $A(\alpha,\beta)$ est à diagonale strictement dominante et la méthode de Gauss-Seidel converge aussi.

Exercice 5.8 : *On garde les notations de l'exercice 5.7 et, en utilisant les notations de l'exercice 2.9 (question 2), on pose $M = M(\alpha,\beta,0)$, $N = M - A(\alpha,\beta)$, et $G = M^{-1}N$ désigne la matrice qui intervient dans la méthode de Gauss-Seidel.*

1. Calculer le polynôme caractéristique de G.

2. Montrer que 0 est valeur propre de G et calculer le produit des valeurs propres non nulles dans le cas où α est non nul.

3. Montrer que si $|\alpha| \geq |\beta|$ alors la méthode de Gauss-Seidel ne converge pas.

Solution :

1. Le polynôme caractéristique de G est :

$$P_G(\lambda) = \det\left(M^{-1}N - \lambda I_n\right) = \det\left(M^{-1}\right)(-1)^n \det(M\lambda - N).$$

Soit, en utilisant les notations et les résultats de l'exercice 2.9 :

$$P_G(\lambda) = \frac{(-1)^n}{\beta^n}\det\big(M(\lambda\alpha,\lambda\beta,\alpha)\big) = \frac{(-1)^n}{\beta^n}\frac{(\beta-\alpha)^n\lambda^n - \lambda(\lambda\beta-\alpha)^n}{1-\lambda}.$$

Ce qui peut aussi s'écrire :

$$P_G(\lambda) = (-1)^n \frac{(\lambda-1)\left(\lambda-\frac{\alpha}{\beta}\right)^n + \left(\lambda-\frac{\alpha}{\beta}\right)^n - \left(1-\frac{\alpha}{\beta}\right)^n\lambda^n}{\lambda-1},$$

soit :

$$P_G(\lambda) = (-1)^n\left\{\left(\lambda-\frac{\alpha}{\beta}\right)^n + \frac{\alpha}{\beta}\sum_{k=1}^{n}\left(\lambda-\frac{\alpha}{\beta}\right)^{n-k}\left(1-\frac{\alpha}{\beta}\right)^{k-1}\lambda^{k-1}\right\}.$$

2. On a :

$$P_G(0) = 0, \quad P'_G(0) = \lim_{\lambda \to 0} \frac{P_G(\lambda)}{\lambda} = -\frac{\alpha^n}{\beta^n} \cdot$$

Si α est non nul il en résulte que 0 est valeur propre simple de G et, en écrivant P_G sous la forme :

$$P_G(\lambda) = (-1)^n \lambda \prod_{i=2}^{n} (\lambda - \lambda_i),$$

on déduit que le produit des valeurs propres non nulles de G est donné par :

$$\prod_{i=2}^{n} \lambda_i = \frac{\alpha^n}{\beta^n} \cdot$$

3. Si $|\alpha| \geq |\beta|$ alors $\left| \prod_{i=2}^{n} \lambda_i \right| \geq 1$, donc $\rho(G) \geq 1$ et la méthode de Gauss-Seidel ne converge pas.

Exercice 5.9 : *Soit :*

$$A = \begin{pmatrix} a_1 & c_1 & 0 & \dots & 0 \\ b_2 & a_2 & c_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{n-1} & a_{n-1} & c_{n-1} \\ 0 & \dots & 0 & b_n & a_n \end{pmatrix}$$

une matrice tridiagonale à coefficients réels ou complexes d'ordre $n \geq 3$, les coefficients a_i étant tous non nuls.

1. Montrer que le polynôme caractéristique de la matrice J qui intervient dans la méthode de Jacobi s'écrit $P_J(\lambda) = P(\lambda^2)\lambda^q$ où P est un polynôme tel que $P(0)$ soit non nul.

2. Montrer que le polynôme caractéristique de la matrice G qui intervient dans la méthode de Gauss-Seidel s'écrit $P_G(\lambda^2) = \lambda^n P_J(\lambda)$.

3. Montrer que $\rho(G) = (\rho(J))^2$ et $R_\infty(G) = 2R_\infty(J)$ et conclure.

Solution : On désigne par D la partie diagonale de A, par E sa partie triangulaire inférieure stricte et par F sa partie triangulaire supérieure stricte, de sorte que la matrice A s'écrit $A = E + D + F$. La matrice qui intervient dans la méthode de Jacobi est alors donnée par $J = -D^{-1}(E + F)$ et celle qui intervient dans la méthode de Gauss-Seidel par $G = -(D + E)^{-1} F$.

1. Le polynôme caractéristique de J est défini par :

$$\begin{aligned} P_J(\lambda) &= \det\left(-D^{-1}(E + F) - \lambda I_n\right) \\ &= (-1)^n \det\left(D^{-1}\right) \det(E + F + \lambda D), \end{aligned}$$

la matrice $M = E + F + \lambda D$ étant tridiagonale. Avec les notations du lemme 5.4, on a $M(-1) = -E - F + \lambda D$ et :

$$\det(M) = \det\left(M(-1)\right) = (-1)^n \det(E + F - \lambda D).$$

On déduit donc que $P_J(-\lambda) = (-1)^n P_J(\lambda)$ et on peut écrire $P_J(\lambda) = P\left(\lambda^2\right)\lambda^q$, où P est un polynôme tel que $P(0)$ soit non nul.

2. Le polynôme caractéristique de G est défini par :

$$\begin{aligned} P_G(\lambda) &= \det\left(-(D + E)^{-1} F - \lambda I_n\right) \\ &= (-1)^n \det\left((D + E)^{-1}\right) \det\left(F + \lambda(D + E)\right), \end{aligned}$$

avec $M = F + \lambda(D + E)$ tridiagonale. Pour $\lambda > 0$, avec les notations du lemme 5.4, on a :

$$M\left(\frac{1}{\sqrt{\lambda}}\right) = \sqrt{\lambda}(E + F) + \lambda D = \sqrt{\lambda}\left(E + F + \sqrt{\lambda} D\right)$$

et

$$\det(M) = \det\left(M\left(\frac{1}{\sqrt{\lambda}}\right)\right) = \left(\sqrt{\lambda}\right)^n \det\left(E + F + \sqrt{\lambda} D\right).$$

Donc :

$$\begin{aligned} P_G\left(\lambda^2\right) &= \frac{(-1)^n}{\det(D + E)} \lambda^n \det(E + F + \lambda D) \\ &= \frac{\lambda^n}{\det(D + E)} P_J(\lambda) \det(D) = \lambda^n P_J(\lambda). \end{aligned}$$

3. C'est une conséquence immédiate de ce qui précède.

La méthode de Jacobi converge si et seulement si $\rho(J) < 1$, ce qui équivaut à $\rho(G) < 1$ donc à la convergence de la méthode de Gauss-Seidel. En conclusion, les méthodes de Jacobi et Gauss-Seidel convergent ou divergent simultanément et dans le cas de la convergence c'est la méthode de Gauss-Seidel qui est la plus rapide.

Exercice 5.10 : *Soit A une matrice réelle d'ordre $n \geq 2$ telle qu'il existe un réel α strictement positif avec :*

$$\forall x \in \mathbb{R}^n, \quad \langle Ax \mid x \rangle \geq \alpha \|x\|_2^2 . \tag{5.15}$$

1. Montrer que pour tout vecteur b dans $\mathbb{R}^n$ le système linéaire $Ax = b$ admet une unique solution $u \in \mathbb{R}^n$.

2. Soient θ un réel et $(x_k)_{k\in\mathbb{N}}$ la suite de vecteurs définie par :

$$\begin{cases} x_0 \in \mathbb{R}^n, \\ \forall k \in \mathbb{N}, \quad x_{k+1} = x_k - \theta\,(Ax_k - b) . \end{cases} \tag{5.16}$$

Montrer que pour $\theta \in \left]0, \dfrac{2\alpha}{\|A\|_2^2}\right[$ la suite $(x_k)_{k\in\mathbb{N}}$ converge vers u.

Solution :

1. Avec l'inégalité (5.15), on déduit immédiatement que l'équation $Ax = 0$ équivaut à $x = 0$, ce qui équivaut à dire que la matrice A est inversible. En conséquence pour tout vecteur b dans $\mathbb{R}^n$ le système linéaire $Ax = b$ admet une unique solution $u \in \mathbb{R}^n$.

2. On note $r_k = Ax_k - b$ le vecteur résidu d'ordre k. La suite $(x_k)_{k\in\mathbb{N}}$ converge vers u si et seulement si la suite $(r_k)_{k\in\mathbb{N}}$ converge vers le vecteur nul. En remarquant que :

$$r_{k+1} = r_k - \theta A r_k = (I_n - \theta A)\, r_k,$$

on déduit que $r_k = (I_n - \theta A)^r\, r_0$ pour tout entier naturel k et la suite $(r_k)_{k\in\mathbb{N}}$ converge vers le vecteur nul quelle que soit la valeur initiale r_0 si et seulement si $\rho\,(I_n - \theta A)$ est strictement inférieur à 1. Avec $\rho\,(I_n - \theta A) \leq \|I_n - \theta A\|_2$, on déduit que si $\|I_n - \theta A\|_2 < 1$ alors la suite $(x_k)_{k\in\mathbb{N}}$ converge vers u.

Pour $x \in \mathbb{R}^n$ tel que $\|x\|_2 = 1$, on a :

$$\begin{aligned} \|(I_n - \theta A)\,x\|_2^2 &= \|Ax\|_2^2\,\theta^2 - 2\,\langle Ax \mid x\rangle\,\theta + \|x\|_2^2 \\ &\leq \|A\|_2^2\,\|x\|_2^2\,\theta^2 - 2\alpha\,\|x\|_2^2\,\theta + \|x\|_2^2 = \|A\|_2^2\,\theta^2 - 2\alpha\theta + 1 \end{aligned}$$

et pour $\|A\|_2^2\,\theta^2 - 2\alpha\theta + 1 < 1$, ce qui équivaut à $\theta < \dfrac{2\alpha}{\|A\|_2^2}$ avec $\theta > 0$, on a $\|I_n - \theta A\|_2 < 1$. En conséquence $\rho\,(I_n - \theta A) < 1$ et la suite $(x_k)_{k\in\mathbb{N}}$ converge vers u.

Exercice 5.11 : *Soient A une matrice réelle d'ordre $n \geq 2$ symétrique définie positive de valeurs propres rangées dans l'ordre croissant :*

$$0 < \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$$

et b un vecteur dans $\mathbb{R}^n$.

1. Montrer que la matrice A vérifie la condition (5.15) de l'exercice précédent.

2. Montrer que la suite de vecteurs $(x_k)_{k\in\mathbb{N}}$ définie par (5.16) dans l'exercice précédent converge vers la solution u du système linéaire $Ax = b$ quelle que soit la valeur initiale x_0 si et seulement si $\theta \in \left]0, \dfrac{2}{\lambda_n}\right[$.

3. Montrer que, pour x_0 donné, le meilleur choix de $\theta \in \left]0, \dfrac{2}{\lambda_n}\right[$ (i.e. celui qui assure la convergence la plus rapide) est $\theta_0 = \dfrac{2}{\lambda_1 + \lambda_n}$ et calculer le rayon spectral ρ_0 de la matrice $I_n - \theta_0 A$.

4. Calculer les valeurs de θ_0 et ρ_0 pour la matrice :

$$A = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & -1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & -1 & 2 & -1 \\ 0 & \cdots & 0 & -1 & 2 \end{pmatrix}.$$

Solution :

1. Voir le lemme 5.10.

2. Dans l'exercice précédent on a vu que la suite $(x_k)_{k\in\mathbb{N}}$ converge vers u quelle que soit la valeur initiale x_0 si et seulement si $\rho(I_n - \theta A)$ est strictement inférieur à 1, ce qui équivaut à dire que pour tout entier j compris entre 1 et n on a :

$$-1 < 1 - \theta\lambda_j < 1,$$

ce qui équivaut à $\theta > 0$ et $\theta < \dfrac{2}{\lambda_n}$, soit à $\theta \in \left]0, \dfrac{2}{\lambda_n}\right[$.

3. Le meilleur choix de θ est celui qui minimise le rayon spectral :

$$\rho(I_n - \theta A) = \max_{1\leq j\leq n} |1 - \theta\lambda_j| = \max(|1 - \theta\lambda_1|, |1 - \theta\lambda_n|).$$

En dessinant le graphe de la fonction :

$$\theta \mapsto \max(|1 - \theta\lambda_1|, |1 - \theta\lambda_n|),$$

on voit facilement que cette fonction atteint son minimum en $\theta_0 = \dfrac{2}{\lambda_1 + \lambda_n}$ avec :

$$\begin{aligned}\rho_0 = \rho\left(I_n - \theta_0 A\right) &= \max_{1\leq j\leq n}\left|1 - \frac{2}{\lambda_1+\lambda_n}\lambda_j\right| \\ &= \max_{1\leq j\leq n}\frac{(\lambda_n-\lambda_j)-(\lambda_j-\lambda_1)}{\lambda_1+\lambda_n} = \frac{\lambda_n-\lambda_1}{\lambda_1+\lambda_n}\,.\end{aligned}$$

Ce qui s'écrit aussi :

$$\rho_0 = \frac{1-\dfrac{\lambda_1}{\lambda_n}}{1+\dfrac{\lambda_1}{\lambda_n}} = \frac{1-\mathrm{cond}_2\left(A\right)}{1+\mathrm{cond}_2\left(A\right)}\,.$$

La méthode (méthode de descente à paramètre constant de Richardson) sera donc rapidement convergente pour une matrice bien conditionnée, c'est-à-dire telle que $\mathrm{cond}_2\left(A\right)$ est voisin de 1.

4. Les valeurs propres de la matrice A sont données par :

$$\lambda_k\left(a,b\right) = 2-2\cos\left(\frac{k\pi}{n+1}\right) = 4\sin^2\left(\frac{k\pi}{2\left(n+1\right)}\right) \quad \left(1\leq k\leq n\right)$$

(exercice 2.10). On a donc :

$$\theta_0 = \frac{1}{2\left(\sin^2\left(\dfrac{\pi}{2\left(n+1\right)}\right)+\sin^2\left(\dfrac{n\pi}{2\left(n+1\right)}\right)\right)},$$

$$\rho_0 = \frac{\sin^2\left(\dfrac{n\pi}{2\left(n+1\right)}\right)-\sin^2\left(\dfrac{\pi}{2\left(n+1\right)}\right)}{\sin^2\left(\dfrac{n\pi}{2\left(n+1\right)}\right)+\sin^2\left(\dfrac{\pi}{2\left(n+1\right)}\right)}\,.$$

6 Calcul approché des valeurs et vecteurs propres

1. Introduction

Le problème de la détermination des valeurs propres d'une matrice est en général beaucoup plus difficile que celui de la résolution des systèmes linéaires.

Tout d'abord nous allons étudier la méthode de la puissance itérée qui permet de calculer la valeur propre de plus grand module d'une matrice réelle sous certaines hypothèses. Puis en itérant ce procédé on peut en déduire les autres quand elles sont toutes distinctes en module. Cette méthode, peu performante, est à utiliser pour les matrices ayant toutes leurs valeurs propres distinctes ou si on en cherche seulement à calculer quelques-unes d'entre elles.

D'autres méthodes sont les méthodes de Rutishauser, de Givens et Householder et de Jacobi, qui reposent sur le principe suivant : on construit une suite $(A_k)_{k\in\mathbb{N}}$ de matrices toutes semblables à la matrice A, donc ayant les mêmes valeurs propres, qui converge vers une matrice plus simple (diagonale pour Jacobi, tridiagonale pour Givens et Householder, triangulaire pour Rutishauser) encore semblable à la matrice A. Les valeurs propres cherchées sont alors les termes diagonaux de la matrice limite, dans le cas des méthodes de Jacobi et de Rutishauser.

Les problèmes posés par ces méthodes sont :

• savoir décrire le passage de A_k à A_{k+1} aussi simplement que possible, c'est-à-dire avec un coût de calcul peu élevé ;

• les valeurs propres doivent être bien conservées (stabilité numérique de la méthode).

Comme pour le chapitre 5, on pourra consulter [16] ou [5] pour la programmation des méthodes décrites dans ce chapitre.

2. Méthode de la puissance itérée

La méthode de la puissance itérée permet le calcul du rayon spectral de la valeur propre de plus grand module d'une matrice réelle quand cette valeur propre dominante est unique.

Par exemple, le théorème de Perron-Frobenius (théorème 4.17) nous dit qu'une matrice réelle à coefficients strictement positifs a une valeur propre dominante unique.

Théorème 6.1 : *Soit A une matrice d'ordre $n \geq 2$ à coefficients réels telle que la valeur propre λ_1 de module maximal soit unique. Cette valeur propre est alors réelle et simple, l'espace propre associé est une droite vectorielle et*

$$\mathbb{R}^n = \ker(A - \lambda_1 I_n) \oplus \operatorname{Im}(A - \lambda_1 I_n),$$

le sous-espace vectoriel $\operatorname{Im}(A - \lambda_1 I_n)$ *étant stable par A.*

Démonstration : On note $\lambda_1, \lambda_2, \ldots, \lambda_n$ les valeurs propres de A dans $\mathbb{C}$, avec :

$$\rho(A) = |\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \ldots \geq |\lambda_n|.$$

La valeur propre λ_1 est nécessairement réelle et simple (la matrice A étant réelle, son spectre est stable par conjugaison complexe et λ_1 est l'unique valeur propre de module maximal).

On note :

$$E_1 = \ker(A - \lambda_1 I_n), \quad F_1 = \operatorname{Im}(A - \lambda_1 I_n).$$

La valeur propre λ_1 étant simple, l'espace propre associé E_1 est nécessairement de dimension 1.

De plus, il est facile de vérifier que les sous-espaces vectoriels E_1 et F_1 sont stables par A.

Pour montrer que E_1 et F_1 sont en somme directe, il suffit de montrer que leur intersection se réduit au vecteur nul (théorème du rang).

Si $y \in E_1 \cap F_1$, il existe alors un vecteur $x \in \mathbb{R}^n$ tel que $y = (A - \lambda_1 I_n) x$ et $(A - \lambda_1 I_n) y = 0$. Le vecteur x est alors dans le noyau de $(A - \lambda_1 I_n)^2$.

Les valeurs propres de $A_1 = A - \lambda_1 I_n$ sont $0, \lambda_2 - \lambda_1, \ldots, \lambda_2 - \lambda_n$, la valeur propre nulle étant simple. La matrice A_1^2 a donc pour valeurs propres $0, (\lambda_2 - \lambda_1)^2, \ldots, (\lambda_2 - \lambda_n)^2$, la valeur propre nulle étant simple. On déduit alors que $\ker(A_1^2)$ est de dimension 1 et, avec $\{0\} \neq \ker(A_1) \subset \ker(A_1^2)$, que $\ker(A_1) = \ker(A_1^2)$. Il en résulte $x \in \ker(A_1)$ et $y = 0$. On a donc $E_1 \cap F_1 = \{0\}$ et $\mathbb{R}^n = E_1 \oplus F_1$. ■

Remarque 6.1 *: De manière plus générale, on peut montrer que si λ est une valeur propre de A d'ordre $p \geq 1$, alors les sous-espaces vectoriels* $\ker(A - \lambda_1 I_n)^p$ *et* $\operatorname{Im}(A - \lambda_1 I_n)^p$ *sont stables par A, supplémentaires dans $\mathbb{R}^n$ et* $\dim(E_p) = p$ *(exercice 6.1).*

Dans ce qui suit, on munit $\mathbb{R}^n$ d'une norme quelconque et A désigne une matrice d'ordre $n \geq 2$ à coefficients réels telle que la valeur propre λ_1 de module maximal soit unique. En particulier cette valeur propre est unique.

On note :

$$E_1 = \ker(A - \lambda_1 I_n), \quad F_1 = \operatorname{Im}(A - \lambda_1 I_n),$$

et on définit la suite $\left(x^{(k)}\right)_{k \in \mathbb{N}}$ de vecteurs de $\mathbb{R}^n$ par :

$$\begin{cases} x^{(0)} = e_1 + f_1 & \text{avec } e_1 \in E_1 - \{0\}, \quad f_1 \in F_1, \\ \forall k \in \mathbb{N}, \quad x^{(k+1)} = \dfrac{1}{\left\|Ax^{(k)}\right\|} Ax^{(k)}. \end{cases}$$

On verra au cours de la démonstration du théorème 6.2 que cette suite est bien définie.

On note, dans la base canonique de $\mathbb{R}^n$, $e_{1,j}$ les composantes du vecteur e_1, $x_j^{(k)}$ celles de $x^{(k)}$ et $\left(Ax^{(k)}\right)_j$ celles de $Ax^{(k)}$.

On a alors le résultat suivant qui permet de calculer une valeur approchée de la valeur propre de plus grand module et d'un vecteur propre associé.

Théorème 6.2 : *Avec les notations ci-dessus, on a :*

(i) $\lim\limits_{k\to+\infty} \left\|Ax^{(k)}\right\| = |\lambda_1| = \rho(A)$.

(ii) $\lim\limits_{k\to+\infty} x^{(2k)} = v_1,\ \lim\limits_{k\to+\infty} x^{(2k+1)} = v_2 = \operatorname{signe}(\lambda_1)\, v_1$ *où v_1 est un vecteur propre non nul associé à λ_1.*

(iii) Pour tout $j \in \{1, 2, \ldots, n\}$ tel que $e_{1,j} \neq 0$, on a $\lim\limits_{k\to+\infty} \dfrac{\left(Ax^{(k)}\right)_j}{x_j^{(k)}} = \lambda_1$.

Démonstration : On identifie une matrice à l'endomorphisme de $\mathbb{R}^n$ qu'elle définit et on note B la restriction de A à F_1. D'après le théorème 6.1, B est un endomorphisme de F_1 de valeurs propres $\lambda_2, \ldots, \lambda_n$.

On a $Ax^{(0)} = \lambda_1 e_1 + Bf_1 \neq 0$, donc $x^{(1)}$ est bien défini. Par récurrence, on voit que $x^{(k)}$ est bien défini avec une projection non nulle sur E_1. De manière plus précise, on a :

$$\forall k \in \mathbb{N}, \begin{cases} A^k x^{(0)} = \lambda_1^k e_1 + B^k f \neq 0, \\ x^{(k)} = \dfrac{1}{\left\|A^k x^{(0)}\right\|} A^k x^{(0)}. \end{cases}$$

En effet, le résultat est vrai pour $k = 1$ et, en le supposant vrai pour $k \geq 1$, on a :

$$A^{k+1} x^{(0)} = \lambda_1^{k+1} e_1 + B^{k+1} f \neq 0$$

et

$$Ax^{(k)} = \frac{1}{\left\|A^k x^{(0)}\right\|} A^{k+1} x^{(0)} \neq 0,$$

de sorte que :

$$x^{(k+1)} = \frac{\left\|A^k x^{(0)}\right\|}{\left\|A^{k+1} x^{(0)}\right\|} \frac{1}{\left\|A^k x^{(0)}\right\|} A^{k+1} x^{(0)} = \frac{1}{\left\|A^{k+1} x^{(0)}\right\|} A^{k+1} x^{(0)}.$$

Pour tout $k \geq 1$, on a :

$$A^k x^{(0)} = \lambda_1^k e_1 + B^k f_1 = \lambda_1^k (e_1 + f_k),$$

avec $\lim\limits_{k\to+\infty} f_k = \lim\limits_{k\to+\infty} \left(\left(\dfrac{1}{\lambda_1} B\right)^k f_1\right) = 0$ du fait que $\rho\left(\dfrac{1}{\lambda_1} B\right) < 1$.

En écrivant :

$$x^{(k)} = \left(\frac{\lambda_1}{|\lambda_1|}\right)^k \frac{1}{\|e_1 + f_k\|}(e_1 + f_k)$$

on déduit que :

$$\left\|Ax^{(k)}\right\| = \frac{1}{\|e_1 + f_k\|}\|\lambda_1 e_1 + Af_k\|$$

et $\lim\limits_{k\to+\infty} \left\|Ax^{(k)}\right\| = |\lambda_1|$. On déduit aussi que :

$$\lim_{k\to+\infty} x^{(2k)} = v_1 = \frac{1}{\|e_1\|}e_1, \quad \lim_{k\to+\infty} x^{(2k+1)} = \frac{\text{signe}(\lambda_1)}{\|e_1\|}e_1.$$

Enfin avec :

$$Ax^{(k)} - \lambda_1 x^{(k)} = \left(\frac{\lambda_1}{|\lambda_1|}\right)^k \frac{1}{\|e_1 + f_k\|}(Af_k - \lambda_1 f_k),$$

on déduit que $\lim\limits_{k\to+\infty}\left(Ax^{(k)} - \lambda_1 x^{(k)}\right) = 0$.

