Algèbre linéaire 3

Polycopié de cours

L2 Intelligence Artificielle et Sciences des Organisations

Dauphine UNIVERSITÉ PARIS | PSL

Ce polycopié est une version modifiée du cours d'algèbre linéaire de Guillaume Legendre donné en L2 de mathématiques à l'université Paris Dauphine - PSL. J'y ai principalement rajouté des exemples d'applications de l'algèbre linéaire à des situations concrètes, tirés des livres *Introduction to Linear Algebra, 4th Edition* de Gilbert Strang et *Linear Algebra and its Applications, 6th Edition* de David C. Lay, Steven R. Lay et Judi J. McDonald.

Vincent Divol

Table des matières

1 Réduction des endomorphismes **3**
1.1 Sous-espaces stables par un endomorphisme . . . 3
1.2 Éléments propres d'un endomorphisme . . . 4
1.3 Polynôme caractéristique . . . 6
1.4 Diagonalisation . . . 8
1.5 Trigonalisation . . . 10
1.6 Polynômes annulateurs . . . 13
1.7 Réduction de Jordan (hors programme) . . . 18
1.7.1 Sous-espaces caractéristiques . . . 18
1.7.2 Décomposition de Jordan–Chevalley . . . 19
1.8 Applications (hors programme) . . . 20
1.8.1 Modèles matriciels d'évolution de population . . . 20
1.8.2 Chaînes de Markov . . . 21

2 Formes bilinéaires et quadratiques **25**
2.1 Généralités sur les applications bilinéaires . . . 25
2.2 Représentation matricielle d'une forme bilinéaire . . . 26
Changement de bases . . . 26
2.3 Non-dégénérescence d'une forme bilinéaire symétrique . . . 27
2.4 Généralités sur les formes quadratiques . . . 27
2.5 Orthogonalité . . . 29
2.6 Formes quadratiques positives . . . 30

3 Espaces euclidiens **33**
3.1 Définitions . . . 33
3.2 Orthogonalité dans les espaces euclidiens . . . 35
3.2.1 Généralités . . . 35
3.2.2 Bases orthonormées . . . 35
3.2.3 Projection orthogonale . . . 39
3.3 Endomorphismes des espaces euclidiens . . . 41
3.3.1 Adjoint d'un endomorphisme . . . 41
3.3.2 Isométries vectorielles . . . 41
3.3.3 Endomorphismes auto-adjoints . . . 44

Dans tout ce document, on désigne par $\mathbb{K}$ un corps qui peut être $\mathbb{R}$, le corps de nombres réels, ou $\mathbb{C}$, le corps des nombres complexes. Si E et F sont deux $\mathbb{K}$-espaces vectoriels, on notera $\mathscr{L}(E, F)$ l'ensemble des morphismes de E vers F. On note $\mathscr{L}(E) = \mathscr{L}(E, E)$ l'ensemble des endomorphismes de E.

Chapitre 1

Réduction des endomorphismes

Réduire un endomorphisme dans un espace vectoriel de dimension finie, c'est notamment trouver une base de l'espace dans laquelle cet endomorphisme est « aisément » étudiable et manipulable. En pratique, ceci revient à déterminer une base par rapport laquelle la matrice représentative de l'endomorphisme possède une forme particulière : diagonale (dans le meilleur des cas), diagonale par blocs ou triangulaire. La réduction des endomorphismes (et des matrices qui leur sont associées) est pour cette raison un outil incontournable de l'étude de ces derniers et ce chapitre lui est consacré.

1.1 Sous-espaces stables par un endomorphisme

La réduction d'un endomorphisme repose sur une propriété de *stabilité* de certains sous-espaces vectoriels sous l'action de cet endomorphisme, notion que nous allons maintenant introduire.

Définition 1.1 (sous-espace vectoriel stable par un endomorphisme) *Soit E un $\mathbb{K}$-espace vectoriel, $u \in \mathscr{L}(E)$ et A un sous-espace vectoriel de E. On dit que A est **stable** (ou **invariant**) par u si et seulement si l'image de A par u est incluse dans A, c'est-à-dire*

$$\forall x \in E,\ x \in A \implies u(x) \in A.$$

Le résultat suivant est immédiat.

Théorème 1.2 *Soit E un $\mathbb{K}$-espace vectoriel et $u \in \mathscr{L}(E)$. Si A est un sous-espace vectoriel stable par u, alors la restriction de u à A, notée $u_{|_A}$, définie un endomorphisme de A.*

Exemple 1.3 *Soit E un $\mathbb{K}$-espace vectoriel et $u \in \mathscr{L}(E)$. Montrer que les sous-espaces suivants sont stables par u :*

1. $\{0\}$;
2. E ;
3. $\ker u$;
4. $\operatorname{Im} u$.