Comme $e_1 \neq 0$, il existe $j \in \{1, 2, \ldots, n\}$ tel que $e_{1,j} \neq 0$ et pour k assez grand on a $x_j^{(k)} \neq 0$. Avec $\lim\limits_{k\to+\infty}\left(\left(Ax^{(k)}\right)_j - \lambda_1 x_j^{(k)}\right) = 0$, on déduit alors que $\lim\limits_{k\to+\infty} \dfrac{\left(Ax^{(k)}\right)_j}{x_j^{(k)}} = \lambda_1$. ■

Remarque 6.2 : *En général on n'a pas d'informations sur le sous-espace propre E_1 de sorte qu'en prenant $x^{(0)}$ quelconque dans $\mathbb{R}^n$ sa projection sur E_1 peut être nulle et théoriquement la méthode décrite ci-dessus ne donnera pas le résultat espéré. En fait F_1 est un hyperplan de $\mathbb{R}^n$, donc de mesure nulle, et en prenant $x^{(0)}$ quelconque dans $\mathbb{R}^n$ on a très peu de chances de tomber sur F_1.*

Si $x^{(0)}$ est dans F_1, à cause des erreurs d'arrondis, au bout de quelques itérations on sort de cet hyperplan, mais la convergence peut être lente. Une autre façon d'éviter ce type de problème est de prendre deux valeurs différentes de $x^{(0)}$ et de garder celle qui donne la convergence la plus rapide.

Remarque 6.3 : *La méthode de la puissance itérée est en fait une adaptation de la méthode de Bernoulli pour calculer la racine de plus grand module d'un polynôme, quand cette dernière est unique.*

Remarque 6.4 : *Si la matrice A est inversible, en appliquant la méthode de la puissance itérée à la matrice A^{-1}, on a un moyen de calculer la valeur propre du plus petit module de A (quand cette dernière est unique). La méthode obtenue est aussi appelée méthode de la puissance inverse. Dans la pratique on évite d'inverser A et on utilise la suite définie par :*

$$\begin{cases} x^{(0)} = e_n + f_n & \text{avec } e_n \in \ker(A - \lambda_n I_n) - \{0\}, \\ & \quad f_n \in \operatorname{Im}(A - \lambda_n I_n), \\ \forall k \in \mathbb{N}, \quad x^{(k+1)} = \dfrac{1}{\left\|u^{(k+1)}\right\|}u^{(k+1)}, & \text{avec } Au^{(k+1)} = x^{(k)}. \end{cases}$$

Le calcul approché des autres valeurs propres peut se faire en utilisant la méthode de déflation dans l'hypothèse où les valeurs propres de A vérifient :

$$|\lambda_1| > |\lambda_2| > |\lambda_3| > \ldots > |\lambda_n|.$$

La matrice A est alors diagonalisable avec des valeurs propres réelles et simples.

Soit e_1 un vecteur propre associé à λ_1 de norme euclidienne égale à 1. On va vérifier que les valeurs propres de la matrice $B = A - \lambda_1 e_1 {}^t e_1 = \left((a_{ij} - \lambda_1 e_{1,i} e_{1,j})\right)$ sont $0, \lambda_2, \ldots, \lambda_n$. On appliquera alors la méthode de la puissance itérée à cette nouvelle matrice pour obtenir une valeur approchée de λ_2, puis on continuera ainsi pour obtenir des valeurs approchées des autres valeurs propres.

Lemme 6.1 : *Avec les notations qui précèdent, si e_1 est un vecteur propre de A associé à λ_1 de norme euclidienne égale à 1 alors les valeurs propres de la matrice $B = A - \lambda_1 e_1 {}^t e_1$ sont $0, \lambda_2, \ldots, \lambda_n$.*

Démonstration : Pour tout $j = 1, 2, \ldots, n$, λ_j est aussi valeur propre de ${}^t A$. On désigne alors par f_j un vecteur propre de ${}^t A$ associé à λ_j.

En écrivant que :

$$\langle e_1 \mid f_j \rangle = \frac{1}{\lambda_1} \langle A e_1 \mid f_j \rangle = \frac{1}{\lambda_1} \langle e_1 \mid {}^t A f_j \rangle = \frac{\lambda_j}{\lambda_1} \langle e_1 \mid f_j \rangle,$$

on déduit que $\langle e_1 \mid f_j \rangle = 0$ pour tout j compris entre 2 et n. On a alors pour tout j compris entre 2 et n :

$${}^t B f_j = {}^t A f_j - \lambda_1 e_1 {}^t e_1 f_j = \lambda_j f_j - \lambda_1 \langle e_1 \mid f_j \rangle e_1 = \lambda_j f_j,$$

c'est-à-dire que λ_j est valeur propre de ${}^t B$. C'est donc aussi une valeur propre de B.

Enfin :

$$B e_1 = A e_1 - \lambda_1 e_1 {}^t e_1 e_1 = \lambda_1 e_1 - \lambda_1 \langle e_1 \mid e_1 \rangle e_1 = 0,$$

c'est-à-dire que 0 est valeur propre de B. ∎

3. Méthode de Jacobi pour les matrices symétriques

On se donne, dans ce paragraphe, une matrice A symétrique réelle. Elle est donc diagonalisable avec ses valeurs propres toutes réelles.

Le principe de la méthode de Jacobi consiste à construire une suite de matrices de rotations planes $\left(R(\theta_k)\right)_{k \in \mathbb{N}}$ telle que pour k tendant vers l'infini, la suite de matrices $(A_k)_{k \in \mathbb{N}}$ définie par :

$$\begin{cases} A_0 = A, \\ A_{k+1} = R(\theta_k)^{-1} A_k R(\theta_k) \quad (k \geq 0), \end{cases}$$

converge vers une matrice diagonale. Comme chacune des matrices A_k est semblable à la matrice A, on déduit que les valeurs propres de A sont les termes diagonaux de la matrice limite.

Cette méthode est bien adaptée aux matrices symétriques de petite taille.

Pour toute matrice $M = \left((m_{ij})\right)_{1\leq i,j\leq n}$ d'ordre $n \geq 2$ et à coefficients réels, on désigne par $\|M\|_s = \mathrm{Tr}\left({}^tMM\right)^{\frac{1}{2}}$ la norme de Schur de M.

On vérifie facilement qu'on a $\|M\|_2 \leq \|M\|_s$, où $\|\cdot\|_2$ désigne la norme matricielle induite par la norme euclidienne et $\|MR\|_s = \|RM\|_s = \|M\|_s$ pour toute matrice orthogonale R.

Pour tout réel $\theta \in [-\pi, \pi[$ et tout couple d'entiers (p,q) tels que $1 \leq p < q \leq n$, on note $R_{p,q}(\theta)$ la matrice de rotation d'angle θ dans le plan défini par les vecteurs e_p et e_q de la base canonique de $\mathbb{R}^n$. Cette matrice s'écrit :

$$R_{p,q}(\theta) = \begin{pmatrix} I_{p-1} & 0 & 0 \\ 0 & \rho_{p,q}(\theta) & 0 \\ 0 & 0 & I_{n-q} \end{pmatrix},$$

où on a posé :

$$\rho_{p,q}(\theta) = \begin{pmatrix} c & 0 & -s \\ 0 & I_{q-p+1} & 0 \\ s & 0 & c \end{pmatrix},$$

avec $c = \cos(\theta)$ et $s = \sin(\theta)$.

Pour toute matrice symétrique réelle d'ordre $n \geq 2$, $M = \left((m_{ij})\right)_{1\leq i,j\leq n}$, on note :

$$M_{p,q}(\theta) = R_{p,q}(\theta)^{-1} M R_{p,q}(\theta).$$

Pour une matrice symétrique M, le produit $M_{p,q}(\theta)$ est facile à calculer. La multiplication à gauche par $R_{p,q}(\theta)^{-1} = {}^tR_{p,q}(\theta)$ modifie seulement les lignes p et q de M et la multiplication à droite par $R_{p,q}(\theta)$ change seulement les colonnes p et q de M. Ce qui donne pour les coefficients de $M_{p,q}(\theta) = \left((m'_{ij})\right)_{1\leq i,j\leq n}$:

$$\begin{cases} m'_{ij} = m_{ij} \quad (i \neq p,\ i \neq q,\ j \neq p,\ j \neq q), \\ m'_{ip} = cm_{ip} + sm_{iq} \quad (i \neq p,\ i \neq q), \\ m'_{pp} = c^2 m_{pp} + s^2 m_{qq} + 2scm_{pq}, \\ m'_{iq} = cm_{iq} - sm_{ip} \quad (i \neq p,\ i \neq q), \\ m'_{qq} = s^2 m_{pp} + c^2 m_{qq} - 2scm_{pq}, \\ m'_{pq} = \left(c^2 - s^2\right) m_{pq} - sc\left(m_{pp} - m_{qq}\right), \end{cases}$$

la matrice $M_{p,q}(\theta)$ étant symétrique.

De plus on a $\|M_{p,q}(\theta)\|_s = \|M\|_s$.

L'idée de la méthode de Jacobi est de déterminer θ de manière à annuler les coefficients m'_{pq} et m'_{qp} de $M_{p,q}(\theta)$.

Lemme 6.2 : *Soit* $(p,q) \in \mathbb{N}^2$ *tel que* $1 \leq p < q \leq n$ *et* $m_{pq} \neq 0$.

(i) *Il existe un unique réel* $\theta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right] - \{0\}$ *tel que* $m'_{pq} = 0$.

(ii) $t = \tan(\theta)$ *est l'unique solution dans* $]-1,1] - \{0\}$ *de l'équation du second degré* $t^2 + 2b_{pq}t - 1 = 0$ *où on a posé* $b_{pq} = \dfrac{m_{pp} - m_{qq}}{2m_{pq}}$.

(iii) $c = \cos(\theta)$ *et* $s = \sin(\theta)$ *peuvent se calculer de manière algébrique avec* $c = \dfrac{1}{\sqrt{1+t^2}}$ *et* $s = tc$.

Démonstration :

(i) La condition $m'_{pq} = 0$ équivaut à :

$$\left(\cos(\theta)^2 - \sin(\theta)^2\right) m_{pq} - \sin(\theta)\cos(\theta)(m_{pp} - m_{qq}) = 0.$$

Si $m_{pq} \neq 0$, alors $\theta \neq 0$. Si $m_{pp} = m_{qq}$, on prend alors $\theta = \frac{\pi}{4}$ et si $m_{pp} \neq m_{qq}$ la condition ci-dessus s'écrit :

$$\tan(2\theta) = \frac{2m_{pq}}{m_{pp} - m_{qq}},$$

et cette équation admet une unique solution $\theta \in \left]-\frac{\pi}{4}, \frac{\pi}{4}\right[- \{0\}$.

(ii) On déduit facilement de (i) que $t = \tan(\theta) \in]-1,1] - \{0\}$ et qu'il est solution de l'équation du second degré $t^2 + 2b_{pq}t - 1 = 0$. Cette équation admet deux racines réelles :

$$\begin{cases} t_1 = -b_{pq} + \sqrt{1 + b_{pq}^2} = \dfrac{1}{b_{pq} + \sqrt{1 + b_{pq}^2}}, \\ t_2 = -\dfrac{1}{t_1} = \dfrac{-1}{\sqrt{1 + b_{pq}^2} - b_{pq}}. \end{cases}$$

La racine de valeur absolue inférieure à 1 est :

$$t = \frac{\operatorname{signe}(b_{pq})}{\sqrt{b_{pq}^2 + 1} + |b_{pq}|},$$

avec la convention $\operatorname{signe}(0) = 1$.

(iii) Avec $\cos(\theta) = \dfrac{1}{\sqrt{1 + \tan^2(\theta)}}$ et $\sin(\theta) = \cos(\theta)\tan(\theta)$, on déduit que $c = \dfrac{1}{\sqrt{1+t^2}}$ et $s = tc$. ∎

Remarque 6.5 : *En remarquant que* $\tau = \tan\left(\dfrac{\theta}{2}\right) = \dfrac{s}{1+c}$, *les formules donnant les coefficients de* $M_{p,q}(\theta)$ *se simplifient pour donner finalement :*

$$\begin{cases} m'_{ij} = m_{ij} \quad (i \neq p,\ i \neq q,\ j \neq p,\ j \neq q), \\ m'_{ip} = m_{ip} + s(m_{iq} - \tau m_{ip}) \quad (i \neq p,\ i \neq q), \\ m'_{pp} = m_{pp} + t m_{pq}, \\ m'_{iq} = m_{iq} - s(m_{ip} + \tau m_{iq}) \quad (i \neq p,\ i \neq q), \\ m'_{qq} = m_{qq} - t m_{pq}, \\ m'_{pq} = 0. \end{cases} \tag{6.1}$$

Dans ce qui suit, A désigne une matrice symétrique réelle d'ordre $n \geq 3$ de valeurs propres (réelles) $\lambda_1, \lambda_2, \ldots, \lambda_n$. On lui associe la suite de matrices symétriques $(A_k)_{k\in\mathbb{N}}$ définie par $A_k = \left(\left(a_{ij}^{(k)}\right)\right)_{1\leq i,j\leq n}$ et

$$\begin{cases} A_0 = A, \\ A_{k+1} = R(\theta_k)^{-1} A_k R(\theta_k) \quad (k \geq 0), \end{cases}$$

où $R(\theta_k)$ est la matrice unité si $a_{pq}^{(k)} = 0$ et une matrice de rotation $R_{p,q}(\theta_k)$ avec $1 \leq p < q \leq n$ tels que :

$$\left|a_{pq}^{(k)}\right| = \max_{1\leq i<j\leq n} \left|a_{ij}^{(k)}\right|$$

et $\theta_k \in \left]-\dfrac{\pi}{4}, \dfrac{\pi}{4}\right] - \{0\}$ tel que $a_{pq}^{(k+1)} = 0$.

On décompose chaque matrice A_k sous la forme $A_k = D_k + E_k$ où D_k est la partie diagonale de A_k.

Lemme 6.3 : *Avec les notations qui précèdent, on a* $\lim\limits_{k\to+\infty} E_k = 0$.

Démonstration : En utilisant les formules simplifiées (6.1), on a :

$$\|D_{k+1}\|_s^2 = \sum_{i=1}^{n} \left|a_{ii}^{(k+1)}\right|^2 = \sum_{\substack{i=1 \\ i\neq p, i\neq q}}^{n} \left|a_{ii}^{(k)}\right|^2 + \left|a_{pp}^{(k+1)}\right|^2 + \left|a_{qq}^{(k+1)}\right|^2,$$

avec :

$$\begin{aligned} \left|a_{pp}^{(k+1)}\right|^2 + \left|a_{qq}^{(k+1)}\right|^2 &= \left|a_{pp}^{(k)} + t_k a_{pq}^{(k)}\right|^2 + \left|a_{qq}^{(k)} - t_k a_{pq}^{(k)}\right|^2 \\ &= \left|a_{pp}^{(k)}\right|^2 + \left|a_{qq}^{(k)}\right|^2 + 2t_k a_{pq}^{(k)} \left(a_{pp}^{(k)} - a_{qq}^{(k)} + t_k a_{pq}^{(k)}\right), \end{aligned}$$

le réel t_k étant l'unique solution dans $]-1,1] - \{0\}$ de l'équation $t_k^2 + 2b_{pq} t_k - 1 = 0$. On a alors :

$$a_{pp}^{(k)} - a_{qq}^{(k)} = \frac{1 - t_k^2}{t_k} a_{pq}^{(k)}$$

et

$$\left|a_{pp}^{(k+1)}\right|^2 + \left|a_{qq}^{(k+1)}\right|^2 = \left|a_{pp}^{(k)}\right|^2 + \left|a_{qq}^{(k)}\right|^2 + 2\left|a_{pq}^{(k)}\right|^2,$$

ce qui donne :

$$\|D_{k+1}\|_s^2 = \|D_k\|_s^2 + 2\left|a_{pq}^{(k)}\right|^2.$$

On peut aussi constater que :

$$\begin{pmatrix} a_{pp}^{(k+1)} & a_{pq}^{(k+1)} \\ a_{qp}^{(k+1)} & a\binom{k+1}{qq} \end{pmatrix} = \begin{pmatrix} \cos(\theta_k) & \sin(\theta_k) \\ -\sin(\theta_k) & \cos(\theta_k) \end{pmatrix} \begin{pmatrix} a_{pp}^{(k)} & a_{pq}^{(k)} \\ a_{qp}^{(k)} & a_{qq}^{(k)} \end{pmatrix} \begin{pmatrix} \cos(\theta_k) & -\sin(\theta_k) \\ \sin(\theta_k) & \cos(\theta_k) \end{pmatrix}$$

et par conservation des normes on a immédiatement :

$$\left|a_{pp}^{(k+1)}\right|^2 + \left|a_{qq}^{(k+1)}\right|^2 + 2\left|a_{pq}^{(k+1)}\right|^2 = \left|a_{pp}^{(k)}\right|^2 + \left|a_{qq}^{(k)}\right|^2 + 2\left|a_{pq}^{(k)}\right|^2.$$

On conclut alors avec $a_{pq}^{(k+1)} = 0$.

D'autre part, on a $\|A_k\|_s^2 = \|D_k\|_s^2 + \|E_k\|_s^2$ avec $\|A_k\|_s = \|A\|_s$ pour tout $k \geq 0$. On déduit alors que :

$$\|E_{k+1}\|_s^2 = \|E_k\|_s^2 - 2\left|a_{pq}^{(k)}\right|^2.$$

En utilisant le fait que a_{pq}^k est de module maximal dans E_k, on déduit que :

$$\|E_k\|_s^2 \leq \left(n^2 - n\right)\left|a_{pq}^{(k)}\right|^2$$

et

$$\|E_{k+1}\|_s^2 \leq \left(1 - \frac{2}{n^2 - n}\right)\|E_k\|_s^2.$$

Par récurrence on a alors :

$$\|E_k\|_s^2 \leq \left(1 - \frac{2}{n^2 - n}\right)^k \|E_0\|_s^2.$$

Il en résulte que $\lim_{k \to +\infty} \|E_k\|_s = 0$ pour $n \geq 3$ et $\lim_{k \to +\infty} E_k = 0$. ■

Pour montrer que la suite des matrices diagonales $(D_k)_{k \in \mathbb{N}}$ est convergente on a besoin du résultat suivant, qui sera aussi utilisé pour calculer les vecteurs propres de la matrice A.

Lemme 6.4 : *Soit $(x_n)_{n \in \mathbb{N}}$ une suite dans un espace vectoriel normé de dimension finie $(E, \|\cdot\|)$. Cette suite converge dans $(E, \|\cdot\|)$ si et seulement si elle est bornée, elle n'admet qu'un nombre fini de valeurs d'adhérence et si $\lim_{n \to +\infty} \|x_{n+1} - x_n\| = 0$.*

Démonstration : Exercice 1.10. ■

Théorème 6.3 : *Avec les notations qui précèdent, la suite $(A_k)_{k\in\mathbb{N}}$ converge vers une matrice diagonale D_σ de termes diagonaux $d_i^\sigma = \lambda_{\sigma(i)}$ pour $i = 1, 2, \ldots, n$, où $\lambda_1, \lambda_2, \ldots, \lambda_n$ sont les valeurs propres de A et σ une permutation de $\{1, 2, \ldots, n\}$.*

Démonstration : D'après le lemme 6.3, il est équivalent de montrer que la suite $(D_k)_{k\in\mathbb{N}}$ converge vers la matrice D_σ.

Avec $\|D_k\|_s \leq \|A_k\|_s = \|A\|_s$, on déduit que la suite $(D_k)_{k\in\mathbb{N}}$ est bornée.

Avec

$$a_{ii}^{(k+1)} - a_{ii}^{(k)} = \begin{cases} 0, & \text{si } i \neq p,\ i \neq q, \\ \tan(\theta_k)\, a_{pq}^{(k)}, & \text{si } i = p, \\ -\tan(\theta_k)\, a_{pq}^{(k)}, & \text{si } i = q, \end{cases}$$

$|\tan(\theta_k)| \leq 1$ et $\left|a_{pq}^{(k)}\right| \leq \|E_k\|_s \underset{k\to+\infty}{\to} 0$, on déduit que $\lim\limits_{k\to+\infty} (D_{k+1} - D_k) = 0$.

Pour montrer que la suite $(D_k)_{k\in\mathbb{N}}$ est convergente il nous reste donc à montrer qu'elle n'a qu'un nombre fini de valeurs d'adhérence (lemme 6.4).

Soit $\left(D_{\varphi(k)}\right)_{k\in\mathbb{N}}$ une suite extraite de $(D_k)_{k\in\mathbb{N}}$ qui converge vers une matrice D diagonale. Avec $\lim\limits_{k\to+\infty} E_{\varphi(k)} = 0$, on déduit que $\lim\limits_{k\to+\infty} A_{\varphi(k)} = D$. Si $P_k(\lambda)$ désigne le polynôme caractéristique de $A_{\varphi(k)}$, on a $\lim\limits_{k\to+\infty} P_k(\lambda) = \det(D - \lambda I_n)$ avec $P_k(\lambda) = \det(A - \lambda I_n)$ puisque A et $A_{\varphi(k)}$ sont semblables. On a donc montré que D et A ont le même polynôme caractéristique, donc les mêmes valeurs propres. La matrice D étant diagonale, les termes diagonaux de D sont donnés par $d_i^\sigma = \lambda_{\sigma(i)}$ pour tout $i = 1, 2, \ldots, n$, où σ est une permutation de $\{1, 2, \ldots, n\}$. Le nombre de ces permutations étant fini, on déduit qu'il n'y a qu'un nombre fini de valeurs d'adhérence. ■

Pour calculer une famille de vecteurs propres de A, on désigne par $(P_k)_{k\in\mathbb{N}}$ la suite de matrices orthogonales définie par $P_k = R(\theta_0)\, R(\theta_1) \ldots R(\theta_k)$ pour tout $k \geq 0$.

On suppose que toutes les valeurs propres de A sont distinctes. On désigne par D la matrice diagonale de termes diagonaux $d_i = \lambda_i$ pour $i = 1, 2, \ldots, n$ et par $P = (C_1, C_2, \ldots, C_n)$ une matrice orthogonale telle que $D = P^{-1}AP$, où C_j désigne la colonne numéro j de P.

Théorème 6.4 : *Avec les notations ci-dessus, on a :*

(i) $\lim\limits_{k\to+\infty} \theta_k = 0$ *et* $\lim\limits_{k\to+\infty} R_k = I_n$.

(ii) La suite $(P_k)_{k\in\mathbb{N}}$ converge vers une matrice $P_\sigma = \left(\pm C_{\sigma(1)}, \pm C_{\sigma(2)}, \ldots, \pm C_{\sigma(n)}\right)$, où σ est une permutation de $\{1, 2, \ldots, n\}$, le vecteur $C_{\sigma(i)}$ étant un vecteur propre associé à la valeur propre $\lambda_{\sigma(i)}$.

Démonstration :

(i) Avec $\lim_{k\to+\infty} D_k = D_\sigma$ et l'hypothèse que les valeurs propres de A sont deux à deux distinctes, on déduit qu'il existe un entier k_0 tel que :

$$\forall k \geq k_0, \quad \left|a_{pq}^{(k)} - a_{qq}^{(k)}\right| > \frac{1}{2} \min_{1\leq i\neq j\leq n} |\lambda_i - \lambda_j| > 0.$$

Avec $\left|a_{pq}^{(k)}\right| \leq \|E_k\|_s \underset{k\to+\infty}{\to} 0$, on déduit alors que :

$$|b_k| = \left|\frac{a_{pp}^{(k)} - a_{qq}^{(k)}}{2a_{pq}^{(k)}}\right| \underset{k\to+\infty}{\to} +\infty$$

et

$$\tan(\theta_k) = \frac{\text{signe}(b_k)}{\sqrt{b_k^2+1} + |b_k|} \underset{k\to+\infty}{\to} 0.$$

On a donc $\lim_{k\to+\infty} \theta_k = 0$ et $\lim_{k\to+\infty} R_k = I_n$.

(ii) On a $\|P_k\|_2 = 1$ (P_k est orthogonale) et, avec $P_{k+1} - P_k = P_k\big(R(\theta_{k+1}) - I_n\big)$, on déduit que $\lim_{k\to+\infty} (P_{k+1} - P_k) = 0$.

Il nous reste donc à montrer que la suite $(P_k)_{k\in\mathbb{N}}$ n'a qu'un nombre fini de valeurs d'adhérence (lemme 6.4). Soit $\big(P_{\varphi(k)}\big)_{k\in\mathbb{N}}$ une suite extraite de $(P_k)_{k\in\mathbb{N}}$ qui converge vers une matrice orthogonale Q. Avec l'égalité $A_{\varphi(k)} = {}^tP_{\varphi(k)}AP_{\varphi(k)}$, on déduit que :

$${}^tP_\sigma AP_\sigma = D_\sigma = \lim_{k\to+\infty} A_{\varphi(k)} = {}^tQAQ.$$

Les valeurs propres de A étant deux à deux distinctes, on a nécessairement $Q = \big(\pm C_{\sigma(1)}, \pm C_{\sigma(2)}, \ldots, \pm C_{\sigma(n)}\big)$. D'où le résultat. ■

D'un point de vue numérique, le calcul du maximum des termes non diagonaux de A_k à chaque étape n'est pas intéressant car il augmente le temps de calcul. On préfère procéder de la manière suivante : à l'étape k du calcul, la matrice A_{k-1} étant construite, on construit la matrice A_k en effectuant $\dfrac{n(n-1)}{2}$ transformations de Jacobi en prenant pour valeurs successives de (p,q) les valeurs $(1,2), (1,3), \ldots, (1,n); (2,3), \ldots, (2,n); \ldots; (n-1,n)$. Un tel calcul est appelé un balayage.

En notant $S_k = \sum_{1\leq i<j\leq n} \left|a_{ij}^{(k)}\right|$, on arrête les itérations quand $S_k < \epsilon$, où ϵ est une précision donnée.

On peut alors montrer que la convergence devient très rapidement quadratique.

En pratique, on n'effectuera pas de transformation de Jacobi si $\left|a_{pq}^{(k)}\right| < \sigma$, où σ est un seuil donné. Usuellement, on prend la valeur $\sigma = \dfrac{S_k}{5n^2}$ seulement pour les trois premiers balayages, puis, à partir du quatrième, on décide que $a_{pq}^{(k)} = 0$ si $\left|a_{pq}^{(k)}\right| < \epsilon' \left|a_{pp}^{(k)}\right|$ et $\left|a_{pq}^{(k)}\right| < \epsilon' \left|a_{qq}^{(k)}\right|$ avec ϵ' assez petit.

Le calcul de la matrice P_σ, donnant les vecteurs propres de A, se fait avec les relations :

$$\begin{cases} p_{ij}^{(k+1)} = p_{ij}^{(k)} & \text{si } j \neq p,\ j \neq q, \\ p_{ip}^{(k+1)} = p_{ip}^{(k)} + s\left(p_{iq}^{(k)} - \tau p_{ip}^{(k)}\right), & \qquad (i = 1, 2, \ldots, n), \\ p_{iq}^{(k+1)} = p_{iq}^{(k)} - s\left(p_{ip}^{(k)} + \tau p_{iq}^{(k)}\right), \end{cases}$$

où $\tau = \tan\left(\dfrac{\theta_k}{2}\right) = \dfrac{s}{1+c}$.

4. La méthode de Givens et Householder

Soit A une matrice symétrique réelle d'ordre $n \geq 3$. Au paragraphe 6, page 93, on a décrit l'algorithme de Householder qui permet de calculer effectivement une matrice T tridiagonale semblable à la matrice A. Le calcul des valeurs propres de T nous donne alors celles de A. La méthode de Givens (ou de bissection) nous fournit, sous certaines conditions, un algorithme dichotomique de calcul des valeurs propres d'une matrice tridiagonale symétrique. La méthode de la puissance inverse permet ensuite d'obtenir des vecteurs propres associés.

Dans ce paragraphe on se donne une matrice :

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{n-2} & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & b_{n-1} & a_n \end{pmatrix}$$

à coefficients réels, tridiagonale symétrique et supposée irréductible, c'est-à-dire telle que :

$$\forall i \in \{1, 2, \ldots, n-1\}, \quad b_i \neq 0.$$

Une telle matrice est de type Hessenberg irréductible et on sait alors qu'elle admet n valeurs propres réelles simples (exercice 6.2).

De plus, avec $\rho(A) \leq \|A\|_\infty$ où $\rho(A)$ désigne le rayon spectral de A, on déduit que toutes les valeurs propres de la matrice A sont dans l'intervalle $[a_0, b_0] = [-\|A\|_\infty, \|A\|_\infty]$.

Pour tout $k = 1, 2, \ldots, n$, on note $A_k = \left((a_{ij})\right)_{1 \leq i,j \leq k}$ la k-ième sous-matrice principale de A et $P_k(\lambda) = \det(A_k - \lambda I_k)$ son polynôme caractéristique. P_k admet donc k racines réelles distinctes (la matrice A_k vérifie les mêmes propriétés que la matrice A) que l'on range comme suit :

$$\lambda_{1,k} < \lambda_{2,k} < \ldots < \lambda_{k,k}.$$

Les valeurs propres de A sont les $\lambda_j = \lambda_{j,n}$ pour $j = 1, 2, \ldots, n$.

En développant $P_k(\lambda)$ suivant la dernière ligne on obtient la relation de récurrence :

$$\begin{cases} P_0(\lambda) = 1, \quad P_1(\lambda) = a_1 - \lambda, \\ P_k(\lambda) = (a_k - \lambda) P_{k-1}(\lambda) - b_{k-1}^2 P_{k-2}(\lambda) \quad (k = 2, \ldots, n). \end{cases}$$

Pour $k \geq 1$, le coefficient dominant de $P_k(\lambda)$ étant égal à $(-1)^k \lambda^k$, on déduit que :

$$\lim_{\lambda \to -\infty} P_k(\lambda) = \lim_{\lambda \to -\infty} (-1)^k \lambda^k = +\infty.$$

Lemme 6.5 : *Pour $k = 2, \ldots, n$, les racines de P_{k-1} séparent celles de P_k dans le sens où :*

$$\begin{cases} \lambda_{1,k} < \lambda_{1,k-1}, \\ \lambda_{j,k-1} < \lambda_{j+1,k} < \lambda_{j+1,k-1} \quad (1 \leq j \leq k-2), \\ \lambda_{k-1,k-1} < \lambda_{k,k}. \end{cases}$$

Démonstration : Pour $k = 1$, on a $P_1(\lambda) = a_1 - \lambda$ qui admet une racine réelle $\lambda_{1,1} = a_1$.

Pour $k = 2$, on a $P_2(\lambda) = (a_1 - \lambda)(a_2 - \lambda) - b_1^2$, avec $\lim_{\lambda \to -\infty} P_2(\lambda) = +\infty$, $P_2(\lambda_{1,1}) = P_2(a_1) = -b_1^2 < 0$ et $\lim_{\lambda \to +\infty} P_2(\lambda) = +\infty$. Avec le théorème des valeurs intermédiaires, on déduit donc que P_2 admet une racine réelle $\lambda_{1,2} < \lambda_{1,1}$ et une racine réelle $\lambda_{2,2} > \lambda_{1,1}$. On a donc montré que la racine $\lambda_{1,1}$ de P_1 sépare les deux racines de P_2, avec $P_0(\lambda_{1,1}) = 1 > 0$. Ce qui est schématisé par la figure 6.1, page suivante.

Supposons le résultat vrai pour $p = 2, \ldots, k-1$, avec $P_{k-2}(\lambda_{j,k-1})$ du signe de $(-1)^{j-1}$ pour tout $j = 1, \ldots, k-1$. Avec $\lim_{\lambda \to -\infty} P_k(\lambda) = +\infty$ et

$$P_k(\lambda_{1,k-1}) = -b_{k-1}^2 P_{k-2}(\lambda_{1,k-1}) < 0,$$

on déduit qu'il existe une racine de P_k, $\lambda_{1,k} < \lambda_{1,k-1}$ telle que $P_{k-1}(\lambda_{1,k}) > 0$, puisque $P_{k-1}(\lambda) > 0$ sur $]-\infty, \lambda_{1,k-1}[$.

Pour $j = 1, 2, \ldots, k-2$, on a $P_k(\lambda_{j,k-1}) = -b_{k-1}^2 P_{k-2}(\lambda_{j,k-1})$ du signe de $(-1)^j$ et $P_k(\lambda_{j+1,k-1}) = -b_{k-1}^2 P_{k-2}(\lambda_{j+1,k-1})$ du signe de $(-1)^{j+1}$. On en déduit alors que P_k admet une racine $\lambda_{j+1,k} \in]\lambda_{j,k-1}, \lambda_{j+1,k-1}[$ (Fig. 6.2, page suivante).

Si j est impair, alors $P_{k-1}(\lambda_{j+1,k}) < 0$ et si j est pair alors $P_{k-1}(\lambda_{j+1,k}) > 0$. C'est-à-dire que $P_{k-1}(\lambda_{j+1,k})$ est du signe de $(-1)^j$.

Enfin $P_k(\lambda_{k-1,k-1}) = -b_{k-1}^2 P_{k-2}(\lambda_{k-1,k-1})$ est du signe de $(-1)^{k-1}$ et $\lim_{\lambda \to +\infty} P_k(\lambda) = (-1)^k \infty$. On a donc une racine de P_k, $\lambda_{k,k} \in]\lambda_{k-1,k-1}, +\infty[$, telle que $P_{k-1}(\lambda_{k,k})$ soit du signe de $(-1)^{k-1}$. D'où le résultat. ■

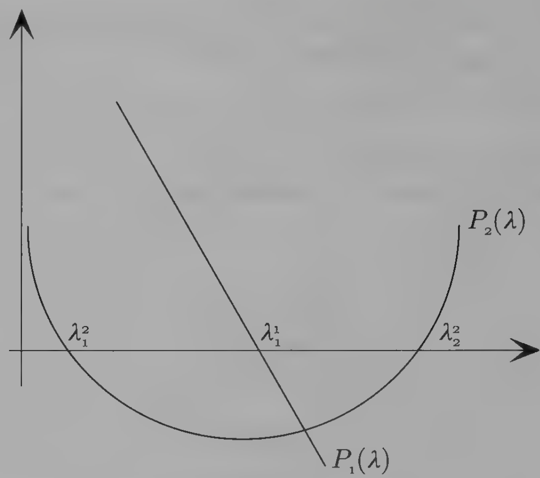

Figure 6.1.

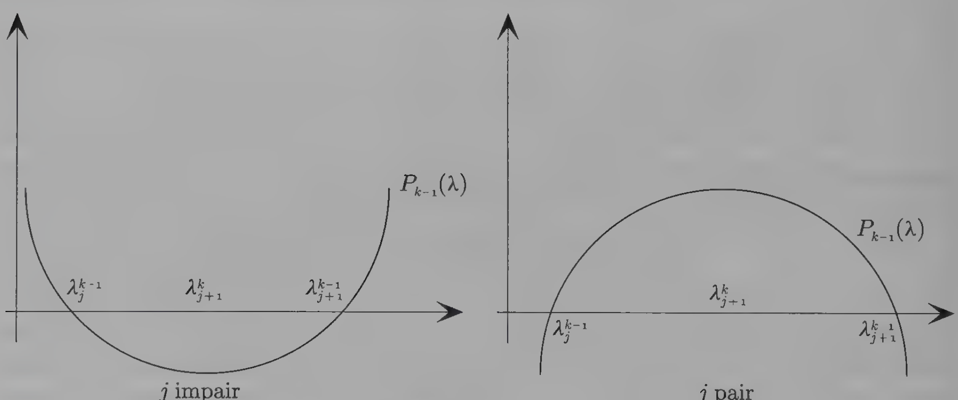

Figure 6.2.

On note, pour tout réel λ et tout entier k compris entre 0 et n :

$$s_k(\lambda) = \begin{cases} \text{signe}\big(P_k(\lambda)\big) & \text{si } P_k(\lambda) \neq 0, \\ s_{k-1}(\lambda) & \text{si } P_k(\lambda) = 0, \end{cases}$$

et $N_k(\lambda)$ désigne le nombre de changements de signes entre deux termes consécutifs de la suite $\big(s_0(\lambda), s_1(\lambda), \ldots, s_k(\lambda)\big)$.

En remarquant que $P_0(\lambda) \neq 0$ et que, pour $k = 1, \ldots, n$, on a $P_{k-1}(\lambda) \neq 0$ si $P_k(\lambda) = 0$, on déduit que la fonction s_k est bien définie et à valeurs dans $\{-1, 1\}$ avec $s_0(\lambda) = 1$ pour tout réel λ.

Lemme 6.6 : *Avec les notations qui précèdent, on a*

$$s_1(\lambda) = \begin{cases} 1 & si\ \lambda \in \left]-\infty, \lambda_{1,1}\right], \\ -1 & si\ \lambda \in \left]\lambda_{1,1}, +\infty\right[, \end{cases}$$

et pour tout $k = 2, \ldots, n$:

$$s_k(\lambda) = \begin{cases} 1 & si\ \lambda \in \left]-\infty, \lambda_{1,k}\right], \\ (-1)^j & si\ \lambda \in \left]\lambda_{j,k}, \lambda_{j+1,k}\right] \quad (1 \leq j \leq k-1), \\ (-1)^k & si\ \lambda \in \left]\lambda_{k,k}, +\infty\right] \end{cases}$$

Démonstration : Le résultat sur s_1 est évident.

Avec $\lim\limits_{\lambda \to -\infty} P_k(\lambda) = +\infty$ on déduit que $P_k(\lambda) > 0$ sur $\left]-\infty, \lambda_{1,k}\right[$ pour tout entier k compris entre 1 et n. Et avec $\lambda_{1,k} < \lambda_{1,k-1}$ on conclut que $s_k(\lambda) = 1$ sur $\left]-\infty, \lambda_{1,k}\right]$.

Pour tout entier k compris entre 2 et n et tout entier j compris entre 1 et $k-1$, P_k est du signe de $(-1)^j$ sur $\left]\lambda_{j,k}, \lambda_{j+1,k}\right[$. En considérant que $\lambda_{j+1,k} \in \left]\lambda_{j,k-1}, \lambda_{j+1,k-1}\right[$, on déduit que $s_k(\lambda) = (-1)^j$ sur $\left]\lambda_{j,k}, \lambda_{j+1,k}\right]$ pour tout $k = 3, \ldots, n$ et tout $j = 1, 2, \ldots, k-1$. Pour $k = 2$, le résultat est encore valable puisque $s_2(\lambda_{2,2}) = s_1(\lambda_{2,2}) = -1$.

Enfin, avec $\lim\limits_{\lambda \to +\infty} P_k(\lambda) = (-1)^k \infty$, on déduit que $s_k(\lambda) = (-1)^k$ sur $\left]\lambda_{k,k}, +\infty\right[$. ∎

Lemme 6.7 : *Pour tout réel λ et tout entier k compris entre 1 et n, $N_k(\lambda)$ est égal au nombre de racines de P_k qui sont strictement inférieures à λ.*

Démonstration : On montre le résultat par récurrence sur $k \geq 1$.

Pour $k = 1$, on a :

$$(s_0(\lambda), s_1(\lambda)) = \begin{cases} (1,1) & \text{si } \lambda \leq a_1 = \lambda_{1,1}, \\ (1,-1) & \text{si } \lambda > a_1 = \lambda_{1,1}, \end{cases}$$

et

$$N_1(\lambda) = \begin{cases} 0 & \text{si } \lambda \leq \lambda_{1,1}, \\ 1 & \text{si } \lambda > \lambda_{1,1}. \end{cases}$$

Le résultat est donc vrai pour $k = 1$.

Supposons le résultat vrai pour $k-1$ compris entre 1 et $n-1$ et soit $\lambda \in \mathbb{R}$.

Si $\lambda \leq \lambda_{1,k} < \lambda_{1,j}$ $(j = 1, 2, \ldots, k-1)$, alors $P_j(\lambda) > 0$ pour tout entier j compris entre 1 et k et $(s_0(\lambda), s_1(\lambda), \ldots, s_k(\lambda)) = (1, 1, \ldots, 1)$. C'est-à-dire que $N_k(\lambda) = 0$ et c'est bien le nombre de racines de P_k strictement inférieures à λ.

Si $\lambda \in \left]\lambda_{j,k}, \lambda_{j+1,k}\right]$ pour un entier $j \in \{1, 2, \ldots, k-1\}$, on distingue alors deux cas en fonction de la position de λ par rapport à $\lambda_{j,k-1} \in \left]\lambda_{j,k}, \lambda_{j+1,k}\right]$.

Si $\lambda \in \left]\lambda_{j,k}, \lambda_{j,k-1}\right]$, on a alors $j-1$ racines de P_{k-1} strictement inférieures à λ. C'est-à-dire, avec l'hypothèse de récurrence, que $N_{k-1}(\lambda) = j-1$. En tenant compte de $s_{k-1}(\lambda) = (-1)^{j-1}$ et $s_k(\lambda) = (-1)^j$, on déduit que $N_k(\lambda) = N_{k-1}(\lambda) + 1 = j$ et c'est bien le nombre de racines de P_k strictement inférieures à λ.

Si $\lambda \in \left]\lambda_{j,k-1}, \lambda_{j+1,k}\right]$, on a alors j racines de P_{k-1} strictement inférieures à λ. C'est-à-dire, avec l'hypothèse de récurrence, que $N_{k-1}(\lambda) = j$. En tenant compte de $s_{k-1}(\lambda) = (-1)^j$ et $s_k(\lambda) = (-1)^j$, on déduit que $N_k(\lambda) = N_{k-1}(\lambda) = j$ et c'est bien le nombre de racines de P_k strictement inférieures à λ.

Enfin, si $\lambda > \lambda_{k,k} > \lambda_{k-1,k-1}$, alors, avec l'hypothèse de récurrence, on déduit que $N_{k-1}(\lambda) = k-1$. Avec $s_{k-1}(\lambda) = (-1)^{k-1}$ et $s_k(\lambda) = (-1)^k$, on déduit que $N_k(\lambda) = N_{k-1}(\lambda) + 1 = k$ et c'est bien le nombre de racines de P_k strictement inférieures à λ. ■

Pour tout réel λ, on désigne par $N(\lambda)$ le nombre de racines de P_n qui sont strictement inférieures à λ et on note $c_0 = \dfrac{a_0 + b_0}{2}$ le milieu de l'intervalle $[a_0, b_0] = [-\|A\|_\infty, \|A\|_\infty]$.

Lemme 6.8 : *Soit i un entier compris entre 1 et n. Si $N(c_0) \geq i$ alors $\lambda_i \in [a_0, c_0]$, sinon $\lambda_i \in [c_0, b_0]$.*

Démonstration : Du lemme 6.7, on déduit que si $\lambda_i \in [c_0, b_0]$, alors $N(c_0) \leq i-1$, et si $\lambda_i \in [a_0, c_0]$, alors $N(c_0) \geq i$. D'où le résultat. ■

Pour i fixé entre 1 et $n-1$, on note $[a_1, b_1]$ la moitié de l'intervalle $[a_0, b_0]$ qui contient la valeur propre λ_i et par récurrence on construit une suite $([a_k, b_k])_{k\geq 1}$ d'intervalles emboîtés telle que :

$$\forall k \geq 1, \begin{cases} b_k - a_k = \dfrac{b_0 - a_0}{2^k}, \\ \lambda_i \in [a_k, b_k]. \end{cases}$$

Il suffit en effet de reprendre le raisonnement du lemme 6.8 avec chaque intervalle $[a_k, b_k]$.

On a alors :

$$\lambda_i = \lim_{k\to+\infty} a_k = \lim_{k\to+\infty} b_k = \lim_{k\to+\infty} c_k,$$

où on a noté c_k le milieu de l'intervalle $[a_k, b_k]$ (théorème des suites adjacentes).

Le calcul des vecteurs propres associés peut se faire ensuite en utilisant la méthode de la puissance inverse.

Soient $i \in \{1, 2, \ldots, n\}$ fixé èt λ un réel qui vérifie :

$$0 < |\lambda - \lambda_i| < \inf_{1\leq j\neq i\leq n} |\lambda - \lambda_j|.$$

On se donne un vecteur x_0 non orthogonal au sous-espace propre associé à λ_i et on définit la suite de vecteurs $(x_k)_{k\in\mathbb{N}}$ par :

$$\forall k \geq 0, \quad (A - \lambda I_n) x_{k+1} = x_k.$$

Lemme 6.9 : *Avec les notations qui précèdent, la suite $\left(\dfrac{\operatorname{signe}(\lambda_i - \lambda)^k}{\|x_k\|} x_k\right)_{k\in\mathbb{N}}$ converge vers un vecteur propre non nul associé à la valeur propre λ_i.*

Démonstration : On remarque que si $x_{k+1} = 0$, alors $x_k = 0$. On déduit donc par récurrence que si $x_0 \neq 0$ alors $x_k \neq 0$ pour tout entier k.

Soit $\{f_1, f_2, \ldots, f_n\}$ une base orthonormée de vecteurs propres avec $Af_j = \lambda_j f_j$ pour tout $j = 1, 2, \ldots, n$. Pour tout entier k, on note $x_k = \sum_{j=1}^{n} x_{j,k} f_j$. Avec $(A - \lambda I_n) x_{k+1} = x_k$, on déduit que $x_{j,k+1} = \dfrac{1}{\lambda_j - \lambda} x_{j,k}$ et par récurrence $x_{j,k} = \dfrac{1}{(\lambda_j - \lambda)^k} x_{j,0}$. On peut alors écrire que $x_k = \dfrac{1}{(\lambda_i - \lambda)^k} y_k$, où y_k est le vecteur de composantes $y_{j,k} = \left(\dfrac{\lambda_i - \lambda}{\lambda_j - \lambda}\right)^k x_{j,0}$. On a alors :

$$\lim_{k \to +\infty} y_{j,k} = \begin{cases} 0 & \text{si } j \neq i, \\ x_{i,0} & \text{si } j = i, \end{cases}$$

avec $x_i^0 \neq 0$. C'est-à-dire que $\lim_{k \to +\infty} y_k = x_{i,0} f_i$ est un vecteur propre non nul associé à la valeur propre λ_i. On déduit alors que :

$$\lim_{k \to +\infty} \frac{\operatorname{signe}(\lambda_i - \lambda)^k}{\|x_k\|} x_k = \lim_{k \to +\infty} \frac{1}{\|y_k\|} y_k = \operatorname{signe}(x_{i,0}) f_i. \qquad \blacksquare$$

Remarque 6.6 : *Si la matrice A n'est pas irréductible, on peut alors la découper en blocs de matrices irréductibles et l'algorithme de Givens permet encore de calculer les valeurs propres.*

5. Exercices

Exercice 6.1 : *Soient $A \in \mathcal{M}_n(\mathbb{K})$ et λ une valeur propre de A d'ordre $p \geq 1$. Montrer que les sous-espaces vectoriels $N_\lambda = \ker(A - \lambda I_n)^p$ et $M_\lambda = \operatorname{Im}(A - \lambda I_n)^p$ sont stables par A, supplémentaires dans $\mathbb{K}^n$ et que N_λ est de dimension p.*

Solution : Avec la commutativité de $\mathbb{K}[A]$ on vérifie facilement que les sous-espaces vectoriels N_λ et M_λ sont stables par A.

On a vu (théorème 2.8) que N_λ est de dimension p et avec le théorème du rang on déduit que M_λ est de dimension $n - p$.

Enfin, en notant $P_A(X) = (X - \lambda)^p Q(X)$ le polynôme caractéristique de A, le polynôme Q étant premier avec $(X - \lambda)^p$, on peut trouver des polynômes U et V tels que :

$$U(X)(X - \lambda)^p + V(X) Q(X) = 1$$

(théorème de Bézout) et tout vecteur x de $\mathbb{K}^n$ s'écrit :

$$x = U(A)(A - \lambda I_n)^p (x) + V(A) Q(A)(x)$$

avec $U(A)(A - \lambda I_n)^p (x) \in M_\lambda$ et $V(A) Q(A)(x) \in N_\lambda$. On a donc $\mathbb{K}^n = N_\lambda + M_\lambda$ et la somme est directe en considérant les dimensions.

Exercice 6.2 : *On appelle matrice de Hessenberg une matrice A à coefficients complexes qui vérifie $a_{ij} = 0$ pour $j < i-1$. On dit que A est irréductible si $a_{i,i-1} \neq 0$ pour tout $i = 2, \ldots, n$.*

1. Montrer que si A est une matrice de Hessenberg irréductible, alors pour toute valeur propre de A, l'espace propre associé est de dimension 1.

2. En déduire que les valeurs propres d'une matrice de Hessenberg irréductible sont simples si et seulement si la matrice est diagonalisable.

3. Soit :

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ b_1 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & b_{n-2} & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & b_{n-1} & a_n \end{pmatrix}$$

une matrice à coefficients réels, tridiagonale symétrique et irréductible.

(a) Montrer que les valeurs propres de A sont simples.