Une situation importante impliquant des sous-espaces vectoriels stables est celle dans laquelle on considère deux endomorphismes commutant entre eux.

Théorème 1.4 *Soit E un $\mathbb{K}$-espace vectoriel et u et v deux endomorphismes de E qui commutent entre eux, i.e. $u \circ v = v \circ u$. L'endomorphisme v laisse stable l'image de u, le noyau de u, et plus généralement tout sous-espace $\ker(u - \lambda\, id_E)$, avec λ un scalaire.*

Démonstration. Puisque les endomorphismes u et v commutent entre eux, on a, pour tout vecteur x de E,

$$v(u(x)) = u(v(x)),$$

le dernier vecteur appartenant à l'image de u. On en déduit donc que $v(\operatorname{Im}(u))$ est inclus dans $\operatorname{Im}(u)$, qui est alors stable par v.

Considérons à présent un vecteur x du noyau de u. On a $u(v(x)) = v(u(x)) = v(0_E) = 0_E$, d'où $v(\ker(u))$ est inclus dans $\ker(u)$, qui est alors stable par v.

Soit enfin un scalaire λ. On a

$$v \circ (u - \lambda\, id_E) = v \circ u - \lambda\, v = u \circ v - \lambda\, v = (u - \lambda\, id_E) \circ v,$$

et les endomorphismes $u - \lambda\, id_E$ et v commutent donc aussi entre eux, ce qui permet de conclure. □

En dimension finie, il est possible d'interpréter matriciellement la propriété de stabilité d'une sous-espace vectoriel par un endomorphisme. Pour le voir, considérons un $\mathbb{K}$-espace vectoriel E de dimension finie égale à n, $u \in \mathscr{L}(E)$, A un sous-espace vectoriel de E stable par u, de dimension égale à p. Soit B un sous-espace vectoriel de E supplémentaire de A, c'est-à-dire tel que $E = A \oplus B$, et $\mathscr{B}$ une base adaptée [1] à une telle décomposition. Dans ce cas, la matrice représentative de l'endomorphisme u dans la base $\mathscr{B}$ est triangulaire supérieure par blocs,

$$\begin{pmatrix} M_{11} & M_{12} \\ 0_{n-p,p} & M_{22} \end{pmatrix},$$

où M_{11} est une matrice de $M_p(\mathbb{K})$, M_{12} est une matrice de $M_{p,n-p}(\mathbb{K})$ et M_{22} est une matrice de $M_{n-p}(\mathbb{K})$. Si l'on suppose de plus que le sous-espace B soit stable par u, alors la matrice de l'endomorphisme dans la base $\mathscr{B}$ est diagonale par blocs,

$$\begin{pmatrix} M_{11} & 0_{p,n-p} \\ 0_{n-p,p} & M_{22} \end{pmatrix}.$$

Plus généralement, si on a décomposé E en une somme directe de sous-espaces stables par u, $E = \oplus_{i=1}^{m} A_i$ avec $u(A_i) \subset A_i$ pour tout entier i appartenant à $\{1, \dots, m\}$, et si $\mathscr{B}$ est une base de E adaptée à cette décomposition, alors la matrice de l'endomorphisme dans la base $\mathscr{B}$ sera de la forme

$$\begin{pmatrix} M_{11} & 0 & \dots & 0 \\ 0 & M_{22} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & M_{mm} \end{pmatrix}.$$

Exemple 1.5 (un exemple en dimension infini) *Soit E l'espace vectoriel des polynômes à coefficients réels. On définit l'endomorphisme u, qui à un polynôme P associe le polynôme dérivée P'. Montrer que le sous-espace A des polynômes de degré au plus 17 est stable par u.*

1.2 Éléments propres d'un endomorphisme

Intéressons nous maintenant aux sous-espaces stables les plus simples possibles : ceux de dimension 1. Soit $x \in E$ un vecteur non nul, et considérons l'espace vectoriel $A = \mathrm{Vect}(x)$ engendré par x. Si A est stable par u, alors $u(x) \in A$. Ainsi, il existe un nombre $\lambda \in \mathbb{K}$ avec $u(x) = \lambda x$. Cette équation

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

est intimement lié aux sous-espaces stables de dimension 1. On donne ainsi un nom aux vecteurs x et aux nombres λ la vérifiant.

Définitions 1.6 (éléments propres d'un endomorphisme) *Soit E un $\mathbb{K}$-espace vectoriel et $u \in \mathscr{L}(E)$. Un scalaire λ est une **valeur propre** de u si et seulement s'il existe un vecteur **non nul** x de E tel que $u(x) = \lambda\, x$. Un vecteur non nul x de E tel qu'il existe un scalaire λ vérifiant $u(x) = \lambda\, x$ est un **vecteur propre** de u associé à la valeur propre λ.*

Il est important de remarquer qu'un vecteur propre est associé à une *unique* valeur propre, mais qu'une valeur propre possède une infinité de vecteurs propres, tout multiple non nul d'un vecteur propre donné étant lui-même un vecteur propre associé à la même valeur propre.

1. Cette adaptation s'entend dans le sens où les p premiers vecteurs composant la base forment une base du sous-espace A et les $n-p$ suivants forment une base du sous-espace B.

On notera également que la définition précédente est donnée pour un espace vectoriel de dimension quelconque. En considérant des espaces de dimension finie non nulle, il est possible de l'étendre à une matrice carrée en posant que les éléments propres de la matrice sont ceux de l'endomorphisme qui lui est canoniquement associé. Ainsi, en considérant une matrice M d'ordre n, on a que le scalaire λ est une valeur propre de M s'il existe une matrice colonne X non nulle de $M_{n,1}(\mathbb{K})$ telle que

$$MX = \lambda X.$$

De même, une matrice colonne X de $M_{n,1}(\mathbb{K})$ sera un vecteur propre de la matrice M si elle est non nulle et qu'il existe un scalaire λ tel que $MX = \lambda X$.

Tous les résultats énoncés pour des endomorphismes dans un espace vectoriel de dimension finie seront par conséquent valables *mutatis mutandis* pour des matrices carrées.

Définition 1.7 (spectre d'un endomorphisme) *L'ensemble des valeurs propres d'un endomorphisme u d'un $\mathbb{K}$-espace vectoriel s'appelle le* ***spectre*** *de u et se note* $\mathrm{Sp}(u)$*.*

Si cette dernière définition s'étend naturellement à toute matrice carrée M, la notation $\mathrm{Sp}(M)$ peut s'avérer ambigüe lorsque les coefficients de M sont des réels. En effet, suivant que l'on conçoit M comme une matrice à coefficients complexes ou réels, on cherchera ses valeurs dans $\mathbb{C}$ ou dans $\mathbb{R}$. Dans ce cas, on a recours aux notations $\mathrm{Sp}_{\mathbb{C}}(M)$ ou $\mathrm{Sp}_{\mathbb{R}}(M)$, qui sont plus explicites.

Remarque 1.8 *Un endomorphisme peut ne pas avoir de valeur propre, comme le montrent les exemples suivants.*

1. *Considérons l'endomorphisme $u : \mathbb{R}^2 \to \mathbb{R}^2$ donné par une rotation d'angle $\frac{\pi}{2}$. Soit $\lambda \in \mathbb{R}$ et $x \in \mathbb{R}^2$ un vecteur non nul. Alors, on ne peut pas avoir $u(x) = \lambda x$, puisque $u(x)$ est perpendiculaire à x. Ainsi, u n'a pas de valeurs propres :* $\mathrm{Sp}(u) = \emptyset$*. On peut aussi faire ce raisonnement d'un point de vue matricielle. La représentation matricielle de u dans la base canonique de $\mathbb{R}^2$ est*
$$M = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$
Pour un vecteur $X = (x_1, x_2) \in \mathbb{R}^2$ non nul et $\lambda \in \mathbb{R}$, cherchons à résoudre l'équation $MX = \lambda X$. Nous avons
$$MX = \begin{pmatrix} -x_2 \\ x_1 \end{pmatrix} = \begin{pmatrix} \lambda x_1 \\ \lambda x_2 \end{pmatrix}.$$
Ainsi, $-x_2 = \lambda x_1 = \lambda^2 x_2$ et $x_1 = \lambda x_2 = -\lambda^2 x_1$. Puisque $X \neq 0$, on a soit $x_1 \neq 0$, soit $x_2 \neq 0$, et on obtient $\lambda^2 = 1$. Cette équation n'a pas de solution dans $\mathbb{R}$: ainsi $\mathrm{Sp}_{\mathbb{R}}(M) = \emptyset$*. Quel est le spectre complexe* $\mathrm{Sp}_{\mathbb{C}}(M)$ *de M ?*
2. *Soit E l'espace vectoriel des suites réelles. On définit l'endomorphisme $u : E \to E$ dit « de décalage », qui à la suite $x = (x_1, x_2, x_3, \dots)$ associe la suite $u(x) = (0, x_1, x_2, x_3, \dots)$. Soit $\lambda \in \mathbb{R}$ et $x \in E$. Peut-on avoir $u(x) = \lambda x$? Si $\lambda = 0$, on a clairement $x = 0$. Sinon, en regardant la première entrée des suites $u(x)$ et λx, on voit que $0 = \lambda x_1$, donc $x_1 = 0$. Ensuite, en regardant la deuxième entrée des deux suites, on obtient l'équation $x_1 = \lambda x_2$, donc $x_2 = 0$. En raisonnant par récurrence, on observe que la suite x est nulle. Ainsi, l'équation $u(x) = \lambda x$ n'a pas de solution non nulle, et l'endomorphisme u n'a pas de valeur propre :* $\mathrm{Sp}(u) = \emptyset$*.*