(b) Décrire un algorithme de calcul de l'espace propre associé à une valeur propre de A.

4. Soit :

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ c_2 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & c_n & a_n \end{pmatrix}$$

une matrice tridiagonale à coefficients complexes.

(a) Donner un algorithme de calcul du polynôme caractéristique de A.

(b) Montrer que A admet les mêmes valeurs propres que la matrice :

$$B = \begin{pmatrix} a_1 & c_2b_1 & 0 & \ldots & 0 \\ 1 & a_2 & c_3b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & a_{n-1} & c_nb_{n-1} \\ 0 & \ldots & 0 & 1 & a_n \end{pmatrix}.$$

(c) Montrer que si $a_i \in \mathbb{R}$ pour tout $i = 1, 2, \ldots, n$ et $c_{i+1}b_i \in \mathbb{R}_+^*$ pour tout $i = 1, 2, \ldots, n-1$, alors A admet n valeurs propres réelles simples et est diagonalisable.

(d) Que peut-on dire si $a_i \in \mathbb{R}$ pour tout $i = 1, 2, \ldots, n$ et $c_{i+1}b_i \in \mathbb{R}_-^*$ pour tout $i = 1, 2, \ldots, n-1$?

Solution :

1. Soient A une matrice de Hessenberg irréductible, λ un complexe et :

$$A_\lambda = A - \lambda I_n = \begin{pmatrix} a_{11}-\lambda & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22}-\lambda & a_{23} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & a_{n-2,n} \\ \vdots & \ddots & a_{n-1,n-2} & a_{n-1,n-1}-\lambda & a_{n-1,n} \\ 0 & \cdots & 0 & a_{n,n-1} & a_{n,n}-\lambda \end{pmatrix}.$$

Si B_λ est la matrice extraite de A_λ en supprimant la première ligne et la dernière colonne, soit :

$$B_\lambda = \begin{pmatrix} a_{21} & a_{22}-\lambda & a_{23} & \cdots & a_{2,n-1} \\ 0 & a_{32} & a_{33}-\lambda & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & a_{n-2,n} \\ \vdots & \ddots & 0 & a_{n-1,n-2} & a_{n-1,n-1}-\lambda \\ 0 & \cdots & 0 & 0 & a_{n,n-1} \end{pmatrix}.$$

On a $\det(B_\lambda) = \prod_{i=2}^{n} a_{i,i-1} \neq 0$. On a donc ainsi montré que :

$$\forall \lambda \in \mathbb{C}, \quad \operatorname{rang}(A_\lambda) \geq n-1,$$

ou encore :

$$\forall \lambda \in \mathbb{C}, \quad \dim\bigl(\ker(A - \lambda I_n)\bigr) \leq 1.$$

En particulier, pour λ valeur propre de A on a :

$$\dim\bigl(\ker(A - \lambda I_n)\bigr) = 1.$$

2. On sait que la matrice A est diagonalisable sur $\mathbb{C}$ si et seulement si pour toute valeur propre λ de A, la dimension de $\ker(A - \lambda I_n)$ est égale à la multiplicité de λ comme racine du polynôme caractéristique de A. Avec ce qui précède on déduit alors que la matrice A est diagonalisable si et seulement si toutes ses valeurs propres sont simples.

3. (a) Une matrice symétrique réelle a toutes ses valeurs propres réelles et est diagonalisable. Si de plus elle est tridiagonale et irréductible alors toutes ses valeurs propres sont simples.

(b) Soit λ une valeur propre de A et $x \in \mathbb{R}^n - \{0\}$ un vecteur propre associé. On a alors :

$$\begin{cases} a_1 x_1 + b_1 x_2 = \lambda x_1, \\ b_{k-1} x_{k-1} + a_k x_k + b_k x_{k+1} = \lambda x_k \quad (2 \leq k \leq n-1), \\ b_{n-1} x_{n-1} + a_n x_n = \lambda x_n. \end{cases}$$

Si $x_n = 0$, alors avec l'hypothèse $b_i \neq 0$ pour tout $i = 1, \ldots, n-1$, on déduit que tous les x_i sont nuls. On peut donc prendre $x_n = 1$ et $(x_1, x_2, \ldots, x_{n-1})$ est solution du système triangulaire supérieur :

$$\begin{cases} b_{k-1}x_{k-1} + (a_k - \lambda)\, x_k + b_k x_{k+1} = 0 \quad (2 \leq k \leq n-2), \\ b_{n-2}x_{n-2} + (a_{n-1} - \lambda)\, x_{n-1} = -b_{n-1}, \\ b_{n-1}x_{n-1} = \lambda - a_n. \end{cases}$$

La solution de ce système peut se calculer avec l'algorithme :

$$\begin{cases} x_{n-1} = \dfrac{\lambda - a_n}{b_{n-1}}, \\ x_{k-1} = -\dfrac{b_k x_{k+1} + (a_k - \lambda)\, x_k}{b_{k-1}} \quad (k = n-1, \ldots, 2). \end{cases}$$

On retrouve le fait que l'espace propre associé à la valeur propre λ est de dimension 1.

4. (a) Si, pour tout $k = 1, 2, \ldots, n$, on désigne par A_k la matrice principale d'ordre k de A, alors la suite $(P_k)_{1 \leq k \leq n}$ des polynômes caractéristiques des A_k vérifie la récurrence :

$$\begin{cases} P_0(\lambda) = 1, \quad P_1(\lambda) = a_1 - \lambda, \\ P_k(\lambda) = (a_k - \lambda)\, P_{k-1}(\lambda) - b_{k-1}c_k P_{k-2}(\lambda) \quad (2 \leq k \leq n). \end{cases}$$

(b) Le polynôme caractéristique de la matrice B s'obtenant avec la même récurrence que celui de A en utilisant les mêmes conditions initiales, ces deux polynômes sont donc identiques.

(c) En utilisant deux fois le résultat précédent, on voit que A admet les mêmes valeurs propres que la matrice :

$$C = \begin{pmatrix} a_1 & \sqrt{c_2 b_1} & 0 & \cdots & 0 \\ \sqrt{c_2 b_1} & a_2 & \sqrt{c_3 b_2} & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \sqrt{c_{n-1} b_{n-2}} & a_{n-1} & \sqrt{c_n b_{n-1}} \\ 0 & \cdots & 0 & \sqrt{c_n b_{n-1}} & a_n \end{pmatrix}.$$

La matrice C étant symétrique tridiagonale et irréductible, elle est donc diagonalisable avec n valeurs propres réelles simples. La matrice A admet donc n valeurs propres réelles simples et est diagonalisable.

(d) Soit $A = \begin{pmatrix} a & -b_1 & 0 \\ 1 & 0 & -b_2 \\ 0 & 1 & a \end{pmatrix}$, avec $a \in \mathbb{R}$, $b_i > 0$ pour $i = 1, 2$. Son polynôme caractéristique est :

$$P(\lambda) = (a - \lambda)\left(\lambda^2 - a\lambda + b_2 + b_1\right)$$

et pour $a^2 - 4(b_1 + b_2) = 0$ la matrice A admet une valeur propre double et donc n'est pas diagonalisable.

Exercice 6.3 : *Soit :*

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ c_2 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & c_{n-1} & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & c_n & a_n \end{pmatrix}$$

une matrice tridiagonale à coefficients complexes telle que $a_i \in \mathbb{R}$ pour tout $i = 1, 2, \ldots, n$ et $c_{i+1} b_i \in \mathbb{R}_+^*$ pour tout $i = 1, 2, \ldots, n-1$.

1. Montrer que A admet les mêmes valeurs propres que la matrice :

$$B = \begin{pmatrix} a_1 & c_2 b_1 & 0 & \ldots & 0 \\ 1 & a_2 & c_3 b_2 & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & \ddots & 1 & a_{n-1} & c_n b_{n-1} \\ 0 & \ldots & 0 & b_n & a_n \end{pmatrix}.$$

Dans tout ce qui suit, on suppose que :

$$A = \begin{pmatrix} a_1 & b_1 & 0 & \ldots & 0 \\ 1 & a_2 & b_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & a_{n-1} & b_{n-1} \\ 0 & \ldots & 0 & 1 & a_n \end{pmatrix}$$

avec $a_i, b_i \in \mathbb{R}$, $b_i > 0$. On note :

$$D = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 1 & 0 & 1 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & 0 & 1 \\ 0 & \ldots & 0 & 1 & 0 \end{pmatrix}.$$

On définit alors une homotopie de D vers A en posant :

$$\forall t \in [0,1], \quad A(t) = (1-t) D + tA.$$

2. Calculer les valeurs propres de D.

3. Montrer que, pour tout t dans $[0,1]$, $A(t)$ admet n valeurs propres réelles simples.

4. Montrer qu'il existe une fonction continûment dérivable :

$$f = (f_1, \ldots, f_n) : [0,1] \to \mathbb{R}^n$$

telle que le polynôme caractéristique de $A(t)$ s'écrive :

$$\forall t \in [0,1], \quad P(\lambda, t) = \prod_{j=1}^{n} (\lambda - f_j(t)).$$

Pour tout t dans $[0,1]$ et $j = 1, 2, \ldots, n$, on pose $\varphi_j(t) = P(f_j(t), t)$ et on a alors $\varphi_j(t) = 0$ sur $[0,1]$.

5. Montrer que pour tout $j = 1, 2, \ldots, n$, f_j est solution sur $[0,1]$ d'une équation différentielle avec condition initiale.

6. Donner un algorithme de calcul de toutes les valeurs propres de la matrice A.

Solution :

1. Voir l'exercice 6.2.

2. En utilisant les résultats de l'exercice 2.10, on déduit que les valeurs propres de la matrice D sont données par :

$$\lambda_k = 2\cos(k\theta) \quad (1 \leq k \leq n),$$

avec $\theta = \dfrac{\pi}{n+1}$.

3. Pour tout t dans $[0,1]$, on a :

$$A(t) = \begin{pmatrix} ta_1 & 1 - t + tb_1 & 0 & \cdots & 0 \\ 1 & ta_2 & 1 - t + tb_2 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & ta_{n-1} & 1 - t + tb_{n-1} \\ 0 & \cdots & 0 & 1 & ta_n \end{pmatrix}$$

et cette matrice admet n valeurs propres réelles simples puisqu'elle est irréductible avec $1 - t + tb_i > 0$ pour tout $t \in [0,1]$ et tout $i = 1, 2, \ldots, n$ (exercice 6.2).

4. Soit $t_0 \in [0,1]$ et λ_0 une valeur propre de $A(t_0)$. On a $P(\lambda_0, t_0) = 0$ et $\dfrac{\partial P}{\partial \lambda}(\lambda_0, t_0) \neq 0$ puisque λ_0 est racine simple de $P(\cdot, t_0)$. Le théorème des fonctions implicites nous dit alors qu'il existe un voisinage ouvert $\mathcal{V}_0$ de t_0 dans $[0,1]$ et une fonction continûment dérivable :

$$f = (f_1, \ldots, f_n) : \mathcal{V}_0 \to \mathbb{R}^n$$

tels que

$$\forall t \in V_0, \quad P(\lambda, t) = \prod_{k=1}^{n} (\lambda - f_k(t)).$$

Les valeurs propres de $A(t)$ étant toutes simples, on peut supposer que $f_1(t) < f_2(t) < \ldots < f_n(t)$. Avec cette condition la fonction f est unique.

L'intervalle $[0,1]$ étant compact, on peut trouver une partition du type :

$$[0,1] = [a_0, b_0[\cup [a_1, b_1[\cup \ldots \cup [a_p, b_p],$$

où

$$0 = a_0 < b_0 = a_1 < b_1 = a_2 < \ldots < b_{p-1} = a_p < b_p = 1,$$

telle que, pour tout $k = 0, 1, \ldots, p$, il existe une unique fonction continûment dérivable :

$$f_k = (f_{1,k}, \ldots, f_{n,k}) : [a_k, b_k] \to \mathbb{R}^n$$

avec :

$$\forall t \in [a_k, b_k], \begin{cases} P(\lambda, t) = \prod_{j=1}^{n} (\lambda - f_{j,k}(t)), \\ f_{1,k}(t) < f_{2,k}(t) < \ldots < f_{n,k}(t). \end{cases}$$

La fonction f définie sur $[0,1]$ par :

$$f(t) = (f_{1,k}(t), \ldots, f_{n,k}(t)) \text{ si } t \in [a_k, b_k]$$

est alors continûment dérivable et $P(\lambda, t) = \prod_{j=1}^{n} (\lambda - f_j(t))$ pour tout $t \in [0,1]$.

5. En dérivant la relation $\varphi_j(t) = 0$ par rapport à t, pour tout $j = 1, 2, \ldots, n$, on déduit alors que f_j est solution du problème de Cauchy :

$$\begin{cases} f_j(0) = a_j, \\ f_j'(t) = -\dfrac{\dfrac{\partial P}{\partial t}(f_j(t), t)}{\dfrac{\partial P}{\partial \lambda}(f_j(t), t)} \quad (t \in [0,1]). \end{cases}$$

La simplicité des valeurs propres de $A(t)$ nous garantit que $\dfrac{\partial P}{\partial \lambda}(f_j(t), t) \neq 0$ sur $[0,1]$.

6. Le calcul du polynôme caractéristique $P(\lambda, t)$ se fait en utilisant la récurrence :

$$\begin{cases} P_0(\lambda, t) = 1, \quad P_1(\lambda, t) = ta_1 - \lambda, \\ P_k(\lambda, t) = (ta_k - \lambda) P_{k-1}(\lambda, t) - (1 - t + tb_{k-1}) P_{k-2}(\lambda, t) \quad (k = 2, \ldots, n). \end{cases}$$

En dérivant par rapport à λ et à t, on obtient l'algorithme suivant de calcul des dérivées partielles de P :

$$\begin{cases} \dfrac{\partial P_0}{\partial \lambda}(\lambda,t) = 0, \qquad \dfrac{\partial P_0}{\partial t}(\lambda,t) = 0, \\ \dfrac{\partial P_1}{\partial \lambda}(\lambda,t) = -1, \qquad \dfrac{\partial P_1}{\partial t}(\lambda,t) = a_1, \\ \dfrac{\partial P_k}{\partial \lambda}(\lambda,t) = (ta_k - \lambda)\dfrac{\partial P_{k-1}}{\partial \lambda}(\lambda,t) - P_{k-1}(\lambda,t) - (1 - t + tb_{k-1})\dfrac{\partial P_{k-2}}{\partial \lambda}(\lambda,t), \\ \dfrac{\partial P_k}{\partial t}(\lambda,t) = (ta_k - \lambda)\dfrac{\partial P_{k-1}}{\partial t}(\lambda,t) + a_k P_{k-1}(\lambda,t) \\ \qquad\qquad - (1 - t + tb_{k-1})\dfrac{\partial P_{k-2}}{\partial t}(\lambda,t) - (b_{k-1} - 1)P_{k-2}(\lambda,t). \end{cases}$$

En utilisant la méthode de Runge-Kutta d'ordre 4, on peut alors calculer des valeurs approchées des valeurs propres $f_j(1)$ de la matrice A.

7 Systèmes différentiels linéaires et exponentielle d'une matrice

Pour $n \in \mathbb{N} - \{0\}$, l'espace vectoriel $\mathbb{C}^n$ est muni d'une norme quelconque notée $\|\cdot\|$.

Étant donnés un ouvert Ω de $\mathbb{R} \times \mathbb{C}^n$ et une fonction continue

$$f : \begin{array}{ccc} \Omega & \longrightarrow & \mathbb{C}^n \\ (t, y) & \mapsto & f(t, y), \end{array}$$

on dit qu'une fonction $y : I \longrightarrow \mathbb{R}$ définie sur un intervalle réel I est solution du système différentiel :

$$y' = f(t, y), \tag{7.1}$$

si cette fonction est dérivable sur I avec :

$$\forall t \in I, \quad \big(t, y(t)\big) \in \Omega \text{ et } y'(t) = f\big(t, y(t)\big).$$

Le problème de Cauchy associé à ce système et aux conditions initiales (t_0, y_0) consiste à rechercher les solutions de (7.1) qui vérifient $y(t_0) = y_0$.

Pour tout intervalle réel $[a, b]$ on désigne par $\mathcal{C}^0([a, b], \mathbb{C}^n)$ l'espace vectoriel des fonctions continues définies sur $[a, b]$ et à valeurs dans $\mathbb{C}^n$. On munit cet espace vectoriel de la norme de la convergence uniforme définie par :

$$\forall y \in \mathcal{C}^0([a, b], \mathbb{C}^n), \quad \|y\|_\infty = \sup_{t \in [a,b]} \|y(t)\|.$$

On rappelle que l'espace vectoriel E muni de cette norme est complet (théorème 1.6).

1. Systèmes différentiels linéaires à coefficients constants

Pour tout réel strictement positif α, on note $E_\alpha = \mathcal{C}^0([-\alpha, \alpha], \mathbb{C}^n)$ et on munit cet espace vectoriel de la norme de la convergence uniforme $\|\cdot\|_\infty$.

On se donne une matrice non nulle $A \in \mathcal{M}_n(\mathbb{C})$, un vecteur $y_0 \in \mathbb{C}^n$ et on définit l'application $\varphi : E_\alpha \to E_\alpha$ par :

$$\forall y \in E_\alpha, \quad \forall t \in [-\alpha, \alpha], \; \varphi(y)(t) = A \int_0^t y(u)\, du + y_0.$$

Lemme 7.1 : *Il existe un entier p strictement positif tel que l'application itérée $\varphi^p = \underbrace{\varphi \circ \cdots \circ \varphi}_{p \text{ fois}}$ soit strictement contractante.*

Démonstration : Pour y, z dans E_α et t dans $[-\alpha, \alpha]$ on a :

$$\begin{aligned}\|\varphi(y)(t) - \varphi(z)(t)\| &= \left\| A \int_0^t (y(u) - z(u))\, \mathrm{d}u \right\| \\ &\leq \|A\| \left| \int_0^t \|y(u) - z(u)\|\, \mathrm{d}u \right| \leq \|A\| \, \|y - z\|_\infty \, |t| ,\end{aligned}$$

où $A \mapsto \|A\|$ est la norme matricielle induite par la norme vectorielle $x \mapsto \|x\|$.
Par récurrence sur $k \geq 1$, on déduit que :

$$\forall k \geq 1, \forall t \in [-\alpha, \alpha] , \ \left\|\varphi^k(y)(t) - \varphi^k(z)(t)\right\| \leq \frac{(\|A\| \, |t|)^k}{k!} \|y - z\|_\infty .$$

En effet, on vient de voir que le résultat est vrai pour $k = 1$ et, en le supposant acquis pour $k \geq 1$, on a pour tout $t \in [-\alpha, \alpha]$:

$$\begin{aligned}\left\|\varphi^{k+1}(y)(t) - \varphi^{k+1}(z)(t)\right\| &\leq \|A\| \left| \int_0^t \left\|\varphi^k(y)(u) - \varphi^k(z)(u)\right\|\, \mathrm{d}u \right| \\ &\leq \|A\| \frac{\|A\|^k}{k!} \|y - z\|_\infty \left| \int_0^t |u|^k \, \mathrm{d}u \right| ,\end{aligned}$$

avec :

$$\left| \int_0^t |u|^k \, \mathrm{d}u \right| = \int_0^{|t|} u^k \, \mathrm{d}u = \frac{|t|^{k+1}}{k+1} .$$

Ce qui donne :

$$\left\|\varphi^{k+1}(y)(t) - \varphi^{k+1}(z)(t)\right\| \leq \frac{(\|A\| \, |t|)^{k+1}}{(k+1)!} \|y - z\|_\infty .$$

On déduit alors que :

$$\forall k \geq 1, \ \left\|\varphi^k(y) - \varphi^k(z)\right\|_\infty \leq \frac{(\|A\| \, \alpha)^k}{k!} \|y - z\|_\infty$$

et, avec $\lim\limits_{k \to +\infty} \dfrac{(\|A\| \, \alpha)^k}{k!} = 0$, on déduit que pour $p \geq 1$ assez grand l'application φ^p est contractante. ■

On considère le problème de Cauchy :

$$\begin{cases} y'(t) = Ay(t) & (t \in [-\alpha, \alpha]), \\ y(0) = y_0. \end{cases} \tag{7.2}$$

Lemme 7.2 : *Si une fonction y définie sur $[-\alpha, \alpha]$ est solution de (7.2), elle est alors de classe $\mathcal{C}^\infty$ sur $[-\alpha, \alpha]$.*

Démonstration : Soit $y \in \mathcal{C}^1([-\alpha, \alpha], \mathbb{C}^n)$ solution de (7.2).

On vérifie, par récurrence sur $k \geq 1$, que y est de classe $\mathcal{C}^k$ pour tout $k \geq 1$. Le résultat est vrai pour $k = 1$. En le supposant acquis pour $k \geq 1$, on déduit de $y' = Ay$ que y est de classe $\mathcal{C}^{k+1}$ avec $y^{(k+1)} = Ay^{(k)}$. ■

Lemme 7.3 : *Une fonction y définie sur $[-\alpha, \alpha]$ est solution de (7.2) si et seulement si c'est un point fixe de l'application φ dans E_α.*

Démonstration : Si $y \in \mathcal{C}^1([-\alpha, \alpha], \mathbb{C}^n)$ est solution du système différentiel (7.2), pour tout $t \in [-\alpha, \alpha]$ on a :

$$y(t) = \int_0^t y'(u)\ du + y(0) = A \int_0^t y(u)\ du + y_0 = \varphi(y)(t),$$

c'est-à-dire que $y = \varphi(y)$.

Réciproquement, si $y \in E_\alpha$ est point fixe de φ, alors $y(0) = y_0$ et y est de classe $\mathcal{C}^1$ sur $[-\alpha, \alpha]$ avec $y' = Ay$. ■

Le problème de Cauchy (7.2) est donc équivalent au problème du point fixe $y = \varphi(y)$ sur E_α.

Lemme 7.4 : *Pour tout réel strictement positif α, le problème (7.2) admet une unique solution y définie par $y = \lim\limits_{k \to +\infty} (y_k)$ où la suite de fonctions $(y_k)_{k \in \mathbb{N}}$ est définie par :*

$$\forall k \in \mathbb{N},\ \forall t \in [-\alpha, \alpha], \quad y_k(t) = \left(\sum_{j=0}^{k} \frac{t^j}{j!} A^j \right) y_0,$$

la convergence de la suite de fonctions $(y_k)_{k \in \mathbb{N}}$ vers y étant uniforme sur $[-\alpha, \alpha]$.

Démonstration : Pour p assez grand, l'itérée φ^p de φ est contractante. Le théorème du point fixe itéré (théorème 1.8) nous dit alors que φ admet un unique point fixe donné par $y = \lim\limits_{k \to +\infty} (y_k)$ dans $(E_\alpha, \|\cdot\|_\infty)$ (c'est-à-dire que la suite de fonctions $(y_k)_{k \in \mathbb{N}}$ converge uniformément vers y sur $[-\alpha, \alpha]$), où $(y_k)_{k \in \mathbb{N}}$ est une suite d'approximations successives de y définie par :

$$\begin{cases} \forall t \in [-\alpha, \alpha], & y_0(t) = y_0, \\ \forall k \in \mathbb{N}, & y_{k+1} = \varphi(y_k). \end{cases}$$

Par récurrence, on a facilement :

$$\forall k \in \mathbb{N},\ \forall t \in [-\alpha, \alpha], \quad y_k(t) = \left(\sum_{j=0}^{k} \frac{t^j}{j!} A^j \right) y_0.$$

En effet, pour $k = 1$, on a $y_1(t) = A \int_0^t y_0 \, \mathrm{d}u + y_0 = (tA + I_n)\, y_0$ et en supposant le résultat acquis pour $k \geq 1$:

$$y_{k+1}(t) = A \int_0^t \left(\sum_{j=0}^{k} \frac{u^j}{j!} A^j \right) y_0 \, \mathrm{d}u + y_0 = \left(\sum_{j=0}^{k+1} \frac{t^j}{j!} A^j \right) y_0. \qquad \blacksquare$$

Théorème 7.1 : *Le problème de Cauchy :*

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(0) = y_0, \end{cases} \tag{7.3}$$

admet une unique solution définie par :

$$\forall t \in \mathbb{R}, \quad y(t) = \lim_{k \to +\infty} \left(\sum_{j=0}^{k} \frac{t^j}{j!} A^j \right) y_0.$$

Démonstration : Pour tout réel $\alpha > 0$, on note y_α la solution de (7.2) sur $[-\alpha, \alpha]$.