Proposition et définition 1.9 (sous-espace propre) *Soit E un $\mathbb{K}$-espace vectoriel, $u \in \mathscr{L}(E)$ et λ un scalaire. Les assertions suivantes sont équivalentes.*

(i) Le scalaire λ est une valeur propre de u.

(ii) L'endomorphisme $u - \lambda\, id_E$ n'est pas injectif.

(iii) Le noyau $\ker(u - \lambda\, id_E)$ *n'est pas réduit à* $\{0_E\}$*.*

Dans ce cas, le sous-espace vectoriel $E_\lambda = \ker(u - \lambda\, id_E)$ est appelé ***sous-espace propre*** *de u associé à la valeur propre λ.*

DÉMONSTRATION. Si λ est une valeur propre de l'endomorphisme u, alors, par définition, il existe un vecteur x non nul tel que $u(x) = \lambda x$. Il existe par conséquent un vecteur x non nul tel que $(u - \lambda\, id_E)(x) = 0_E$, ce qui signifie encore que l'endomorphisme $u-\lambda\, id_E$ n'est pas injectif, ce qui signifie encore que le noyau de cet endomorphisme n'est pas réduit au vecteur nul. □

Lorsque E est de dimension finie, on a l'équivalence

$$\lambda \in \mathrm{Sp}(u) \iff u-\lambda\, id_E \notin GL(E),$$

où $GL(E)$ est le *groupe général linéaire de* E, c'est-à-dire l'ensemble des automorphismes de E (l'ensemble des endomorphismes inversibles de E).

Proposition 1.10 *Soit E un $\mathbb{K}$-espace vectoriel, $u \in \mathscr{L}(E)$. Soit k un entier naturel non nul et $\lambda_1, \dots, \lambda_k$ des valeurs propres de u deux à deux distinctes. Alors, les sous-epaces propres E_{λ_i}, $i = 1, \dots, k$, sont en somme directe.*

DÉMONSTRATION. On raisonne par récurrence sur l'entier k. Pour $k = 1$, il n'y a rien à montrer. Pour k un entier naturel non nul, on suppose que si $\lambda_1, \dots, \lambda_k$ sont des valeurs propres de u deux à deux distinctes et $x_1, \dots, x_k$ sont des éléments des sous-espaces propres qui leur sont respectivement associés, alors l'égalité $x_1+\dots+x_k = 0_E$ implique que $x_1 = \dots = x_k = 0_E$. Soit alors λ_{k+1} une valeur propre de u distincte des précédentes et x_{k+1} un vecteur propre associé. Dans ce cas, supposons que $x_1+\dots+x_{k+1} = 0_E$, ce qui équivaut à $x_{k+1} = -(x_1+\dots+x_k)$, d'où $u(x_{k+1}) = -u(x_1)-\dots-u(x_k)$, par linéarité de u. On a par conséquent $\lambda_{k+1}\, x_{k+1} = -\lambda_1\, x_1-\dots-\lambda_k\, x_k$, soit encore $0_E = (\lambda_{k+1}-\lambda_1)\, x_1 + \dots + (\lambda_{k+1}-\lambda_k)\, x_k$. L'hypothèse de récurrence conduit alors à $(\lambda_{k+1}-\lambda_i)\, x_i = 0_E$ pour tout entier i dans $\{1, \dots, k\}$ et, puisque $\lambda_{k+1}-\lambda_i \neq 0$ pour tout entier i dans $\{1, \dots, k\}$, on en déduit que $x_i = 0_E$ pour tout entier i dans $\{1, \dots, k\}$, d'où $x_{k+1} = 0_E$. □

Corollaire 1.11 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle égale à n et $u \in \mathscr{L}(E)$. Alors, l'endomorphisme u possède au plus n valeurs propres distinctes.*

Exemple 1.12 *Soit $M = \begin{pmatrix} 4 & -1 & 6 \\ 2 & 1 & 6 \\ 2 & -1 & 8 \end{pmatrix}$. Montrer que 2 est une valeur propre de M, et trouver une base de l'espace propre associé.*

1.3 Polynôme caractéristique

On a vu que le scalaire λ était une valeur propre d'un endomorphisme u d'un espace vectoriel E si et seulement si l'endomophisme $u-\lambda\, id_E$ n'est pas injectif. Si l'espace E est de dimension finie, ceci équivaut encore à dire que le déterminant de $u-\lambda\, id_E$ est nul. Trouver les valeurs propres d'un endomorphisme en dimension finie revient donc à résoudre une équation polynomiale.