Si $0 < \alpha < \beta$, alors y_α et la restriction de y_β à $[-\alpha, \alpha]$ sont solutions du même problème de Cauchy. Une telle solution étant unique on déduit que les fonctions y_α et y_β coïncident sur $[-\alpha, \alpha]$. On peut donc définir sur $\mathbb{R}$ la fonction y par :

$$\forall t \in \mathbb{R},\ y(t) = y_\alpha(t) = \lim_{k \to +\infty} \left(\sum_{j=0}^{k} \frac{t^j}{j!} A^j \right) y_0,$$

où $\alpha > 0$ est tel que $t \in [-\alpha, \alpha]$. Cette fonction est bien solution de (7.3) sur $\mathbb{R}$. $\blacksquare$

Corollaire 7.1 : *L'ensemble S_A des solutions du système différentiel $y' = Ay$ est un sous-espace vectoriel de dimension n de $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^n)$.*

Si $\{e_1, e_2, \ldots, e_n\}$ est une base de $\mathbb{C}^n$ et si, pour tout entier k compris entre 1 et n, la fonction y_k est la solution du problème de Cauchy :

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(0) = e_k, \end{cases}$$

alors $\{y_1, y_2, \ldots, y_n\}$ est une base de S_A.

Démonstration : Il est clair que l'ensemble S_A des solutions sur $\mathbb{R}$ du système différentiel $y' = Ay$ est un sous-espace vectoriel de $\mathcal{C}^\infty(\mathbb{R}, \mathbb{C}^n)$. Le théorème 7.1 se traduit en disant que l'application $y \longmapsto y(0)$ réalise un isomorphisme de S_A sur $\mathbb{C}^n$. Donc S_A est de dimension n.

Pour montrer que le système $\{y_1, y_2, \ldots, y_n\}$ est une base de S_A, il suffit de montrer qu'il est libre.

L'égalité $\sum_{k=1}^{n} \lambda_k y_k = 0$ équivaut à $\sum_{k=1}^{n} \lambda_k y_k(t) = 0$ pour tout réel t et $t = 0$ donne $\sum_{k=1}^{n} \lambda_k e_k = 0$, ce qui entraîne $\lambda_k = 0$ pour tout $k \in \{1, 2, \ldots, n\}$ puisque $\{e_1, e_2, \ldots, e_n\}$ est libre. Le système $\{y_1, y_2, \ldots, y_n\}$ est donc une base de S_A. ■

2. L'exponentielle d'une matrice

Théorème 7.2 : *Pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, on peut définir une unique fonction de classe $\mathcal{C}^\infty$, $E_A : \mathbb{R} \to \mathcal{M}_n(\mathbb{C})$, telle que :*

$$\begin{cases} E_A(0) = I_n, \\ \forall t \in \mathbb{R}, \quad E'_A(t) = AE_A(t). \end{cases} \tag{7.4}$$

Cette fonction est définie par :

$$\forall t \in \mathbb{R}, \quad E_A(t) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} A^k.$$

Démonstration : On note $\{e_1, e_2, \ldots, e_n\}$ la base canonique de $\mathbb{C}^n$ et on définit la fonction matricielle $E_A = (C_1, C_2, \ldots, C_n)$ en disant que la colonne numéro j de E_A, notée C_j, est la solution sur $\mathbb{R}$ du problème de Cauchy :

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(0) = e_j, \end{cases} \tag{7.5}$$

ce qui se traduit matriciellement par (7.4).

Réciproquement, si E est solution de (7.4), alors ses colonnes vérifient le système (7.5) et elles sont uniquement déterminées.

Pour $j \in \{1, 2, \ldots, n\}$, la solution de (7.5) est de classe $\mathcal{C}^\infty$ et est donnée par :

$$C_j(t) = \lim_{p \to +\infty} \left(\sum_{k=0}^{p} \frac{t^k}{k!} A^k \right) e_j.$$

On déduit donc que la fonction E_A est de classe $\mathcal{C}^\infty$ avec, pour tout réel t :

$$E_A(t) = \lim_{p \to +\infty} \left(\sum_{k=0}^{p} \frac{t^k}{k!} A^k \right) = \sum_{k=0}^{+\infty} \frac{t^k}{k!} A^k,$$

la convergence étant uniforme sur tout compact de $\mathbb{R}$. ■

On note, pour tout réel t, $E_A(t) = e^{tA}$ et la matrice

$$e^A = \sum_{k=0}^{+\infty} \frac{1}{k!} A^k$$

est appelée l'exponentielle de la matrice A. La convergence de cette série est uniforme sur tout compact de $\mathcal{M}_n(\mathbb{C})$. En effet, si on note $\|\cdot\|$ une norme sous-multiplicative sur $\mathcal{M}_n(\mathbb{C})$, pour tout compact C de $\mathcal{M}_n(\mathbb{C})$ il existe une constante $M > 0$ telle que $\|A\| \leq M$ pour toute matrice A dans C et, avec :

$$\sum_{k=0}^{+\infty} \frac{1}{k!} \left\|A^k\right\| \leq \sum_{k=0}^{+\infty} \frac{1}{k!} \|A\|^k \leq \sum_{k=0}^{+\infty} \frac{1}{k!} M^k = e^M,$$

on déduit que la série de terme général $\frac{1}{k!} A^k$ est uniformément convergente sur C.

On en déduit alors que l'application $A \mapsto e^A$ est de classe $\mathcal{C}^\infty$ sur $\mathcal{M}_n(\mathbb{C})$.

Le calcul de l'exponentielle d'une matrice est simplifié dans le cas d'une matrice diagonalisable. En effet, si la matrice $A \in \mathcal{M}_n(\mathbb{C})$ est diagonalisable, il existe alors une matrice inversible P et une matrice diagonale :

$$D = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{pmatrix},$$

telles que $A = PDP^{-1}$. En utilisant la continuité du produit matriciel, on déduit facilement que $e^A = Pe^D P^{-1}$ est également diagonalisable avec :

$$e^D = \begin{pmatrix} e^{\lambda_1} & 0 & \dots & 0 \\ 0 & e^{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & e^{\lambda_n} \end{pmatrix}.$$

Ce calcul se simplifie également dans le cas des matrices nilpotentes. En effet, si la matrice A est nilpotente d'ordre p on a :

$$e^A = \sum_{k=0}^{p} \frac{1}{k!} A^k.$$

***Remarque 7.1** : Pour toute matrice A dans $\mathcal{M}_n(\mathbb{C})$, la matrice e^A étant limite d'une suite de polynômes en A, c'est donc un élément de l'adhérence de $\mathbb{C}[A]$ dans $\mathcal{M}_n(\mathbb{C})$. Le sous-espace vectoriel $\mathbb{C}[A]$ étant de dimension finie dans $\mathcal{M}_n(\mathbb{C})$ est fermé et donc confondu avec son adhérence. On déduit donc que e^A est un polynôme en A, c'est-à-dire qu'il existe un polynôme R_A dépendant de A tel que $e^A = R_A(A)$. En fait il n'est pas possible de trouver un polynôme R tel que $e^A = R(A)$ pour toute* *matrice A. En effet, si un tel polynôme existe en notant p son degré on a pour toute matrice A et tout réel t :*

$$e^{tA} = \sum_{k=0}^{p+1} \frac{t^k}{k!} A^k + o\left(t^{p+1}\right) = \sum_{k=0}^{p} r_k t^k A^k$$

et, avec l'unicité du développement limité d'ordre $p+1$, on déduit que nécessairement $A^{p+1} = 0$ pour toute matrice A, ce qui est impossible (prendre par exemple $A = I_n$).

Avec la continuité de l'application $M \longmapsto My_0$ et le théorème 7.1, on déduit que la solution du problème (7.2) est la fonction y définie par :

$$\forall t \in \mathbb{R}, \quad y(t) = e^{tA} y_0.$$

De manière plus générale, pour tout $t_0 \in \mathbb{R}$, le problème de Cauchy :

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(t_0) = y_0, \end{cases} \tag{7.6}$$

admet pour unique solution la fonction y définie par :

$$\forall t \in \mathbb{R}, \quad y(t) = e^{(t-t_0)A} y_0.$$

En effet, la fonction y est solution de (7.6) si et seulement si la fonction z définie par :

$$\forall t \in \mathbb{R}, \quad z(t) = y(t + t_0)$$

est solution de (7.3), et on sait que ce problème admet pour unique solution la fonction z définie par :

$$\forall t \in \mathbb{R}, \quad z(t) = e^{tA} y_0.$$

C'est-à-dire que la solution du problème (7.6) est donnée par :

$$\forall t \in \mathbb{R}, \quad y(t) = z(t - t_0) = e^{(t-t_0)A} y_0.$$

Lemme 7.5 : *Si A, B sont deux matrices qui commutent dans $\mathcal{M}_n(\mathbb{C})$, on a alors :*

$$e^{A+B} = e^A e^B = e^B e^A.$$

Démonstration : On définit les fonctions Y et Z par :

$$\forall t \in \mathbb{R}, \quad Y(t) = e^{t(A+B)}, \quad Z(t) = e^{tA} e^{tB}.$$

Ces fonctions sont dérivables sur $\mathbb{R}$ avec :

$$\forall t \in \mathbb{R}, \quad Y'(t) = (A+B) Y(t), \quad Z'(t) = Ae^{tA} e^{tB} + e^{tA} B e^{tB}.$$

Du fait que A et B commutent et que l'application $X \mapsto XB$ est continue, on déduit que $e^{tA} B = B e^{tA}$. On a donc $Z'(t) = (A+B) Z(t)$. Finalement, avec $Y(0) = Z(0) = I_n$, on déduit que $Y = Z$ (unicité de la solution du problème de Cauchy), c'est-à-dire que $e^{t(A+B)} = e^{tA} e^{tB}$ pour tout réel t. On démontre de manière analogue que $e^{t(A+B)} = e^{tB} e^{tA}$ (on peut aussi utiliser la commutativité de A et B et la continuité du produit matriciel pour montrer que e^A et e^B commutent). ■

Si les matrices A et B ne commutent pas, en général il n'y pas d'égalité entre e^{A+B}, $e^A e^B$ et $e^B e^A$ (exercices 7.1, 7.2 et 7.3).

En utilisant la décomposition de Dunford-Schwarz $A = D + V$, avec D diagonalisable, V nilpotente et $DV = VD$, on a, en notant p l'indice de nilpotence de la matrice V :

$$e^A = e^D e^V = e^D \sum_{k=0}^{p} \frac{1}{k!} V^k,$$

la matrice e^D étant facilement diagonalisable si on sait diagonaliser D.

En fait la décomposition de Dunford-Schwarz de la matrice A permet d'obtenir celle de e^A.

Lemme 7.6 : *Si $A = D + V$ est la décomposition de Dunford-Schwarz de la matrice A, alors celle de e^A est donnée par :*

$$e^A = e^D + e^D \left(e^V - I_n\right),$$

avec e^D diagonalisable et $e^D \left(e^V - I_n\right)$ nilpotente.

Démonstration : Comme les matrices D et V commutent, on a :

$$e^A = e^D e^V = e^D + e^D \left(e^V - I_n\right).$$

On a vu que la matrice e^D est diagonalisable. D'autre part, avec la continuité du produit matriciel on déduit que les matrices V et e^D commutent et on peut écrire que :

$$e^D \left(e^V - I_n\right) = e^D V \sum_{k=1}^{p} \frac{1}{k!} V^{k-1} = V \left(e^D \sum_{k=1}^{p} \frac{1}{k!} V^{k-1}\right),$$

la matrice V commutant avec $e^D \sum\limits_{k=1}^{p} \frac{1}{k!} V^{k-1}$. Il en résulte que la matrice $e^D \left(e^V - I_n\right)$ est nilpotente. Les matrices e^D et $e^D \left(e^V - I_n\right)$ commutant, on a obtenu ainsi la décomposition de Dunford-Schwarz de e^A (cette décomposition est unique). ■

Lemme 7.7 : *L'application* $\exp : A \mapsto e^A$ *est de classe $\mathcal{C}^\infty$ de $\mathcal{M}_n(\mathbb{C})$ dans $GL_n(\mathbb{C})$. Plus précisément, pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$, la matrice e^A est inversible d'inverse e^{-A}.*

Démonstration : Résulte de :

$$I_n = e^{A-A} = e^A e^{-A}.$$

■

Le calcul du déterminant de e^A pour toute matrice A permet de retrouver le caractère inversible de e^A.

Lemme 7.8 : *Pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ on a* $\det\left(e^A\right) = e^{\mathrm{Tr}(A)}$.

Démonstration : Pour $A \in \mathcal{M}_n(\mathbb{C})$ diagonalisable on a $A = PDP^{-1}$ avec P inversible et D diagonale de valeurs propres $\lambda_1, \ldots, \lambda_n$. On a alors $e^A = Pe^D P^{-1}$ et

$$\det\left(e^A\right) = \det\left(e^D\right) = \prod_{k=1}^{n} e^{\lambda_k} = e^{\sum\limits_{k=1}^{n} \lambda_k} = e^{\mathrm{Tr}(A)}.$$

Si A est matrice quelconque dans $\mathcal{M}_n(\mathbb{C})$, on peut l'écrire $A = \lim\limits_{k \to +\infty} A_k$, où $(A_k)_{k\in\mathbb{N}}$ est une suite de matrices diagonalisables. Avec la continuité des applications exponentielles complexe et matricielle et de l'application trace, on déduit que :

$$\det\left(e^A\right) = \lim_{k\to+\infty} \det\left(e^{A_k}\right) = \lim_{k\to+\infty} e^{\mathrm{Tr}(A_k)} = e^{\mathrm{Tr}(A)}.$$

On peut aussi utiliser une trigonalisation de la matrice A. ■

Lemme 7.9 : *Si $\lambda \in \mathbb{C}$ est une valeur propre de la matrice $A \in \mathcal{M}_n(\mathbb{C})$, alors pour tout réel t, $e^{\lambda t}$ est valeur propre de e^{tA} et pour tout vecteur propre associé $x \in \mathbb{C}^n$, la solution du problème de Cauchy :*

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(0) = x, \end{cases} \tag{7.7}$$

est la fonction y définie par :

$$\forall t \in \mathbb{R}, \quad y(t) = e^{\lambda t} x.$$

Démonstration : Soit $x \in \mathbb{C}^n - \{0\}$ un vecteur propre associé à la valeur propre λ de A. Pour tout $k \in \mathbb{N}$, on a $A^k x = \lambda^k x$ et, avec la continuité de $M \longmapsto Mx$, $e^{tA}x = \sum\limits_{k=0}^{+\infty} \frac{t^k}{k!} A^k x = \sum\limits_{k=0}^{+\infty} \frac{t^k}{k!} \lambda^k x = e^{t\lambda} x$. C'est-à-dire que x est un vecteur propre de e^{tA} associé à la valeur propre $e^{t\lambda}$.

La solution du problème (7.7) est $y(t) = e^{tA}x = e^{t\lambda}x$. ■

Théorème 7.3 : *Si $A \in \mathcal{M}_n(\mathbb{C})$ est diagonalisable de valeurs propres deux à deux distinctes $\lambda_1, \ldots, \lambda_p$ $(1 \leq p \leq n)$, alors les solutions du système différentiel $y' = Ay$ sont les fonctions définies par :*

$$\forall t \in \mathbb{R}, \quad y(t) = \sum_{k=1}^{p} e^{\lambda_k t} C_k,$$

où les C_k sont des vecteurs de $\mathbb{C}^n$.

Démonstration : On désigne par $\{\mu_1, \ldots, \mu_n\}$ l'ensemble des valeurs propres de A et par $\{e_1, \ldots, e_n\}$ une base de vecteurs propres associée. Avec le corollaire 7.1 et le lemme 7.9 on déduit que $\{e^{t\mu_1}e_1, \ldots, e^{t\mu_n}e_n\}$ est une base de l'espace des solutions de $y' = Ay$, c'est-à-dire que toute solution de ce système s'écrit $y(t) = \sum\limits_{k=1}^{n} \alpha_k e^{\mu_k t} e_k$, ou encore, en regroupant les valeurs propres identiques, que $y(t) = \sum\limits_{k=1}^{p} e^{\lambda_k t} C_k$ avec C_k dans $\ker(A - \lambda_k I_n)$ pour tout $k \in \{1, \ldots, p\}$. ■

Lemme 7.10 : *Soient $A \in \mathcal{M}_n(\mathbb{C})$ admettant une seule valeur propre $\lambda \in \mathbb{C}$ et $\pi_A(X) = (X - \lambda)^\alpha$ le polynôme minimal de A avec $1 \leq \alpha \leq n$. Les solutions du système différentiel $y' = Ay$ sont les fonctions définies par :*

$$\forall t \in \mathbb{R}, \quad y(t) = e^{\lambda t} \sum_{j=0}^{\alpha-1} \frac{t^j}{j!} (A - \lambda I_n)^j y_0,$$

où $y_0 \in \mathbb{C}^n$. Ce qui revient à dire que $e^{tA} = e^{\lambda t} \sum\limits_{k=0}^{\alpha-1} \frac{t^k}{k!} (A - \lambda I_n)^k$.

Démonstration : Toute solution de $z' = (A - \lambda I_n) z$ s'écrit :

$$z(t) = e^{t(A-\lambda I_n)} z_0 = \sum_{k=0}^{+\infty} \frac{t^k}{k!} (A - \lambda I_n)^k z_0 = \sum_{k=0}^{\alpha-1} \frac{t^k}{k!} (A - \lambda I_n)^k z_0,$$

du fait que $(A - \lambda I_n)$ est nilpotente d'ordre α.

On en déduit que toute solution de $y' = Ay$ s'écrit :

$$y(t) = e^{tA} y_0 = e^{t(A-\lambda I_n)} e^{t\lambda I_n} y_0 = e^{\lambda t} \sum_{k=0}^{\alpha-1} \frac{t^k}{k!} (A - \lambda I_n)^k y_0.$$

■

Théorème 7.4 : *Soit $A \in \mathcal{M}_n(\mathbb{C})$ de polynôme minimal $\pi_A(X) = \prod\limits_{k=1}^{p} (X - \lambda_k)^{\alpha_k}$, les racines $\lambda_1, \ldots, \lambda_p$ étant deux à deux distinctes dans $\mathbb{C}$. Les solutions du système différentiel $y' = Ay$ sont les fonctions définies par :*

$$\forall t \in \mathbb{R}, \quad y(t) = \sum_{k=1}^{p} e^{\lambda_k t} P_k(t),$$

où, pour tout $k \in \{1, \ldots, p\}$, P_k désigne une fonction polynomiale de degré inférieur ou égal à $\alpha_k - 1$ et à valeurs dans $\mathbb{C}^n$.

Démonstration : Pour tout $k \in \{1, \ldots, p\}$ on note $E_k = \ker (A - \lambda_k I_n)^{\alpha_k}$ le sous-espace caractéristique associé à la valeur propre λ_k. D'après le théorème de décomposition des noyaux, on a $\mathbb{C}^n = \bigoplus\limits_{k=1}^{p} E_k$.

Soit y solution de $y' = Ay$. On peut écrire, de manière unique :

$$y(0) = \sum_{k=1}^{p} x_k,$$

avec $x_k \in E_k$ pour tout $k \in \{1, \ldots, p\}$. Si, pour $k \in \{1, \ldots, p\}$, on désigne par y_k la solution du problème de Cauchy :

$$\begin{cases} y'(t) = Ay(t) \quad (t \in \mathbb{R}), \\ y(0) = x_k, \end{cases}$$

alors $z = \sum_{k=1}^{p} y_k$ est solution du même problème de Cauchy que y. On a donc, du fait de l'unicité de cette solution, $y = \sum_{k=1}^{p} y_k$. Mais, d'autre part, on a :

$$y_k(t) = e^{tA}x_k = e^{\lambda_k t}e^{t(A-\lambda_k I_n)}x_k = e^{\lambda_k t}\sum_{j=0}^{\alpha_k-1}\frac{t^j}{j!}(A-\lambda_k I_n)^j x_k,$$

du fait que $(A-\lambda_k I_n)^j x_k = 0$ pour tout $j > \alpha_k$. Ce qui donne, en notant :

$$P_k(t) = \sum_{j=0}^{\alpha_k-1}\frac{t^j}{j!}(A-\lambda_k I_n)^j x_k,$$

$y(t) = \sum_{k=1}^{p} e^{\lambda_k t}P_k(t)$. ∎

Dans la démonstration précédente, le vecteur x_k est la projection de $y(0)$ sur $\ker(A-\lambda_k I_n)^{\alpha_k}$. En notant π_k cette projection (projecteur spectral), on en déduit que :

$$e^{tA} = \sum_{k=1}^{p} e^{\lambda_k t}\left(\sum_{j=0}^{\alpha_k-1}\frac{t^j}{j!}(A-\lambda_k I_n)^j\right)\pi_k.$$

On peut retrouver ce résultat en utilisant la décomposition de Dunford-Schwarz (paragraphe 6, page 58). En effet la matrice A s'écrit $A = D + V$, avec D diagonalisable, V nilpotente et $DV = VD$. On a alors $e^A = e^D e^V$. Puis, en écrivant que :

$$\forall r \geq 0, \begin{cases} D^r = \sum_{k=1}^{p} \lambda_k^r \pi_k, \\ V^r = \sum_{k=1}^{p} (A-\lambda_k I_n)^r \pi_k, \end{cases}$$

on obtient :

$$e^D = \sum_{r=0}^{+\infty}\frac{1}{r!}D^r = \sum_{r=0}^{+\infty}\frac{1}{r!}\left(\sum_{k=1}^{p}\lambda_k^r\pi_k\right) = \sum_{k=1}^{p}\left(\sum_{r=0}^{+\infty}\frac{\lambda_k^r}{r!}\right)\pi_k = \sum_{k=1}^{p}e^{\lambda_k}\pi_k,$$

et

$$\begin{cases} e^V = \sum_{r=0}^{+\infty}\frac{1}{r!}V^r = \sum_{r=0}^{+\infty}\frac{1}{r!}\left(\sum_{k=1}^{p}(A-\lambda_k I_n)^r\pi_k\right) \\ \quad = \sum_{k=1}^{p}\left(\sum_{r=0}^{\alpha_k-1}\frac{1}{r!}(A-\lambda_k I_n)^r\pi_k\right) = \sum_{k=1}^{p}\left(\sum_{r=0}^{\alpha_k-1}\frac{1}{r!}(A-\lambda_k I_n)^r\right)\pi_k \end{cases}$$

(la restriction de $A-\lambda_k I_n$ à $\ker(A-\lambda_k I_n)^{\alpha_k}$ est nilpotente d'indice α_k).

Enfin, avec les propriétés des projecteurs, on déduit que :

$$e^A = \sum_{k=1}^{p}e^{\lambda_k}\left(\sum_{r=0}^{\alpha_k-1}\frac{1}{r!}(A-\lambda_k I_n)^r\right)\pi_k.$$

3. Un algorithme de calcul de l'exponentielle d'une matrice

Soit $A \in \mathcal{M}_n(\mathbb{C})$ de valeurs propres $\lambda_1, \ldots, \lambda_n$ dans $\mathbb{C}$. On lui associe la suite de matrices $(B_k)_{1\leq k\leq n}$ définie par :

$$\begin{cases} B_1 = I_n, \\ B_{k+1} = (A - \lambda_k I_n) B_k \quad (1 \leq k \leq n-1). \end{cases}$$

On a alors $(A - \lambda_n I_n) B_n = 0$. En effet, on a $B_k = \prod_{j=1}^{k-1} (A - \lambda_j I_n)$ pour tout $k \in \{1, \ldots, n\}$ et pour $k = n$:

$$(A - \lambda_n I_n) B_n = \prod_{j=1}^{n} (A - \lambda_j I_n) = P_A(A) = 0,$$

d'après le théorème de Cayley-Hamilton.

Théorème 7.5 : *Avec les notations qui précèdent, si* $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ *est la solution du problème de Cauchy :*

$$\begin{cases} y'(t) = Jy(t) \quad (t \in \mathbb{R}), \\ y(0) = e_1, \end{cases} \tag{7.8}$$

où

$$J = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 1 & \lambda_2 & 0 & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & 1 & \lambda_{n-1} & 0 \\ 0 & \ldots & 0 & 1 & \lambda_n \end{pmatrix}, \quad e_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

on a alors :

$$\forall t \in \mathbb{R}, \quad e^{tA} = \sum_{k=1}^{n} y_k(t) B_k.$$

Démonstration : Il suffit de montrer que la fonction $y = \sum\limits_{k=1}^{n} y_k B_k$ est solution du même problème de Cauchy que e^{tA}.