Définition 1.13 (polynôme caractéristique) *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. On appelle **polynôme caractéristique** de u le polynôme χ_u de $\mathbb{K}[X]$ défini par $\chi_u(X) = \det(X\, id_E - u)$.*

Remarque 1.14 *On trouve parfois la définition $\chi_u(X) = \det(u - X\, id_E)$, qui a comme défaut le fait que le polynôme ainsi défini n'est pas nécessairement unitaire, le coefficient du terme de plus haut degré étant égal à -1 élevé à une puissance égale à la dimension de l'espace.*

Pour une matrice M d'ordre n, on a $\chi_M(X) = \det(X\, I_n - M)$. Compte tenu de cette définition et des propriétés du déterminant, il est clair que **deux matrices semblables possèdent le même polynôme caractéristique**. La réciproque est fausse, comme le montre l'exemple des matrices I_2 et $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$.

Théorème 1.15 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. Un scalaire λ est une valeur propre de u si et seulement si $\chi_u(\lambda) = 0$.*

Exemple 1.16 *Trouver les valeurs propres de la matrice $M = \begin{pmatrix} 2 & 3 \\ 3 & -6 \end{pmatrix}$ en calculant son polynôme caractéristique.*

Remarque 1.17 *Les valeurs propres d'une matrice triangulaire sont ses coefficients diagonaux.*

Définition 1.18 (ordre de multiplicité algébrique d'une valeur propre) *On appelle **ordre de multiplicité algébrique** d'une valeur propre son ordre de multiplicité en tant que racine du polynôme caractéristique.*

Il découle des dernières définitions que deux matrices ayant le même polynôme caractéristique ont les mêmes valeurs propres, avec les mêmes ordres de multiplicité algébrique.

Proposition 1.19 (degré et coefficients du polynôme caractéristique d'une matrice) *Soit M une matrice d'ordre n à coefficients dans le corps $\mathbb{K}$. Le polynôme caractéristique de M est de degré égal à n et l'on a*

$$\chi_M(X) = X^n - \operatorname{tr}(M)X^{n-1} + \cdots + (-1)^n \det(M).$$

DÉMONSTRATION. Posons $\alpha = X\,I_n - M$. Par la formule de Leibniz pour le déterminant, on a

$$\chi_M(X) = \det(\alpha) = \sum_{\sigma\in\Sigma_n} \varepsilon(\sigma)\prod_{i=1}^{n} \alpha_{\sigma(i)\,i}.$$

Chacun des $n!$ termes de cette somme est le produit de n termes polynomiaux en X de degré inférieur ou égal à un. Le degré de χ_M est donc inférieur ou égal à n. De plus, un des termes de la somme est exactement de degré n si et seulement si chacun des facteurs le composant est de degré exactement égal à un. Ceci équivaut au fait que $\sigma = id_{\{1,\dots,n\}}$. Ainsi, la somme correspondant au polynôme caractéristique de M contient un unique terme de degré n et des termes de degré inférieur ou égal à $n-1$. C'est ainsi un polynôme de degré n, s'écrivant sous la forme

$$\begin{aligned}\chi_M(X) &= (X - m_{11})\dots(X - m_{nn}) + \text{termes de degré inférieur ou égal à } n-1\\ &= X^n + \text{termes de degré inférieur ou égal à } n-1.\end{aligned}$$

C'est donc bien un polynôme unitaire. Le coefficient du terme de degré nul de χ_M est par ailleurs donné par sa valeur en 0, qui est $\det(-M) = (-1)^n \det(M)$.

Enfin, pour déterminer le coefficient du terme de degré $n-1$, on observe que, quand la permutation σ n'est pas égale à l'identité, il existe un entier i de $\{1,\dots,n\}$ tel que $\sigma(i) \neq i$. En posant alors $j = \sigma^{-1}(i)$, on a que $\alpha_{\sigma(i)\,i} = -m_{\sigma(i)\,i}$ et $\alpha_{\sigma(j)\,j} = -m_{\sigma(j)\,j}$, et le terme $\varepsilon(\sigma)\prod_{i=1}^{n} \alpha_{\sigma(i)\,i}$ correspondant à cette permutation est de degré inférieur ou égal à $n-2$. Il en découle que

$$\begin{aligned}\chi_M(X) &= (X - m_{11})\dots(X - m_{nn}) + \text{termes de degré inférieur ou égal à } n-2\\ &= X^n - (m_{11} + \cdots + m_{nn})X^{n-1} + \text{termes de degré inférieur ou égal à } n-2\\ &= X^n - \operatorname{tr}(M)X^{n-1} + \text{termes de degré inférieur ou égal à } n-2.\end{aligned}$$

□

Remarque 1.20 *Pour une matrice M d'ordre deux, on a en particulier $\chi_M(X) = X^2 - \operatorname{tr}(M)X + \det(M)$.*

On retrouve avec ce dernier résultat qu'une matrice M d'ordre n possède au plus n valeurs propres. Par ailleurs, si $\mathbb{K} = \mathbb{C}$, cette matrice admet exactement n valeurs propres comptées avec leurs ordres de multiplicité respectifs. Il en va de même si $\mathbb{K} = \mathbb{R}$ et si le polynôme caractéristique est *scindé*, c'est-à-dire décomposable en un produit de facteurs de degré un. Dans ces deux cas, on peut écrire

$$\chi_M(X) = (X - \lambda_1)\dots(X - \lambda_n)$$

où les scalaires $\lambda_1,\dots,\lambda_n$ sont les valeurs propres, distinctes ou confondues, de la matrice M, ou bien

$$\chi_M(X) = (X - \lambda_1)^{m_{\lambda_1}}\dots(X - \lambda_p)^{m_{\lambda_p}}$$

où l'entier naturel p est inférieur ou égal à n, les scalaires $\lambda_1,\dots,\lambda_p$ sont les valeurs propres, distinctes deux à deux, de M et les entiers naturels $m_{\lambda_1},\dots,m_{\lambda_p}$ sont leurs ordres de multiplicité respectifs.

Proposition 1.21 (propriétés du polynôme caractéristique d'une matrice) *Soit A et B deux matrices carrées de même ordre. On a*

$$\chi_{A^\top} = \chi_A \text{ et } \chi_{AB} = \chi_{BA}.$$

DÉMONSTRATION. Soit A et B deux matrices d'ordre n. Le déterminant d'une matrice étant égal à celui de sa transposée, on a

$$\chi_{A^\top}(X) = \det\left(X\,I_n - A^\top\right) = \det\left((X\,I_n - A)^\top\right) = \det(X\,I_n - A) = \chi_A(X).$$

Pour la seconde assertion, si l'une des matrices, A par exemple, est inversible, alors $A^{-1}(AB)A = BA$, d'où AB et BA sont semblables. Dans le cas général, on considère les matrices de $M_{2n}(\mathbb{K})$ décomposées par blocs

$$M = \begin{pmatrix} BA & -B \\ 0 & 0 \end{pmatrix},\ N = \begin{pmatrix} 0 & -B \\ 0 & AB \end{pmatrix} \text{ et } P = \begin{pmatrix} I_n & 0 \\ A & I_n \end{pmatrix}.$$

On vérifie que $PN = MP = \left(\begin{smallmatrix} 0 & -B \\ 0 & 0 \end{smallmatrix}\right)$ et que P est une matrice triangulaire inférieure à diagonale non nulle, donc inversible. Les matrice M et N sont donc semblables. Le calcul par blocs de leurs polynômes caractéristiques respectifs laisse alors apparaître que $\chi_M(X) = \chi_{BA}(X)X^n$ et $\chi_N(X) = X^n\,\chi_{AB}(X)$, d'où la conclusion. □

1.4 Diagonalisation

Soit M une matrice. Pour plusieurs applications décrites dans la section 1.8, la suite des puissances itérées M^k pour $k \geq 0$ permet de comprendre le comportement d'un système (biologique, économique, etc.) en temps long. Cependant, calculer des multiplications matricielles est lourd, que ce soit pour un·e étudiant·e devant multiplier des matrices d'ordre 4, ou pour un ordinateur devant calculer des produits de matrices d'ordre atteignant dans certaines applications la dizaine de milliers. Calculer des puissances d'une matrice diagonale est par contre particulièrement facile : si $D = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$, alors $D^k = \begin{pmatrix} \lambda_1^k & 0 & 0 \\ 0 & \lambda_2^k & 0 \\ 0 & 0 & \lambda_3^k \end{pmatrix}$. De plus, si la matrice M est semblable à une matrice diagonale D (via une matrice de passage P), alors

$$M^3 = (PDP^{-1})(PDP^{-1})(PDP^{-1}) = PD(P^{-1}P)D(P^{-1}P)DP^{-1} = PD^3P^{-1},$$

et de manière plus générale, $M^k = PD^kP^{-1}$. Ainsi, si on trouve une matrice diagonale semblable à M, calculer M^k devient beaucoup plus facile : on passe du calcul (lourd) de k multiplications matricielles au calcul de D^k (qui est immédiat. On voit avec cet exemple que trouver une matrice diagonale semblable à une matrice M est particulièrement pratique pour effectuer des calculs. Si une telle matrice existe, on dit alors que M est diagonalisable.