Pour $t = 0$ on a $y(0) = \sum\limits_{k=1}^{n} y_k(0) B_k = B_1 = I_n$ et pour tout $t \in \mathbb{R}$:

$$\begin{aligned} y'(t) &= \sum_{k=1}^{n} y_k'(t) B_k = \lambda_1 y_1(t) + \sum_{k=2}^{n} \big(y_{k-1}(t) + \lambda_k y_k(t)\big) B_k \\ &= \sum_{k=1}^{n} \lambda_k y_k(t) B_k + \sum_{k=1}^{n-1} y_k(t) B_{k+1}, \end{aligned}$$

avec :

$$\begin{aligned} \sum_{k=1}^{n-1} y_k(t) B_{k+1} &= \sum_{k=1}^{n-1} y_k(t) \left(A - \lambda_k I_n\right) B_k \\ &= A \sum_{k=1}^{n-1} y_k(t) B_k - \sum_{k=1}^{n-1} \lambda_k y_k(t) B_k. \end{aligned}$$

Ce qui donne en définitive :

$$\begin{aligned} y'(t) &= \lambda_n y_n(t) B_n + A y(t) - y_n(t) A B_n \\ &= A y(t) - y_n(t) \left(A - \lambda_n I_n\right) B_n = A y(t). \end{aligned}$$

On peut donc conclure que $e^{tA} = \sum\limits_{k=1}^{n} y_k(t) B_k$. ∎

4. Équations différentielles linéaires d'ordre n

On considère l'équation différentielle d'ordre n linéaire à coefficients complexes constants :

$$y^{(n)}(t) = a_1 y(t) + a_2 y'(t) + \cdots + a_n y^{(n-1)}(t) \quad (t \in \mathbb{R}). \tag{7.9}$$

Cette équation différentielle peut se ramener à un système différentiel linéaire $Y' = AY$ de matrice :

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & \ddots & \ddots & . \\ 0 & 0 & \dots & 0 & 1 \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix}.$$

En effet, si $y \in \mathcal{C}^n(\mathbb{R}, \mathbb{C})$ est une solution de l'équation différentielle (7.9), alors la fonction vectorielle $Y \in \mathcal{C}^1(\mathbb{R}, \mathbb{C}^n)$ définie par :

$$\forall t \in \mathbb{R}, \quad Y(t) = \begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix} = \begin{pmatrix} y(t) \\ y'(t) \\ \vdots \\ y^{(n-1)}(t) \end{pmatrix}$$

est solution de :

$$\forall t \in \mathbb{R}, \quad Y'(t) = \begin{pmatrix} y_2(t) \\ \vdots \\ y_n(t) \\ a_1 y_1(t) + a_2 y_2(t) + \cdots + a_n y_n(t) \end{pmatrix} = AY(t),$$

où la matrice A est définie par :

$$A = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & \ddots & \ddots & . \\ 0 & 0 & \dots & 0 & 1 \\ a_1 & a_2 & \dots & a_{n-1} & a_n \end{pmatrix}.$$

Réciproquement, soit $Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{C}^1(\mathbb{R}, \mathbb{C}^n)$ solution du système différentiel $Y' = AY$. En posant $y = y_1$, on a :

$$y' = y_2, \dots, y^{(n-1)} = y_n \text{ et } y^{(n)} = a_1 y_1 + a_2 y_2 + \cdots + a_n y_n.$$

C'est-à-dire que y est solution de (7.9).

Théorème 7.6 : *Les solutions de l'équation différentielle (7.9) sont de la forme*

$$y(t) = \sum_{k=1}^{p} e^{\lambda_k t} P_k(t),$$

où $\lambda_1, \dots, \lambda_p$ sont les racines deux à deux distinctes du polynôme :

$$\pi_A(X) = X^n - \sum_{k=1}^{n} a_k X^{k-1}$$

de multiplicités respectives $\alpha_1, \dots, \alpha_p$ et chaque P_k est une fonction polynomiale de degré inférieur ou égal à $\alpha_k - 1$.

Démonstration : On sait que toute solution du système différentiel $Y' = AY$ s'écrit :

$$Y(t) = \sum_{k=1}^{p} e^{\lambda_k t} H_k(t),$$

où $\lambda_1, \ldots, \lambda_p$ sont les valeurs propres deux à deux distinctes de A et, pour tout $k \in \{1, \ldots, p\}$, H_k est un polynôme à coefficients dans $\mathbb{C}^n$ de degré strictement inférieur à la multiplicité α_k de λ_k comme racine du polynôme minimal de A.

Avec les notations de ce paragraphe, le polynôme minimal de la matrice A est défini par :

$$\pi_A(X) = X^n - \sum_{k=1}^{n} a_k X^{k-1}$$

(exercice 2.12). On déduit donc que toute solution de l'équation différentielle (7.9) est de la forme :

$$y(t) = \sum_{k=1}^{p} e^{\lambda_k t} P_k(t),$$

où $\lambda_1, \ldots, \lambda_p$ sont les racines deux à deux distinctes du polynôme π_A et, pour tout $k \in \{1, \ldots, p\}$, P_k est un polynôme de degré strictement inférieur à α_k. ■

Le polynôme π_A est appelé le polynôme caractéristique de l'équation différentielle d'ordre n (7.9).

5. Systèmes différentiels linéaires à coefficients non constants

Soient $[\alpha, \beta]$ un intervalle réel fermé borné, $t_0 \in [\alpha, \beta]$, $K : [\alpha, \beta]^2 \to \mathbb{C}^n$ et $h : [\alpha, \beta] \to \mathbb{C}^n$ deux applications continues. On désigne par $E = \mathcal{C}^0([\alpha, \beta], \mathbb{C}^n)$ l'espace des fonctions continues de $[\alpha, \beta]$ dans $\mathbb{C}^n$ muni de la norme de la convergence uniforme et on définit l'application $\varphi : E \to E$ par :

$$\forall y \in E, \quad \forall t \in [\alpha, \beta], \ \varphi(y)(t) = \int_{t_0}^{t} K(t, u)\, y(u)\, du + h(t).$$

Lemme 7.11 : *Avec les notations qui précèdent, il existe un entier p strictement positif tel que l'application $\varphi^p = \underbrace{\varphi \circ \cdots \circ \varphi}_{p \text{ fois}}$ soit contractante. En conséquence, l'application φ admet un unique point fixe dans E.*

Démonstration : Il est clair que pour toute fonction y continue sur $[\alpha, \beta]$ la fonction $\varphi(y)$ est également continue sur $[\alpha, \beta]$.

Pour y, z, dans E et t dans $[\alpha, \beta]$ on a :

$$\begin{aligned}\|\varphi(y)(t) - \varphi(z)(t)\| &= \left\| \int_{t_0}^{t} K(t,u)\left(y(u) - z(u)\right) du \right\| \\ &\leq \lambda \left| \int_{t_0}^{t} \|y(u) - z(u)\| \, du \right| \\ &\leq \lambda \|y - z\|_\infty |t - t_0|,\end{aligned}$$

où $\lambda = \sup\limits_{(t,u)\in[\alpha,\beta]^2} \|K(t,u)\|$. Par récurrence sur $k \geq 1$, on déduit que :

$$\forall t \in [\alpha, \beta], \ \left\|\varphi^k(y)(t) - \varphi^k(z)(t)\right\| \leq \frac{(\lambda |t - t_0|)^k}{k!} \|y - z\|_\infty .$$

En effet, on vient de voir que le résultat est vrai pour $k = 1$ et, en le supposant acquis pour $k \geq 1$, on a pour tout $t \in [\alpha, \beta]$:

$$\begin{aligned}\left\|\varphi^{k+1}(y)(t) - \varphi^{k+1}(z)(t)\right\| &\leq \lambda \left| \int_{t_0}^{t} \left\|\varphi^k(y)(u) - \varphi^k(z)(u)\right\| \, du \right| \\ &\leq \lambda \frac{\lambda^k}{k!} \|y - z\|_\infty \left| \int_{t_0}^{t} |u - t_0|^k \, du \right|,\end{aligned}$$

avec :

$$\left| \int_{t_0}^{t} |u - t_0|^k \, du \right| = \frac{|t - t_0|^{k+1}}{k+1} .$$

Ce qui donne :

$$\left\|\varphi^{k+1}(y)(t) - \varphi^{k+1}(z)(t)\right\| \leq \frac{(\lambda |t - t_0|)^{k+1}}{(k+1)!} \|y - z\|_\infty .$$

On déduit alors que :

$$\forall k \geq 1, \ \left\|\varphi^k(y) - \varphi^k(z)\right\|_\infty \leq \frac{\left(\lambda (\beta - \alpha)\right)^k}{k!} \|y - z\|_\infty$$

et, avec $\lim\limits_{k\to+\infty} \dfrac{\left(\lambda(\beta - \alpha)\right)^k}{k!} = 0$, on déduit que pour $k \geq 1$ assez grand l'application φ^k est contractante.

Le théorème du point fixe itéré (l'espace $(E, \|\cdot\|_\infty)$ est complet) nous dit alors que φ admet un unique point fixe dans E. ■

Lemme 7.12 : *Soient $[\alpha, \beta]$ un intervalle réel fermé borné et*

$$A : [\alpha, \beta] \to \mathcal{M}_n(\mathbb{C}), \quad b : [\alpha, \beta] \to \mathbb{C}^n,$$

deux applications continues. Pour tout $t_0 \in [\alpha, \beta]$ et tout $y_0 \in \mathbb{C}^n$ le problème de Cauchy :

$$\begin{cases} y'(t) = A(t)y(t) + b(t) \quad (t \in [\alpha, \beta]), \\ y(t_0) = y_0, \end{cases} \tag{7.10}$$

admet une unique solution.

Démonstration : Dire que $y \in \mathcal{C}^1([\alpha, \beta], \mathbb{C}^n)$ est solution de (7.10) équivaut à dire que pour tout $t \in [\alpha, \beta]$ on a :

$$y(t) = \int_{t_0}^{t} y'(u)\, du + y(t_0) = \int_{t_0}^{t} A(u)\, y(u)\, du + \int_{t_0}^{t} b(u)\, du + y_0,$$

c'est-à-dire que $y = \varphi(y)$ où φ est définie par :

$$K : (t, u) \mapsto A(u), \quad h : t \mapsto \int_{t_0}^{t} b(u)\, du + y_0.$$

On déduit alors de ce qui précède que (7.10) a une unique solution. ∎

Dans ce qui suit I est un intervalle réel et $A : I \to \mathcal{M}_n(\mathbb{C})$, $b : I \to \mathbb{C}^n$ sont deux applications continues.

Théorème 7.7 : *Pour tout $t_0 \in I$ et tout $y_0 \in \mathbb{C}^n$ le problème de Cauchy :*

$$\begin{cases} y'(t) = A(t)y(t) + b(t) \quad (t \in I), \\ y(t_0) = y_0, \end{cases} \tag{7.11}$$

admet une unique solution.

Démonstration : Pour tout intervalle fermé borné $J \subset I$ contenant t_0, on note y_J la solution de (7.10) sur J. Si J_1 et J_2 sont deux tels intervalles, du fait de l'unicité de la solution du problème de Cauchy sur un intervalle compact on déduit que $y_{J_1}(t) = y_{J_2}(t)$ pour tout $t \in J_1 \cap J_2$. On peut donc définir sur I la fonction y par $y(t) = y_J(t)$ où J est n'importe quel intervalle compact contenu dans I et tel que $t_0 \in J$. Il est alors clair que y est solution de (7.11) et que ce problème admet une unique solution dans $\mathcal{C}^1(I, \mathbb{C}^n)$. ∎

Corollaire 7.2 : *L'ensemble $\mathcal{S}_I$ des solutions sur l'intervalle I du système différentiel $y'(t) = A(t)y(t)$ est un espace vectoriel de dimension n sur $\mathbb{C}$.*

Un ensemble de solutions $\{y_1, \ldots, y_n\}$ est une base de $\mathcal{S}_I$ si et seulement si il existe $t_0 \in I$ tel que le système de vecteurs $\{y_1(t_0), \ldots, y_n(t_0)\}$ soit une base de $\mathbb{C}^n$.

Démonstration : Du théorème 7.7 avec $b = 0$, on déduit que $\mathcal{S}_I$ est non vide et que, pour tout $t_0 \in I$, l'application linéaire $\psi : y \mapsto y(t_0)$ réalise un isomorphisme de $\mathcal{S}_I$ sur $\mathbb{C}^n$. Il en résulte que $\mathcal{S}_I$ est un espace vectoriel de dimension n sur $\mathbb{C}$.

Si le système $\{y_1(t_0), \ldots, y_n(t_0)\}$ est une base de $\mathbb{C}^n$ alors son image réciproque $\{y_1, \ldots, y_n\}$ par l'isomorphisme ψ est une base de $\mathcal{S}_I$. ∎

Remarque 7.2 : *Ce résultat peut se traduire en disant que si l'application $w : t \mapsto \det(y_1(t), \ldots, y_n(t))$ s'annule en un point alors elle est identiquement nulle.*

Corollaire 7.3 : *Soit u une solution particulière sur l'intervalle I du système différentiel avec second membre :*

$$y'(t) = A(t)y(t) + b(t). \tag{7.12}$$

Toute autre solution de ce système s'écrit $z = u + y$ où y est solution du système différentiel sans second membre (ou système homogène) $y'(t) = A(t)y(t)$ sur l'intervalle I.

Démonstration : Il est clair que pour toute solution z de (7.12) la fonction $y = z - u$ est solution de $y' = Ay$. On déduit que si $\{y_1, \ldots, y_n\}$ est une base de solutions de $y' = Ay$ sur I alors toute solution de $y' = Ay + b$ s'écrit :

$$z = \sum_{k=1}^{n} \lambda_k y_k + u,$$

où les scalaires λ_k sont uniquement déterminés par $z(t_0)$, le point t_0 étant donné dans I. ∎

Définition 7.1 : *Soit $\{y_1, \ldots, y_n\}$ une base de l'espace vectoriel des solutions sur l'intervalle I du système différentiel $y'(t) = A(t)y(t)$. On appelle wronskien de cet ensemble de solution, la fonction :*

$$\begin{array}{rccl} w: & I & \rightarrow & \mathbb{C} \\ & t & \longmapsto & w(t) = \det\big(y_1(t), \ldots, y_n(t)\big). \end{array}$$

Théorème 7.8 : *Avec les notations qui précèdent, le wronskien w est solution de l'équation différentielle :*

$$w'(t) = \mathrm{Tr}\big(A(t)\big)w(t) \quad (t \in I).$$

Pour tout $t_0 \in I$, on a $w(t) = w(t_0)\, e^{g(t)}$ $(t \in I)$, où :

$$g(t) = \int_{t_0}^{t} \mathrm{Tr}\big(A(x)\big)\, \mathrm{d}x.$$

Le wronskien d'une base de solutions sur I du système $y'(t) = A(t)y(t)$ ne s'annule jamais.

Démonstration : Les y_j étant de classe $\mathcal{C}^1$ et l'application déterminant de classe $\mathcal{C}^\infty$ on déduit que w est de classe $\mathcal{C}^1$ avec :

$$\begin{aligned} w' &= \sum_{j=1}^{n} \det\left(y_1, \ldots, y_{j-1}, y'_j, \ldots, y_n\right) \\ &= \sum_{j=1}^{n} \det\left(y_1, \ldots, y_{j-1}, Ay_j, \ldots, y_n\right) \end{aligned}$$

(exercice 3.5).

Pour tout $t \in I$ l'application :

$$\varphi_t : (y_1, \ldots, y_n) \mapsto \sum_{j=1}^{n} \det\bigl(y_1, \ldots, y_{j-1}, A(t)y_j, \ldots, y_n\bigr)$$

étant n-linéaire alternée sur $(\mathbb{C}^n)^n$ est proportionnelle au déterminant, c'est-à-dire qu'il existe une constante $\lambda(t) \in \mathbb{C}$ telle que :

$$\forall Y = (y_1, \ldots, y_n) \in (\mathbb{C}^n)^n, \quad \varphi_t(Y) = \lambda(t) \det(Y).$$

Cette constante est égale à $\varphi_t(I_n)$, soit, en notant $\{e_1, \ldots, e_n\}$ la base canonique de $\mathbb{C}^n$:

$$\begin{aligned} \lambda(t) &= \sum_{j=1}^{n} \det\bigl(e_1, \ldots, e_{j-1}, A(t)\, e_j, \ldots, e_n\bigr) \\ &= \sum_{j=1}^{n} a_{jj}(t) = \operatorname{Tr}\bigl(A(t)\bigr). \end{aligned}$$

On a donc en définitive :

$$\forall t \in I, \quad w'(t) = \operatorname{Tr}\bigl(A(t)\bigr) w(t).$$

Soit, pour t_0 fixé dans I, $g : t \mapsto \displaystyle\int_{t_0}^{t} \operatorname{Tr}\bigl(A(x)\bigr)\, dx$ la primitive de $\operatorname{Tr}(A)$ nulle en t_0.

On a $w'(t) = g'(t)w(t)$ et $\left(w(t)e^{-g(t)}\right)' = 0$ pour tout $t \in I$. On déduit donc qu'il existe une constante $C \in \mathbb{C}$ telle que $w(t) = Ce^{g(t)}$ et $t = t_0$ donne $C = w(t_0)$. On a donc :

$$\forall t \in I, \quad w(t) = w(t_0)\, e^{g(t)}.$$

On sait que $\{y_1, \ldots, y_n\}$ est une base de solutions si et seulement si le système de vecteurs $\{y_1(t_0), \ldots, y_n(t_0)\}$ est une base de $\mathbb{C}^n$, ce qui équivaut à dire que $w(t_0) \neq 0$, qui est encore équivalent à $w(t) \neq 0$ pour tout $t \in I$. On a donc ainsi montré que le wronskien d'une base de solutions ne s'annule jamais sur I. ■

6. Méthode de variation des constantes

Théorème 7.9 : *Si le système $\{y_1, \ldots, y_n\}$ est une base de l'espace vectoriel des solutions du système $y'(t) = A(t)y(t)$ sur l'intervalle I, alors la fonction y définie sur I par :*

$$y(t) = \sum_{i=1}^{n} \lambda_i(t) y_i(t)$$

est solution particulière sur l'intervalle I du système différentiel avec second membre $y'(t) = A(t)\,y(t) + b(t)$ si et seulement si, pour tout $t \in I$, le vecteur $\left(\lambda_1'(t), \ldots, \lambda_n'(t)\right)$ est solution du système linéaire :

$$\sum_{i=1}^{n} \lambda_i'(t) y_i(t) = b(t).$$

Démonstration : $y = \sum_{i=1}^{n} \lambda_i y_i$ est solution particulière sur l'intervalle I du système différentiel $y' = Ay + b$ si et seulement si :

$$\sum_{i=1}^{n} \lambda_i' y_i + \sum_{i=1}^{n} \lambda_i y_i' = Ay + b,$$

avec $\sum_{i=1}^{n} \lambda_i y_i' = \sum_{i=1}^{n} \lambda_i A y_i = Ay$. Ce qui équivaut à dire que pour tout $t \in I$ le vecteur $\left(\lambda_1'(t), \ldots, \lambda_n'(t)\right)$ est solution du système linéaire :

$$\sum_{i=1}^{n} \lambda_i'(t) y_i(t) = b(t).$$

Le déterminant de ce système étant le wronskien de la base de solutions $\{y_1, \ldots, y_n\}$, il n'est donc jamais nul et on a une unique solution. ∎

Corollaire 7.4 : *Soient $p, q, r : I \to \mathbb{C}$ continues.*

L'ensemble des solutions sur l'intervalle I de l'équation différentielle d'ordre 2 :

$$y'' = py' + qy \tag{7.13}$$

est un espace vectoriel de dimension 2 sur $\mathbb{C}$.

Si $\{u, v\}$ est une base de solutions sur l'intervalle I de l'équation différentielle $y'' = py' + qy$, alors la fonction y définie par $y(t) = \lambda(t)\,u(t) + \mu(t)v(t)$ est solution sur l'intervalle I de l'équation différentielle avec second membre $y'' = py' + qy + r$ si, pour tout $t \in I$, $\left(\lambda'(t), \mu'(t)\right)$ est solution du système linéaire de deux équations à deux inconnues :

$$\begin{cases} \lambda'(t)u(t) + \mu'(t)\,v(t) = 0, \\ \lambda'(t)u'(t) + \mu'(t)v'(t) = r(t). \end{cases}$$

Démonstration : La fonction y est solution de (7.13) si et seulement si la fonction $z = \begin{pmatrix} y \\ y' \end{pmatrix}$ est solution du système différentiel $z' = Az$ où $A = \begin{pmatrix} 0 & 1 \\ q & p \end{pmatrix}$ est continue de I dans $M_2(\mathbb{C})$. On déduit alors que l'ensemble des solutions de (7.13) est un espace vectoriel sur $\mathbb{C}$ de dimension 2.

Si $\{u, v\}$ est une base de solutions sur I de $y'' = py' + qy$ alors le couple $\left\{ z = \begin{pmatrix} u \\ u' \end{pmatrix}, w = \begin{pmatrix} v \\ v' \end{pmatrix} \right\}$ est une base de solutions sur I du système différentiel $z' = Az$ (il est facile de vérifier que le système $\{z, w\}$ est libre).

La fonction $Z = \lambda z + \mu w$ est solution de $Z' = AZ + b$ si et seulement si pour tout $t \in I$ le vecteur $\begin{pmatrix} \lambda'(t) \\ \mu'(t) \end{pmatrix}$ est solution de $\lambda'(t)z(t) + \mu'(t)w(t) = b(t)$, soit :

$$\begin{cases} \lambda'(t)u(t) + \mu'(t)\,v(t) = 0, \\ \lambda'(t)u'(t) + \mu'(t)v'(t) = r(t). \end{cases}$$

Dans ce cas on a alors, en posant $y = \lambda u + \mu v$:

$$Z = \lambda z + \mu w = \begin{pmatrix} \lambda u + \mu v \\ \lambda u' + \mu v' \end{pmatrix} = \begin{pmatrix} y \\ y' \end{pmatrix}$$

et

$$Z' = \begin{pmatrix} y' \\ y'' \end{pmatrix} = AZ + b = \begin{pmatrix} y' \\ py' + qy + r \end{pmatrix}.$$

C'est-à-dire que la fonction $y = \lambda u + \mu v$ est solution sur I de $y'' = py' + qy + r$. ∎

Remarque 7.3 *: Le wronskien du système différentiel $z' = Az$ est $w(t) = w(t_0)\,e^{g(t)}$ où $g(t) = \int_{t_0}^{t} \mathrm{Tr}\big(A(x)\big)\, \mathrm{d}x = \int_{t_0}^{t} p(x)\, \mathrm{d}x$. Et en considérant que ce wronskien est $w(t) = \det\big(z(t), w(t)\big)$ on déduit que :*

$$u(t)v'(t) - u'(t)v(t) = w(t_0)\,e^{g(t)}.$$

Ce résultat peut être utilisé pour calculer une solution de base v connaissant une solution u.

7. Surjectivité et injectivité de l'exponentielle matricielle

On a vu que pour toute matrice $A \in \mathcal{M}_n(\mathbb{K})$ la matrice e^A est inversible (lemme 7.7) et avec $\det\left(e^A\right) = e^{\mathrm{Tr}(A)}$, on déduit que si A est à coefficients réels alors e^A est dans l'ensemble $GL_n^+(\mathbb{R})$ des matrices réelles d'ordre n de déterminant strictement positif.

On note $\mathcal{N}_n(\mathbb{K})$ l'ensemble des matrices à coefficients réels ou complexes d'ordre n qui sont nilpotentes.

Définition 7.2 : *On dit qu'une matrice $A \in \mathcal{M}_n(\mathbb{K})$ est unipotente si la matrice $A - I_n$ est nilpotente.*

On note $\mathcal{L}_n(\mathbb{K})$ l'ensemble des matrices à coefficients réels ou complexes d'ordre n qui sont unipotentes.

Lemme 7.13 : *Si $f(z) = \sum\limits_{k=0}^{+\infty} a_k z^k$ est une série entière de rayon de convergence $R > 0$, alors pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ telle que $\rho(A) < R$ la série de terme général $a_k A^k$ est convergente.*

Démonstration : Voir l'exercice 4.14. ∎

***Remarque 7.4** : Avec les notations du lemme précédent, pour $\rho(A) < R$, la matrice $f(A) = \sum\limits_{k=0}^{+\infty} a_k A^k$ est en fait un polynôme en A, dont les coefficients dépendent de A. En effet, $f(A)$ est dans l'adhérence de $\mathbb{C}[A]$ qui est fermé dans $\mathcal{M}_n(\mathbb{C})$ en tant que sous-espace vectoriel d'un espace vectoriel normé de dimension finie.*

En partant du développement en série entière de la fonction logarithme complexe :

$$\ln(1+z) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} z^k \quad (|z| < 1)$$

on peut définir la fonction logarithme matricielle, sur l'ensemble des matrices $A \in \mathcal{M}_n(\mathbb{C})$ telles que $\rho(A) < 1$ par :

$$\ln(I_n + A) = \sum_{k=1}^{+\infty} \frac{(-1)^{k-1}}{k} A^k.$$

En particulier on a $\ln(I_n) = 0$ et pour toute matrice nilpotente A, en notant p un entier naturel tel que $A^{p+1} = 0$, on a :

$$\ln(I_n + A) = \sum_{k=1}^{p} \frac{(-1)^{k-1}}{k} A^k.$$

Lemme 7.14 : *Pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ telle que $\rho(A) < 1$, on a :*

$$e^{\ln(I_n+A)} = I_n + A.$$

Démonstration : Pour tout $x \in \mathbb{R}$, on a :

$$e^x = \sum_{k=0}^{+\infty} \alpha_k x^k,$$

avec $\alpha_k = \dfrac{1}{k!}$ pour tout entier naturel k et pour tout réel x tel que $|x| < 1$, on a :

$$\ln(1+x) = \sum_{j=1}^{+\infty} \beta_j x^j,$$

avec $\beta_j = \dfrac{(-1)^{j-1}}{j}$ pour tout entier naturel j non nul.

On peut alors écrire pour $k \geq 1$ et $|x| < 1$:

$$\left(\ln(1+x)\right)^k = \sum_{j=k}^{+\infty} \beta_{k,j} x^j$$

et

$$\begin{aligned} e^{\ln(1+x)} &= 1 + \sum_{k=1}^{+\infty} \alpha_k \left(\sum_{j=k}^{+\infty} \beta_{k,j} x^j \right) \\ &= 1 + x + \sum_{j=2}^{+\infty} \left(\sum_{k=1}^{j} \alpha_k \beta_{k,j} \right) x^j. \end{aligned}$$

Avec $e^{\ln(1+x)} = 1 + x$, on déduit alors que :

$$\forall k \geq 2, \quad \sum_{k=1}^{j} \alpha_k \beta_{k,j} = 0.$$

En écrivant, pour $A \in \mathcal{M}_n(\mathbb{C})$ telle que $\rho(A) < 1$ que :

$$e^{\ln(I_n + A)} = I_n + A + \sum_{j=2}^{+\infty} \left(\sum_{k=1}^{j} \alpha_k \beta_{k,j} \right) A^j,$$

on déduit que $e^{\ln(I_n+A)} = I_n + A$. ■

Lemme 7.15 : *Pour toute matrice $A \in \mathcal{N}_n(\mathbb{C})$ on a $e^A \in \mathcal{L}_n(\mathbb{C})$ et*

$$\forall t \in \mathbb{R}, \quad \ln\left(e^{tA}\right) = tA.$$

Démonstration : Si $A \in \mathcal{N}_n(\mathbb{C})$, il existe $p \in \mathbb{N}$ tel que $A^{p+1} = 0$ et $e^A = I_n + V$, avec :

$$V = A \sum_{k=1}^{p} \frac{1}{k!} A^{k-1}$$

qui est nilpotente. On a donc $e^A \in \mathcal{L}_n(\mathbb{C})$.

Pour tout réel t, on a également $e^{tA} = I_n + V(t) \in \mathcal{L}_n(\mathbb{C})$ avec :

$$V(t) = \sum_{k=1}^{p} \frac{1}{k!} t^k A^k$$

telle que $V(t)^{p+1} = 0$. La fonction V est dérivable sur $\mathbb{R}$ ainsi que la fonction φ définie par :

$$\varphi(t) = \ln\left(e^{tA}\right) - tA = \sum_{k=1}^{p+1} \frac{(-1)^{k-1}}{k} V(t)^k - tA$$

avec :

$$\varphi'(t) = V'(t)\left(\sum_{k=1}^{p+1} (-1)^{k-1} V(t)^{k-1}\right) - A.$$

Il est facile de vérifier que :

$$(I_n + V(t))\left(\sum_{k=1}^{p+1} (-1)^{k-1} V(t)^{k-1}\right) = I_n,$$

c'est-à-dire que :

$$\sum_{k=1}^{p+1} (-1)^{k-1} V(t)^{k-1} = (I_n + V(t))^{-1} = \left(e^{tA}\right)^{-1} = e^{-tA}$$

et avec :

$$V'(t) = \left(e^{tA} - I_n\right)' = Ae^{tA},$$

on déduit que $\varphi'(t) = 0$ pour tout réel t. On a donc :

$$\forall t \in \mathbb{R}, \quad \varphi(t) = \varphi(0) = \ln(I_n) = 0,$$

ce qui équivaut à $\ln\left(e^{tA}\right) = tA$ pour tout réel t. ■

Théorème 7.10 : *L'exponentielle matricielle réalise une bijection de $\mathcal{N}_n(\mathbb{C})$ sur $\mathcal{L}_n(\mathbb{C})$ d'inverse le logarithme matricielle.*

Démonstration : On sait déjà que l'exponentielle matricielle envoie $\mathcal{N}_n(\mathbb{C})$ dans $\mathcal{L}_n(\mathbb{C})$ et que pour toute matrice $A \in \mathcal{N}_n(\mathbb{C})$ on a $e^{\ln(I_n+A)} = I_n + A$ avec $B = \ln(I_n + A) \in \mathcal{N}_n(\mathbb{C})$, ce qui prouve que l'exponentielle matricielle réalise une surjection de $\mathcal{N}_n(\mathbb{C})$ sur $\mathcal{L}_n(\mathbb{C})$.

Si A_1, A_2 dans $\mathcal{N}_n(\mathbb{C})$ sont telles que $e^{A_1} = e^{A_2}$, alors $\ln\left(e^{A_1}\right) = \ln\left(e^{A_2}\right)$, c'est-à-dire, d'après le lemme précédent avec $t = 1$, que $A_1 = A_2$. L'exponentielle matricielle restreinte à $\mathcal{N}_n(\mathbb{C})$ est donc injective. ■

Corollaire 7.5 : *Pour tout nombre complexe λ non nul et pour toute matrice $A \in \mathcal{N}_n(\mathbb{C})$ il existe une matrice $X \in \mathcal{M}_n(\mathbb{C})$ telle que :*

$$e^X = \lambda I_n + A.$$

Démonstration : Soient $\lambda \in \mathbb{C}$ et $A \in \mathcal{N}_n(\mathbb{C})$. On sait que la fonction exponentielle complexe est surjective de $\mathbb{C}$ sur $\mathbb{C}^*$, il existe donc un nombre complexe μ tel que $\lambda = e^\mu$ et en posant :

$$X = \mu I_n + \ln\left(I_n + \frac{1}{\lambda}A\right),$$

on a :

$$e^X = e^{\mu I_n} e^{\ln\left(I_n + \frac{1}{\lambda}A\right)} = \lambda I_n\left(I_n + \frac{1}{\lambda}A\right) = \lambda I_n + A. \quad ■$$

Théorème 7.11 : *L'exponentielle matricielle réalise une surjection de $\mathcal{M}_n(\mathbb{C})$ sur $GL_n(\mathbb{C})$.*

Démonstration : Le théorème de réduction de Jordan nous dit que toute matrice $A \in GL_n(\mathbb{C})$ est semblable à une matrice bloc de la forme :

$$J = \begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & J_p \end{pmatrix},$$

avec $J_k = \lambda_k I_n + V_k$, pour tout entier k compris entre 1 et p, la matrice V_k étant nilpotente et λ_k étant valeur propre de A. Comme la matrice A est inversible, tous les λ_k sont non nuls et on peut trouver des matrices à coefficients complexes X_k telles que $e^{X_k} = J_k$.

Si $A = PJP^{-1}$ avec P inversible, en définissant la matrice X par :

$$X = P \begin{pmatrix} X_1 & 0 & \cdots & 0 \\ 0 & X_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & X_p \end{pmatrix} P^{-1},$$

on a :

$$e^X = P \begin{pmatrix} e^{X_1} & 0 & \cdots & 0 \\ 0 & e^{X_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & e^{X_p} \end{pmatrix} P^{-1} = PJP^{-1} = A. \qquad \blacksquare$$

Remarque 7.5 : *Il est facile de vérifier que l'exponentielle matricielle définie sur $\mathcal{M}_n(\mathbb{C})$ n'est pas injective. Par exemple, pour tout entier relatif k on a $e^{2ik\pi I_n} = I_n$, c'est-à-dire que l'équation $e^X = I_n$ a une infinité de solutions dans $\mathcal{M}_n(\mathbb{C})$. En fait les solutions sont les matrices semblables aux matrices $2i\pi K$ où K est une matrice diagonale d'ordre n à coefficients entiers relatifs (exercice 7.14).*

Le théorème précédent nous permet de retrouver la connexité de $GL_n(\mathbb{C})$.

Corollaire 7.6 : *$GL_n(\mathbb{C})$ est connexe par arcs.*

Démonstration : Soient A_1, A_2 deux matrices dans $GL_n(\mathbb{C})$. Il existe deux matrices X_1, X_2 dans $\mathcal{M}_n(\mathbb{C})$ telles que $e^{X_1} = A_1$ et $e^{X_2} = A_2$. L'application φ définie sur $[0,1]$ par :

$$\varphi(t) = e^{(1-t)X_1 + tX_2}$$

est alors un chemin continu dans $GL_n(\mathbb{C})$ qui relie A_1 et A_2. $\blacksquare$

On déduit aussi l'existence d'une racine p-ième pour toute matrice inversible.

Corollaire 7.7 : *Soit p un entier naturel non nul. Pour toute matrice $A \in GL_n(\mathbb{C})$ il existe une matrice $X \in GL_n(\mathbb{C})$ telle que $X^p = A$.*

Démonstration : Il existe une matrice $Y \in \mathcal{M}_n(\mathbb{C})$ telle que $e^Y = A$. En posant $X = e^{\frac{1}{p}Y}$, on a $X \in GL_n(\mathbb{C})$ et $X^p = e^Y = A$. ■

Si on se restreint aux matrices complexes hermitiennes [resp. réelles symétriques] l'exponentielle est injective et son image est l'ensemble des matrices complexes hermitiennes définies positives [resp. réelles symétriques définies positives].

Théorème 7.12 : *L'exponentielle matricielle réalise une bijection de l'ensemble des matrices complexes hermitiennes sur l'ensemble des matrices complexes hermitiennes définies positives.*

Démonstration : Il est facile de vérifier que l'exponentielle d'une matrice hermitienne est également hermitienne. De plus si X est hermitienne alors ses valeurs propres $\lambda_1, \ldots, \lambda_n$ sont réelles et celles de e^X sont les réels strictement positifs $e^{\lambda_1}, \ldots, e^{\lambda_n}$, en conséquence la matrice e^X est définie positive.

Soit $A = UDU^*$ une matrice hermitienne définie positive où U est unitaire et D diagonale de valeurs propres réelles strictement positives $\mu_1, \ldots, \mu_n$. On note P le polynôme d'interpolation de Lagrange défini par $P(\mu_i) = \ln(\mu_i)$ pour tout $i \in \{1, \ldots, n\}$ (le degré de P est $p-1$ où p est le nombre de valeurs propres distinctes de A). La matrice $X = P(A) = UP(D)U^*$ est hermitienne et telle que $e^X = A$. Si Y est une autre matrice hermitienne telle que $e^Y = A$, alors Y commute avec A, donc avec X qui est polynomiale en A. En définitive les matrices X et Y commutent et sont diagonalisables, on sait alors qu'elles sont simultanément diagonalisables (exercice 3.3), c'est-à-dire qu'il existe une matrice P dans $GL_n(\mathbb{C})$ telle que $X = P\Delta P^{-1}$ et $Y = P\Lambda P^{-1}$ avec Δ et Λ diagonales à coefficients réels. De $e^X = e^Y$ on déduit alors que $e^\Delta = e^\Lambda$ et $\Delta = \Lambda$ du fait que ces matrices sont diagonales à coefficients réels et que l'exponentielle réelle est bijective de $\mathbb{R}$ sur $\mathbb{R}_+^*$. On a donc $X = Y$, c'est à dire que l'équation $e^X = A$ a une unique solution hermitienne. ■

8. Exercices

Exercice 7.1 : *Comparer e^{A+B}, $e^A e^B$ et $e^B e^A$ pour les matrices :*

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}.$$

Solution : Les deux matrices ont même polynôme caractéristique :

$$P(X) = X(X-1).$$

Avec le théorème de Cayley-Hamilton on déduit qu'elles vérifient les équations $M^2 = M$ et donc $M^k = M$ pour tout entier k supérieur ou égal à 2. Il en résulte que :

$$e^A = I_2 + (e-1)A = \begin{pmatrix} e & e-1 \\ 0 & 1 \end{pmatrix}, \quad e^B = I_2 + (e-1)B = \begin{pmatrix} e & 1-e \\ 0 & 1 \end{pmatrix}.$$

Ce qui donne :

$$e^A e^B = \begin{pmatrix} e^2 & -(e-1)^2 \\ 0 & 1 \end{pmatrix}, \quad e^B e^A = \begin{pmatrix} e^2 & (e-1)^2 \\ 0 & 1 \end{pmatrix} \neq e^A e^B.$$

D'autre part, la matrice $A + B = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}$ a pour polynôme caractéristique $P(X) = X(X-2)$, ce qui entraîne, en utilisant le théorème de Cayley-Hamilton, $(A+B)^2 = 2(A+B)$, et

$$e^{A+B} = I_2 + \frac{1}{2}\left(e^2 - 1\right)(A+B) = \begin{pmatrix} e^2 & 0 \\ 0 & 1 \end{pmatrix}$$

est différent de $e^A e^B$ et de $e^B e^A$.

Exercice 7.2 : *Soient A, B dans $\mathcal{M}_n(\mathbb{C})$. Montrer que $e^{t(A+B)} = e^{tA}e^{tB}$ pour tout réel t si et seulement si A et B commutent.*

Solution : Pour tout réel t on a :

$$e^{t(A+B)} = I_n + t(A+B) + \frac{t^2}{2}(A+B)^2 + o\left(t^2\right)$$
$$e^{tA}e^{tB} = I_n + t(A+B) + \frac{t^2}{2}(A + 2AB + B) + o\left(t^2\right)$$

et, du fait de l'unicité du développement limité d'ordre 2, l'égalité $e^{t(A+B)} = e^{tA}e^{tB}$ est réalisée si et seulement si $(A+B)^2 = A+2AB+B$, ce qui équivaut à $AB = BA$.

Exercice 7.3 : *Pour toutes matrices A, B dans $\mathcal{M}_n(\mathbb{C})$ on note :*

$$[A,B] = AB - BA.$$

1. Soient A, B dans $\mathcal{M}_n(\mathbb{C})$ telles que $\left[A, [A,B]\right] = 0$. Montrer que :

$$\forall t \in \mathbb{R}, \quad e^{tA}Be^{-tA} = B + t[A,B].$$

2. Soient A, B dans $\mathcal{M}_n(\mathbb{C})$ telles que $\left[A, [A,B]\right] = \left[B, [B,A]\right] = 0$. Montrer que :

$$\forall t \in \mathbb{R}, \quad e^{-t(A+B)}e^{tA}e^{tB} = e^{\frac{t^2}{2}[A,B]}.$$

3. Soient A, B dans $\mathcal{M}_n(\mathbb{C})$ telles que $\left[A, [A,B]\right] = \left[B, [B,A]\right] = 0$. Montrer que :

$$e^{A+B} = e^A e^B e^{-\frac{1}{2}[A,B]}.$$

Solution :

1. On note $Y(t) = e^{tA}Be^{-tA}$ pour tout réel t et on a :

$$\forall t \in \mathbb{R}, \quad Y'(t) = e^{tA}\,[A,B]\,e^{-tA}.$$

Si de plus on a $\left[A,[A,B]\right] = 0$, alors les matrices A et $[A,B]$ commutent, ce qui entraîne que les matrices e^{tA} et $[A,B]$ commutent de sorte que :

$$\forall t \in \mathbb{R}, \quad Y'(t) = [A,B]\,e^{tA}e^{-tA} = [A,B]$$

et, avec $Y(0) = B$, on déduit que $Y(t) = B + t\,[A,B]$.

2. On note $Y(t) = e^{-t(A+B)}e^{tA}e^{tB}$ et $Z(t) = e^{\frac{t^2}{2}[A,B]}$ pour tout réel t. On a :

$$\forall t \in \mathbb{R}, \quad Z'(t) = t\,[A,B]\,Z(t),$$

$$\begin{aligned}\forall t \in \mathbb{R}, \quad Y'(t) &= e^{-t(A+B)}\left(-Be^{tA}e^{tB} + e^{tA}Be^{tB}\right)\\ &= e^{-t(A+B)}\left(-B + e^{tA}Be^{-tA}\right)e^{tA}e^{tB}.\end{aligned}$$

Avec l'hypothèse $\left[A,[A,B]\right] = 0$ on déduit que $-B + e^{tA}Be^{-tA} = t\,[A,B]$ et :

$$\forall t \in \mathbb{R}, \quad Y'(t) = te^{-t(A+B)}\,[A,B]\,e^{tA}e^{tB}.$$

Enfin, avec l'hypothèse $\left[A,[A,B]\right] = \left[B,[B,A]\right] = 0$, on déduit que $[A,B]$ et $A+B$ commutent, ce qui entraîne que $[A,B]$ et $e^{-t(A+B)}$ commutent de sorte que :

$$\forall t \in \mathbb{R}, \quad Y'(t) = t\,[A,B]\,Y(t).$$

Avec $Y(0) = Z(0) = I_n$ on déduit que Y et Z sont solutions du même problème de Cauchy, ce qui donne $Y(t) = Z(t)$ pour tout réel t.

3. En faisant $t = 1$ dans l'égalité précédente on a :

$$e^{-(A+B)}e^{A}e^{B} = e^{\frac{1}{2}[A,B]}$$

ce qui équivaut à :

$$e^{A+B} = e^{A}e^{B}e^{-\frac{1}{2}[A,B]}.$$

Exercice 7.4 : *Montrer que pour toute matrice $A \in \mathcal{M}_n(\mathbb{C})$ on a :*

$$\lim_{p \to +\infty}\left(I_n + \frac{1}{p}A\right)^p = e^A.$$

Solution : On pose, pour tout entier naturel p :

$$A_p = \left(I_n + \frac{1}{p}A\right)^p = \sum_{k=0}^{p} \frac{C_p^k}{p^k} A^k.$$

On note $\|\cdot\|$ une norme sous-multiplicative sur $\mathcal{M}_n(\mathbb{C})$.

Pour tout entier naturel p, on a :

$$\left\|e^A - A_p\right\| \leq \left\|\sum_{k=0}^{p} \left(\frac{1}{k!} - \frac{C_p^k}{p^k}\right) A^k\right\| + \sum_{k=p+1}^{+\infty} \frac{1}{k!} \|A\|^k,$$

avec :

$$\epsilon_p = \sum_{k=p+1}^{+\infty} \frac{1}{k!} \|A\|^k \underset{p \to +\infty}{\to} 0.$$

En remarquant que :

$$\frac{1}{k!} - \frac{C_p^k}{p^k} = \frac{1}{k!}\left(1 - \frac{p(p-1)\dots(p-k+1)}{p^k}\right) \geq 0,$$

on déduit que :

$$\left\|e^A - A_p\right\| \leq \sum_{k=0}^{p} \left(\frac{1}{k!} - \frac{C_p^k}{p^k}\right) \|A\|^k + \epsilon_p = e^{\|A\|} - \left(1 + \frac{\|A\|}{p}\right)^p \underset{p \to +\infty}{\to} 0.$$

Exercice 7.5 : *Soit $A \in \mathcal{M}_n(\mathbb{C})$ diagonalisable de valeurs propres deux à deux distinctes $\lambda_1, \dots, \lambda_p$ $(1 \leq p \leq n)$. Montrer que :*

$$\forall t \in \mathbb{R}, \quad e^{tA} = \sum_{k=1}^{p} \alpha_k e^{\lambda_k t} \prod_{\substack{j=1 \\ j \neq k}}^{p} (A - \lambda_j I_n),$$

où les coefficients $\alpha_k \in \mathbb{C}$ s'expriment en fonction des valeurs propres de A.

Solution : En reprenant les notations du théorème 7.3, le vecteur C_k est la projection de $y(0)$ sur $\ker(A - \lambda_k I_n)$. En désignant par p_k cette projection, on déduit alors du théorème 7.3 que $e^{tA} = \sum_{k=1}^{p} e^{\lambda_k t} p_k$. D'autre part, on sait que ces projecteurs s'écrivent $p_k = \alpha_k \prod_{\substack{j=1 \\ j \neq k}}^{p} (A - \lambda_j I_n)$ où $\alpha_k = \dfrac{1}{\prod_{\substack{j=1 \\ j \neq k}}^{p} (\lambda_k - \lambda_j)}$ (exercice 3.2).

Ce qui donne en définitive :

$$e^{tA} = \sum_{k=1}^{p} \alpha_k e^{\lambda_k t} \prod_{\substack{j=1 \\ j \neq k}}^{p} (A - \lambda_j I_n).$$

Exercice 7.6 : *Soit* $A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in M_4(\mathbb{C})$.

1. Calculer e^{tA} *pour tout* $t \in \mathbb{R}$.

2. Résoudre le système différentiel $y' = Ay$.

Solution :

1. Le polynôme caractéristique de A est $P_A(X) = X(X-1)^3$. L'espace propre associé à la valeur propre 0 est $E_0 = \mathbb{C}\begin{pmatrix} 1 \\ 0 \\ 0 \\ -1 \end{pmatrix}$ et l'espace propre associé à la valeur propre 1 est $E_1 = \mathbb{C}\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \oplus \mathbb{C}\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$. La matrice A n'est donc pas diagonalisable et, comme

$$\mathbb{C}^4 = \ker(A) \oplus \ker(A - I_4)^3,$$

on déduit que :

$$2 < \dim\left(\ker(A - I_4)^2\right) \leq \dim\left(\ker(A - I_4)^3\right) = 3.$$

Donc $\ker(A - I_4)^2 = \dim\left(\ker(A - I_4)^3\right)$ est de dimension 3 et

$$\mathbb{C}^4 = \ker A \oplus \ker(A - I_4)^2$$

(le polynôme minimal de A est $\pi_A(X) = X(X-1)^2$).

Pour tout $x \in \ker A$ et pour tout $y \in \ker(A - I_4)^2$, on a :

$$\begin{aligned} e^{tA}x &= x; \\ e^{tA}y &= e^t e^{t(A-I_4)}y = e^t\big(y + t(A - I_4)y\big). \end{aligned}$$

Tout vecteur $z \in \mathbb{C}^4$ s'écrit $z = x + y$ où :

$$\begin{aligned} x &= (A - I_4)^2 z \in \ker A, \\ y &= A(2I_4 - A)z \in \ker(A - I_4)^2 \end{aligned}$$

(l'identité de Bézout $(-X+2)X + (X-1)^2 = 1$ permet d'obtenir les projections sur les sous-espaces caractéristiques) et :

$$\begin{aligned} e^{tA}z &= x + e^t\big(y + t(A - I_4)y\big) \\ &= (A - I_4)^2 z + e^t\big(A(2I_4 - A)z + t(A - I_4)A(2I_4 - A)z\big). \end{aligned}$$

En écrivant que :

$$A(2I_4 - A) = A - A(I_4 - A),$$
$$(A - I_4)A(2I_4 - A) = -A(A - I_4)^2 + A(A - I_4) = A(A - I_4),$$

on déduit que :

$$e^{tA}z = (A - I_4)^2 z + e^t\big(Az + (t-1)A(A - I_4)z\big).$$

Enfin, de :

$$(A - I_4)^2 = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \end{pmatrix}, \quad (A - I_4)A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

on déduit que :

$$e^{tA} = \begin{pmatrix} e^t & 0 & 1 - e^t & e^t - 1 \\ 0 & e^t & te^t & 0 \\ 0 & 0 & e^t & 0 \\ 0 & 0 & e^t - 1 & 1 \end{pmatrix}.$$

2. Les solutions de $y' = Ay$ sont alors données par :

$$y(t) = e^{tA}y_0 = \begin{pmatrix} \alpha e^t + (\gamma - \delta)(1 - e^t) \\ (\beta + \gamma t)e^t \\ \gamma e^t \\ \gamma(e^t - 1) + \delta \end{pmatrix}.$$

Exercice 7.7 : *Soit* $A = \begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \in M_4(\mathbb{C})$*. Calculer* e^{tA} *en utilisant l'algorithme du paragraphe 3, page 272.*

Solution : On a $P_A(X) = X(X-1)^3$, donc $\lambda_1 = 0$, $\lambda_2 = \lambda_3 = \lambda_4 = 1$ et

$$B_1 = I_4, \quad B_2 = A, \quad B_3 = (A - I_4)A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

$$B_4 = (A - I_4)^2 A = 0$$

(le polynôme minimal de A est $X(X-1)^2$).