Proposition 1.22 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. Si λ est une valeur propre de u d'ordre de multiplicité algébrique m_λ, alors on a*

$$1 \leq \dim(E_\lambda) \leq m_\lambda.$$

DÉMONSTRATION. Soit $(e_1, \ldots, e_{d_\lambda})$ une base du sous-espace propre E_λ, l'entier d_λ étant la dimension de E_λ, que l'on complète, si nécessaire, en une base $\mathscr{B}$ de E. La matrice représentative M de u dans la base $\mathscr{B}$ s'écrit

$$M = \begin{pmatrix} \lambda\,I_{d_\lambda} & M_{12} \\ 0 & M_{22} \end{pmatrix},$$

où la matrice M_{22} est d'ordre $n - d_\lambda$, avec n la dimension de E. En utilisant cette écriture par blocs, on trouve que

$$\chi_u(X) = \det(X\,I_n - M) = \det\left(\begin{pmatrix} (X-\lambda)\,I_{d_\lambda} & -M_{12} \\ 0 & X\,I_{n-d_\lambda} - M_{22} \end{pmatrix}\right) = (X-\lambda)^{d_\lambda}\chi_{M_{22}}(X).$$

Par conséquent, le polynôme $(X-\lambda)^{d_\lambda}$ divise $\chi_u(X)$ et l'ordre de multiplicité de λ est donc supérieur ou égal à d_λ. □

Remarque 1.23 *On a en particulier que le sous-espace propre E_λ est de dimension égale à un si λ est racine simple du polynôme caractéristique.*

On nomme parfois *ordre de multiplicité géométrique* de la valeur propre λ l'entier $\dim(E_\lambda)$, en contraste avec l'ordre de multiplicité algébrique m_λ.

Définition 1.24 (endomorphisme diagonalisable) *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. On dit que u est* ***diagonalisable*** *s'il existe une base de E formée de vecteurs propres de u.*

Il résulte de la définition précédente qu'une matrice carrée est diagonalisable si elle représente un endomorphisme diagonalisable.

Le résultat suivant donne une justification au vocabulaire employé.

Proposition 1.25 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. Les conditions suivantes sont équivalentes.*

(i) L'endomorphisme u est diagonalisable.

(ii) Il existe une base de E dans laquelle la matrice représentative de u est diagonale.

DÉMONSTRATION. On note n la dimension de E. Si $\mathscr{B} = (e_1, \dots, e_n)$ est une base de E formée de vecteurs propres de u associés aux valeurs propres $\lambda_1, \dots, \lambda_n$, comptées avec leurs ordres de multiplicité algébrique respectifs, la matrice représentative de u dans la base $\mathscr{B}$ est la matrice diagonale

$$\begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{pmatrix}.$$

Réciproquement, si $\mathscr{B}$ est une base de E, dire que la matrice représentative de u dans la base $\mathscr{B}$ est la matrice diagonale

$$\begin{pmatrix} \mu_1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & \mu_n \end{pmatrix}$$

revient à dire que $u(e_1) = \mu_1 e_1, \dots, u(e_n) = \mu_n e_n$. Les vecteurs de cette base sont donc des vecteurs propres de u et l'endomorphisme est donc diagonalisable. □

On dira qu'une matrice M d'ordre n est diagonalisable si et seulement si elle est semblable à une matrice diagonale. En notant $\mathscr{U}$ une base de $M_{n,1}(K)$ formée de vecteurs propres associés aux valeurs propres $\lambda_1, \dots, \lambda_n$ de M, $\mathscr{B}$ la base canonique de $M_{n,1}(K)$, P la matrice de passage de la base $\mathscr{B}$ à la base $\mathscr{U}$ et D la matrice diagonale telle que $d_{ii} = \lambda_i$, $i = 1, \dots, n$, on a alors l'égalité

$$M = PDP^{-1}.$$

Théorème 1.26 (condition nécessaire et suffisante de diagonalisabilité) *Soit un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle. Cet endomorphisme est diagonalisable si et seulement si son polynôme caractéristique est scindé sur $\mathbb{K}$ et l'ordre de multiplicité algébrique de chaque valeur propre est égal à la dimension du sous-espace propre correspondant.*

DÉMONSTRATION. Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle égale à n et $u \in \mathscr{L}(E)$.