La solution de (7.8) est donnée par :

$$\begin{cases} y_1(t) = 1, \\ y_2(t) = e^t - 1, \\ y_3(t) = (t-1)\, e^t + 1, \\ y_4(t) = \left(\dfrac{(t-1)^2 + 1}{2} \right) e^t - 1. \end{cases}$$

(le calcul de y_4 n'est pas nécessaire puisque $B_4 = 0$).

On a alors :

$$\begin{aligned} e^{tA} &= I_4 + \left(e^t - 1\right) A + \left((t-1)\, e^t + 1\right) (A - I_4)\, A \\ &= \begin{pmatrix} e^t & 0 & 1 - e^t & e^t - 1 \\ 0 & e^t & te^t & 0 \\ 0 & 0 & e^t & 0 \\ 0 & 0 & e^t - 1 & 1 \end{pmatrix}. \end{aligned}$$

Exercice 7.8 : *En utilisant les résultats de l'exercice 2.8, calculer l'exponentielle de la matrice* $A(\alpha, \beta) = \left((a_{ij})\right)_{1 \le i,j \le n}$ *d'ordre* n *supérieur ou égal à 3 définie par :*

$$\forall i \in \{1, 2, \ldots, n\}, \quad \begin{cases} a_{ii} = \beta, \\ a_{ij} = \alpha & si\ j \in \{1, 2, \ldots, n\} - \{i\}, \end{cases}$$

où α, β *sont des nombres complexes donnés.*

Solution : Pour $\alpha = 0$, on $A(\alpha, \beta) = \beta I_n$ et $e^{A(\alpha,\beta)} = e^{\beta} I_n$.

On suppose que α est non nul et dans ce cas on a (exercice 2.8) :

$$\forall k \in \mathbb{N}, \quad A(\alpha, \beta)^k = \frac{1}{n} \left(\lambda_2^k A(1,1) - \lambda_1^k A(1, 1-n) \right)$$

avec $\lambda_1 = \beta + (n-1)\, \alpha$ et $\lambda_2 = \beta - \alpha$. Ce qui donne :

$$e^{A(\alpha,\beta)} = \frac{1}{n} \left(e^{\lambda_2} A(1,1) - e^{\lambda_1} A(1, 1-n) \right).$$

Exercice 7.9 : *Montrer que le système différentiel* $y' = Ay$ *a toutes ses solutions bornées sur* $\mathbb{R}$ *si et seulement si* A *est diagonalisable avec toutes ses valeurs propres imaginaires pures.*

Démonstration : Si A est diagonalisable à valeurs propres imaginaires pures alors toute solution de $y' = Ay$ est de la forme $y(t) = \sum_{k=1}^{p} e^{i\mu_k t} C_k$ avec les μ_k réels et pour tout $t \in \mathbb{R}$ on a :

$$\|y(t)\| \leq \sum_{k=1}^{p} \left|e^{i\mu_k t}\right| \|C_k\| = \sum_{k=1}^{p} \|C_k\| = M.$$

C'est-à-dire que toute solution de $y' = Ay$ est bornée.

Si A admet une valeur propre λ de partie réelle non nulle, alors pour tout vecteur propre associé $x \in \mathbb{C}^n - \{0\}$ la fonction y définie par $y(t) = e^{\lambda t} x$ est une solution de $y' = Ay$ telle que :

$$\forall t \in \mathbb{R}, \quad \|y(t)\| = e^{\mathcal{R}(\lambda)t} \|x\|$$

et pour $\mathcal{R}(\lambda) > 0$ [resp. $\mathcal{R}(\lambda) < 0$] on a $\lim_{t \longrightarrow +\infty} \|y(t)\| = +\infty$ [resp. $\lim_{t \longrightarrow -\infty} \|y(t)\| = +\infty$]. C'est donc une solution non bornée.

Si A n'est pas diagonalisable, il existe une valeur propre λ de A telle que $\ker(A - I_n) \subsetneq \ker(A - I_n)^2$ et on peut trouver un vecteur non nul $x \in \mathbb{C}^n$ tel que $(A - I_n)x \neq 0$ et $(A - I_n)^2 x = 0$. La solution du problème de Cauchy $y' = Ay$, $y(0) = x$, est alors :

$$y(t) = e^{\lambda t} e^{t(A - I_n)} x = e^{\lambda t}\big(x + t(A - I_n)x\big)$$

et

$$\|y(t)\| \geq e^{\mathcal{R}(\lambda)t}\big(|t| \|(A - I_n)x\| - \|x\|\big).$$

C'est donc une solution non bornée.

On a donc ainsi montré que le système différentiel $y' = Ay$ a toutes ses solutions bornées sur $\mathbb{R}$ si et seulement si A est diagonalisable avec toutes ses valeurs propres imaginaires pures. ■

Exercice 7.10 : *Montrer que si $A \in \mathcal{M}_n(\mathbb{C})$ est telle que $A^* = -A$ alors toutes les solutions de $y' = Ay$ sont bornées sur $\mathbb{R}$.*

Solution : Si $A^* = -A$ alors la matrice iA est hermitienne. Elle est donc diagonalisable à valeurs propres réelles. Les valeurs propres de A sont donc imaginaires pures et avec le résultat de l'exercice 7.9 on déduit que toutes les solutions de $y' = Ay$ sont bornées sur $\mathbb{R}$.

Exercice 7.11 : *Donner une condition nécessaire et suffisante sur la matrice A pour que le système $y' = Ay$ ait toutes ses solutions bornées sur $\mathbb{R}^+$.*

Solution : On a vu avec l'exercice 7.9 que s'il existe une valeur propre λ de A de partie réelle strictement positive, alors on peut construire une solution de $y' = Ay$ non bornée sur $\mathbb{R}^+$.

Supposons que toutes les valeurs propres de A soient de partie réelle négative ou nulle. On note $\lambda_1, \ldots, \lambda_r$ les valeurs propres de partie réelle nulle (quand il en existe) et $\lambda_{r+1}, \ldots, \lambda_p$ celles de partie réelle strictement négative (quand il en existe), les λ_k étant deux à deux distinctes.

Le polynôme minimal de A s'écrit alors $\pi_A(X) = \prod_{k=1}^{p} (X - \lambda_k)^{\alpha_k}$ et toute solution de $y' = Ay$ est de la forme $y(t) = \sum_{k=1}^{p} e^{\lambda_k t} P_k(t)$ avec $P_k(t) = \sum_{j=0}^{\alpha_k - 1} \frac{t^j}{j!} (A - \lambda_k I_n)^j x_k$. Une telle fonction est bornée sur $\mathbb{R}^+$ si et seulement si $\alpha_1 = \ldots = \alpha_r = 1$.

En définitive toutes les solutions de $y' = Ay$ sont bornées sur $\mathbb{R}^+$ si et seulement si toutes les valeurs propres de A sont de partie réelle négative ou nulle, celles de partie réelle nulle étant racines simples du polynôme minimal de A. Ce qui est encore équivalent à dire que la matrice A est semblable à une matrice de l'une des trois formes suivantes :

$$\begin{pmatrix} e^{i\theta_1} & 0 & \ldots & 0 \\ 0 & e^{i\theta_2} & \ldots & 0 \\ . & . & \ddots & . \\ 0 & \ldots & 0 & e^{i\theta_n} \end{pmatrix}, \quad \begin{pmatrix} e^{i\theta_1} & 0 & . & . & . & 0 \\ . & \ddots & . & . & . & . \\ 0 & \ldots & e^{i\theta_r} & 0 & \ldots & 0 \\ 0 & \ldots & 0 & J_{r+1} & \ldots & 0 \\ . & . & . & . & \ddots & . \\ 0 & . & . & . & 0 & J_p \end{pmatrix},$$

ou

$$\begin{pmatrix} J_1 & 0 & \ldots & 0 \\ 0 & J_2 & \ldots & 0 \\ . & . & \ddots & . \\ 0 & \ldots & 0 & J_p \end{pmatrix}$$

$$\text{avec } J_k = \begin{pmatrix} \lambda_k & \epsilon_1 & & \ldots & 0 \\ 0 & \lambda_k & \epsilon_2 & \ldots & 0 \\ . & . & \ddots & \ddots & . \\ 0 & \ldots & 0 & \lambda_k & \epsilon_{n-1} \\ 0 & & \ldots & 0 & \lambda_k \end{pmatrix}, \mathcal{R}(\lambda_k) < 0, \epsilon_i \in \{0, 1\}.$$

Exercice 7.12 : *Pour A et B dans $\mathcal{M}_n(\mathbb{C})$ on désigne par $\Phi_{A,B}$ l'application de $\mathcal{M}_n(\mathbb{C})$ dans $\mathcal{M}_n(\mathbb{C})$ définie par :*

$$\Phi_{A,B}(X) = AX + XB.$$

1. *Donner les solutions du problème de Cauchy :*

$$\begin{cases} Y'(t) = \Phi_{A,B}(Y)(t) \quad (t \in \mathbb{R}), \\ Y(0) = C, \end{cases} \tag{7.14}$$

où C est donnée dans $\mathcal{M}_n(\mathbb{C})$ (on peut utiliser $Z(t) = Y(t)e^{-tB}$).

2. *On suppose que toutes les valeurs propres de A et B sont de partie réelle strictement négative. On se donne une matrice $C \in \mathcal{M}_n(\mathbb{C})$ et Y est la solution de (7.14). L'espace vectoriel $\mathcal{M}_n(\mathbb{C})$ est muni d'une norme matricielle induite par une norme vectorielle sur $\mathbb{C}^n$.*

(a) *Montrer qu'il existe deux constantes $\alpha > 0$ et $\beta > 0$ telles que :*

$$\forall t \in \mathbb{R}, \quad \|Y(t)\| \leq \beta e^{-\alpha t}.$$

(b) *Montrer qu'il existe $X \in \mathcal{M}_n(\mathbb{C})$ telle que $C = \Phi_{A,B}(X)$.*

(c) *En déduire que $\Phi_{A,B}$ est un isomorphisme.*

Solution :

1. Le problème (7.14) étant un problème de Cauchy ($\mathcal{M}_n(\mathbb{C})$ est identifié à $\mathbb{C}^{n^2}$), il admet donc une unique solution $Y \in \mathcal{C}^\infty(\mathbb{R}, \mathcal{M}_n(\mathbb{C}))$. On définit $Z \in \mathcal{C}^\infty(\mathbb{R}, \mathcal{M}_n(\mathbb{C}))$ par :

$$\forall t \in \mathbb{R}, \quad Z(t) = Y(t)e^{-tB}$$

et pour tout $t \in \mathbb{R}$ on a :

$$\begin{aligned} Z'(t) &= Y'(t)e^{-tB} - Y(t)Be^{-tB} \\ &= (AY(t) + Y(t)B)e^{-tB} - Y(t)Be^{-tB} = AZ(t), \end{aligned}$$

avec $Z(0) = C$. On a donc $Z(t) = e^{tA}C$ et la solution du problème (7.14) est $Y(t) = e^{tA}Ce^{tB}$.

2. (a) En reprenant les notations de la démonstration du théorème 7.4, on a

$$\begin{aligned} \left\|e^{tA}\right\| &= \left\| \sum_{k=1}^{p} e^{\lambda_k t} \sum_{j=0}^{\alpha_k - 1} \frac{t^j}{j!} (A - \lambda_k I_n)^j p_k \right\| \\ &\leq e^{-\mu t} \nu \sum_{j=0}^{n-1} \frac{t^j}{j!}, \end{aligned}$$

où $-\mu = \max\limits_{1 \leq k \leq p} \mathcal{R}(\lambda_k) < 0$, $\nu = \max\limits_{\substack{1 \leq k \leq p \\ 0 \leq j \leq \alpha_k - 1}} \|A - \lambda_k I_n\|^j \|p_k\|$. De :

$$\lim_{t \to +\infty} e^{\frac{\mu}{2} t} \left\|e^{tA}\right\| = 0,$$

on déduit qu'il existe $\beta_1 > 0$ tel que $\left\|e^{tA}\right\| \leq e^{-\alpha_1 t}\beta_1$ où $\alpha_1 = \frac{\mu}{2}\cdot$ On a une majoration analogue pour la matrice B avec des constantes α_2 et β_2. On déduit alors la majoration de la solution de (7.14) :

$$\forall t \in \mathbb{R}, \quad \|Y(t)\| \leq \left\|e^{tA}\right\| \left\|e^{tB}\right\| \|C\| \leq e^{-\alpha t}\beta,$$

avec $\alpha = \alpha_1 + \alpha_2$ et $\beta = \beta_1\beta_2 \|C\|$.

(b) Soit Y solution de (7.14). En intégrant on a :

$$Y(t) = A\int_0^t Y(u)\ du + \int_0^t Y(u)\ du\, B + C.$$

De $\|Y(t)\| \leq e^{-\alpha t}\beta$ on déduit que l'intégrale $\int_0^{+\infty} Y(u)\ du$ est convergente et que $\lim_{t\to+\infty} Y(t) = 0$. En faisant tendre t vers l'infini dans l'identité ci-dessus, on déduit alors que :

$$-C = A\int_0^{+\infty} Y(u)\ du + \int_0^{+\infty} Y(u)\ du\, B,$$

c'est-à-dire $C = \Phi_{A,B}(X)$ avec $X = -\int_0^{+\infty} Y(u)\ du \in \mathcal{M}_n(\mathbb{C})$.

(c) Le résultat de la question précédente se traduit en disant que $\Phi_{A,B}$ est surjective. Cette application étant un endomorphisme de $\mathcal{M}_n(\mathbb{C})$ on déduit que c'est un isomorphisme.

***Remarque 7.6** : Une autre démonstration consiste à utiliser le fait que* $\mathrm{sp}(\Phi_{A,B}) = \mathrm{sp}(A) + \mathrm{sp}(B)$.

Exercice 7.13 : *Montrer que les fonctions :*

$$u(t) = \int_0^{+\infty} e^{-x^2}\cos\left(tx^2\right)\ dx, \quad v(t) = \int_0^{+\infty} e^{-x^2}\sin\left(tx^2\right)\ dx,$$

sont bien définies, de classe C^1 sur $\mathbb{R}$ et solutions d'un système différentiel que l'on résoudra.

Solution : De la convergence de l'intégrale $\int_0^{+\infty} e^{-x^2}\,\mathrm{d}x$, on déduit que la fonction :

$$w : t \mapsto \int_0^{+\infty} e^{(it-1)x^2}\,\mathrm{d}x$$

est bien définie sur tout $\mathbb{R}$. Il en résulte que $u = \Re(w)$ et $v = \Im(w)$ sont bien définies sur $\mathbb{R}$.

En dérivant formellement, on a $w'(t) = \int_0^{+\infty} ix^2 e^{(it-1)x^2}\,\mathrm{d}x$. De la convergence de l'intégrale $\int_0^{+\infty} x^2 e^{-x^2}\,\mathrm{d}x$ on déduit que la fonction

$$y : t \mapsto \int_0^{+\infty} ix^2 e^{(it-1)x^2}\,\mathrm{d}x$$

est bien définie sur tout $\mathbb{R}$.

Pour $t \in \mathbb{R}$ et $h \in \mathbb{R}^*$, on a :

$$\frac{w\,(t+h) - w(t)}{h} - y(t) = \int_0^{+\infty} e^{(it-1)x^2} \left(\frac{e^{ihx^2} - 1}{h} - ix^2 \right) \mathrm{d}x,$$

avec :

$$\left| \frac{e^{ihx^2} - 1}{h} - ix^2 \right| = \left| \frac{1}{h} \sum_{n \geq 2} \frac{\left(ihx^2\right)^n}{n!} \right| \leq |h|\, x^4 \sum_{n \geq 2} \frac{\left(|h|\, x^2\right)^{n-2}}{n!} \leq |h|\, x^4 e^{|h| x^2}$$

et pour $|h| \leq \frac{1}{2}$:

$$\left| \frac{e^{ihx^2} - 1}{h} - ix^2 \right| \leq |h|\, x^4 e^{\frac{x^2}{2}}.$$

Ce qui donne en définitive :

$$\left| \frac{w\,(t+h) - w(t)}{h} - y(t) \right| \leq |h| \int_0^{+\infty} x^4 e^{-\frac{x^2}{2}}\,\mathrm{d}x \underset{h \to 0}{\longrightarrow} 0.$$

La fonction w est donc dérivable sur $\mathbb{R}$ de dérivée y. On en déduit alors que les fonctions u et v sont dérivables sur $\mathbb{R}$ de dérivées :

$$u'(t) = -\int_0^{+\infty} x^2 e^{-x^2} \sin\left(tx^2\right)\,\mathrm{d}x, \quad v'(t) = \int_0^{+\infty} x^2 e^{-x^2} \cos\left(tx^2\right)\,\mathrm{d}x.$$

Une intégration par parties (où on dérive $e^{(it-1)x^2}$ par rapport à la variable x) donne :

$$w(t) = -2\,(t+i)\,w'(t), \tag{7.15}$$

ce qui équivaut au système :

$$\begin{cases} u(t) = -2\big(tu'(t) - v'(t)\big), \\ v(t) = -2\big(u'(t) + tv'(t)\big), \end{cases}$$

encore équivalent au système différentiel :

$$\begin{cases} 2\left(1+t^2\right) u'(t) = -tu(t) - v(t), \\ 2\left(1+t^2\right) v'(t) = u(t) - tv(t). \end{cases}$$

On a de plus les conditions initiales :

$$u(0) = \int_0^{+\infty} e^{-x^2}\, dx = \frac{\sqrt{\pi}}{2}, \quad v(0) = 0.$$

L'équation (7.15) entraîne $\left(w^2(t+i)\right)' = 0$, soit $w^2(t+i) = C$, où C est une constante complexe. Avec $w(0) = \dfrac{\sqrt{\pi}}{2}$, on déduit que $C = i\dfrac{\pi}{4}$ et :

$$\begin{aligned} u^2 - v^2 &= \frac{\pi}{4}\frac{1}{1+t^2}, \\ 2uv &= \frac{\pi}{4}\frac{t}{(1+t^2)}. \end{aligned} \tag{7.16}$$

C'est à dire que $\left(u^2, -v^2\right)$ est solution de l'équation polynomiale de degré 2 :

$$r^2 - \frac{\pi}{4}\frac{1}{1+t^2} r - \left(\frac{\pi}{8}\right)^2 \frac{t^2}{\left(1+t^2\right)^2} = 0.$$

Ce qui donne :

$$\begin{aligned} u^2(t) &= \frac{\pi}{8}\frac{1+\sqrt{1+t^2}}{(1+t^2)}, \\ v^2(t) &= \frac{\pi}{8}\frac{\sqrt{1+t^2}-1}{(1+t^2)}. \end{aligned}$$

La fonction continue u ne s'annule jamais, elle garde donc un signe constant et ce signe est celui de $u(0)$, c'est-à-dire le signe positif. De (7.16) on déduit que la fonction uv est impaire du signe de t. Ce qui donne en définitive :

$$\begin{aligned} u(t) &= \frac{1}{2}\sqrt{\frac{\pi}{2}}\sqrt{\frac{1+\sqrt{1+t^2}}{(1+t^2)}}, \\ v(t) &= -\frac{\text{signe}(t)}{2}\sqrt{\frac{\pi}{2}}\sqrt{\frac{\sqrt{1+t^2}-1}{(1+t^2)}}. \end{aligned}$$

Exercice 7.14 : *Déterminer toutes les solutions dans $\mathcal{M}_n(\mathbb{C})$ de l'équation $e^X = I_n$.*

Solution : Soit $X \in \mathcal{M}_n(\mathbb{C})$ telle que $e^X = I_n$. On a la décomposition de Dunford-Schwarz $X = D + V$ avec D diagonalisable, V nilpotente et $DV = VD$. Cette décomposition permet d'obtenir celle de e^X :

$$e^X = e^D + e^D \left(e^V - I_n\right),$$

avec e^D diagonalisable et $e^D \left(e^V - I_n\right)$ nilpotente (lemme 7.6). Avec l'unicité de cette décomposition, on déduit que $e^D = I_n$ et $e^V = I_n$. La première égalité avec D diagonalisable entraîne que D est semblable à une matrice $2i\pi K$ où K est une matrice diagonale à coefficients entiers et la deuxième égalité entraîne $V = 0$. En effet si p est le plus petit entier naturel non nul tel que $V^p = 0$, en supposant $p \geq 2$, l'égalité :

$$e^V = I_n + \sum_{k=1}^{p-1} \frac{1}{k!} V^k = I_n$$

entraîne $\sum_{k=1}^{p-1} \frac{1}{k!} V^k = 0$ et en multipliant par V^{p-2}, on aboutit à $V^{p-1} = 0$ ce qui est en contradiction avec la définition de p. On a donc $p = 1$ et $V = 0$.

On a donc ainsi montré que les solutions dans $\mathcal{M}_n(\mathbb{C})$ de l'équation $e^X = I_n$ sont les matrices semblables à une matrice $2i\pi K$ où K est une matrice diagonale à coefficients entiers.

Bibliographie

[1] R. Burlisch, J. Stoer — *Introduction to numerical analysis.* Springer-Verlag (1980).

[2] A. Chambert-Loir, S. Fermigier, V. Maillot — *Exercices de mathématiques pour l'agrégation. Analyse 1.* Masson (1995).

[3] P.G. Ciarlet — *Introduction à l'analyse numérique matricielle et à l'optimisation.* Masson (1982).

[4] J. Dieudonné — *Eléments d'analyse. Fondements de l'analyse moderne.* Gauthier-Villars (1972).

[5] B. Flannery, W. Press, S. Teukolsky, W. Vetterling — *Numerical recipes.* Cambridge University Press (1988).

[6] S. Gonnord, N. Tosel — *Topologie et analyse fonctionnelle.* Ellipses (1996).

[7] A. Gramain — *Géométrie élémentaire.* Hermann (1997).

[8] R.A. Horn, C.A. Johnson — *Matrix analysis.* Cambridge University Press (1985).

[9] A.S. Householder — *The theory of matrices in numerical analysis.* Dover (1975).

[10] P. Lascaux, R. Theodor — *Analyse numérique matricielle appliquée à l'art de l'ingénieur.* Masson (1987).

[11] R. Mnéimné, F. Testard — *Introduction à la théorie des groupes de Lie classiques.* Hermann (1986).

[12] J. Moussa — *Une preuve élémentaire des formules de Newton.* Revue de mathématiques spéciales (novembre-décembre 1995).

[13] M. Parodi — *La localisation des valeurs caractéristiques des matrices et ses applications.* Gauthier-Villars (1959).

[14] F. Rellich — *Perturbation theory of eigenvalues problems.* Gordon and Breach (1950).

[15] F. RIDEAU — *Exercices de calcul différentiel.* Hermann (1979).

[16] J.E. ROMBALDI — *Algorithmique numérique et Ada.* Masson (1994).

[17] J.E. ROMBALDI — *Problèmes corrigés d'analyse numérique.* Masson (1996).

[18] P. TAUVEL - *Exercices de mathématiques pour l'agrégation. Algèbre 2.* Masson (1994).

[19] P. TAUVEL — *Algèbre pour l'agrégation interne.* Masson (1996).

[20] C. TISSERON — *Géométries affine, projective et euclidienne.* Hermann (1983).

Index

Imprimerie STEDI
1, bd Ney, 75018 Paris
Dépôt légal : 6334

Analyse matricielle - Cours et exercices résolus

Cet ouvrage est consacré à l'étude de l'espace vectoriel $\mathcal{M}_n(\mathbb{K})$ des matrices à coefficients réels ou complexes du point de vue algébrique et topologique, préalable nécessaire à tout cours d'analyse numérique. La synthèse réalisée par l'auteur permet aux étudiants d'approfondir leurs connaissances sur les espaces vectoriels normés et l'algèbre linéaire, des notions de base en algèbre linéaire étant suffisantes pour la lecture de l'ouvrage.

Le public visé est celui des candidats à l'agrégation (interne et externe), mais également les étudiants de licence et maîtrise de mathématiques. Chaque chapitre est suivi d'une série d'exercices corrigés. Les résultats classiques sont illustrés par des exemples qui peuvent trouver leur place dans les leçons d'oral des concours.

Jean-Étienne Rombaldi est professeur agrégé de mathématiques à l'université d'Aix-Marseille III.

9 782868 834256
ISBN : 2-86883-425-6
www.edpsciences.org
EDP SCIENCES