Supposons tout d'abord que le polynôme caractéristique χ_u soit scindé sur $\mathbb{K}$ et que l'ordre de multiplicité algébrique de chaque valeur propre soit égal à la dimension du sous-espace propre correspondant. Posons $\chi_u(X) = \prod_{i=1}^{p}(X - \lambda_i)^{m_{\lambda_i}}$, où les scalaires λ_i, $i = 1, \dots, p$, sont les valeurs propres deux à deux distinctes de u et les entiers m_{λ_i}, $i = 1, \dots, p$, sont leurs ordres de multiplicité respectifs. Pour tout entier i de $\{1, \dots, p\}$, posons par ailleurs $n_{\lambda_i} = \dim(E_{\lambda_i})$. Par hypothèse, on a, pour tout entier i de $\{1, \dots, p\}$, $n_{\lambda_i} = m_{\lambda_i}$, ce qui implique alors que

$$n = \deg(\chi_u) = \sum_{i=1}^{p} m_{\lambda_i} = \sum_{i=1}^{p} n_{\lambda_i}.$$

Les sous-espaces propres étant en somme directe en vertu de la proposition 1.10, il existe donc une base de E formée de vecteurs propres de u et la proposition 1.25 permet alors d'affirmer que l'endomorphisme est diagonalisable.

Réciproquement, supposons que l'endomorphisme u soit diagonalisable. L'espace vectoriel E est somme directe des sous-espaces propres de u, c'est-à-dire

$$E = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_p},$$

où les scalaires λ_i, $i = 1, \dots, p$, sont les valeurs propres deux à deux distinctes de u. Dans une base de E associée à cette décomposition en somme directe, la matrice représentative de u s'écrit

$$D = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_p & & \\ & & & & & \ddots & \\ & & & & & & \lambda_p \end{pmatrix},$$
$$\underbrace{\qquad}_{\dim(E_{\lambda_1})\text{ fois}} \qquad \underbrace{\qquad}_{\dim(E_{\lambda_p})\text{ fois}}$$

ce qui implique que

$$\chi_u(X) = \chi_D(X) = \prod_{i=1}^{p} (X - \lambda_i)^{\dim(E_{\lambda_i})}.$$

Le polynôme caractéristique de u est donc scindé sur $\mathbb{K}$ et l'ordre de multiplicité de chaque valeur propre correspond à la dimension du sous-espace propre associé. □

On a évidemment un énoncé similaire pour une matrice carrée. La preuve du résultat suivant est laissée au lecteur.

Corollaire 1.27 (condition suffisante de diagonalisabilité) *Soit un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle égale à n. Si cet endomorphisme possède n valeurs propres distinctes, alors il est diagonalisable et ses sous-espaces propres sont des droites vectorielles.*

La diagonalisation d'une matrice en pratique

Soit M une matrice d'ordre n à coefficients dans le corps $\mathbb{K}$.

1. Calculer le polynôme caractéristique χ_M, trouver ses racines et une factorisation associée du polynôme pour obtenir les valeurs propres avec leurs ordres de multiplicité respectifs. Si le polynôme n'est pas scindé, la matrice n'est pas diagonalisable.
2. Si le polynôme caractéristique est scindé, décrire pour chaque valeur propre le sous-espace propre qui lui est associé.
3. Pour chacune des valeurs propres, comparer l'ordre de multiplicité algébrique de la valeur propre avec la dimension du sous-espace propre correspondant. Si ces deux nombres sont égaux pour *toutes* les valeurs propres, alors la matrice est diagonalisable. S'il existe au moins une valeur propre pour laquelle ce n'est pas le cas, la matrice n'est pas diagonalisable.
4. Si la matrice est diagonalisable et si $\{V_1, \dots, V_n\}$ est une base de $M_{n,1}(\mathbb{K})$ formée de vecteurs propres associés aux valeurs propres $\lambda_1, \dots, \lambda_n$ de M comptées avec leur ordre de multiplicité, alors la matrice P dont les colonnes sont les vecteurs V_i, $i = 1, \dots, n$, est telle que

$$M = P \operatorname{diag}(\lambda_1, \dots, \lambda_n) P^{-1}.$$

1.5 Trigonalisation

On a vu dans la section précédente que tout endomorphisme n'est pas nécessairement diagonalisable. On va maintenant montrer que l'on peut en revanche toujours trouver une base de l'espace par rapport à laquelle la matrice d'un endomorphisme défini sur un $\mathbb{C}$-espace vectoriel est une matrice triangulaire.

Définition 1.28 (endomorphisme trigonalisable) *Soit un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et un endomorphisme de E. On dit que cet endomorphisme est **trigonalisable** si et seulement s'il existe une base de l'espace dans laquelle sa matrice est triangulaire supérieure.*

On notera que, dans la définition précédente, on aurait pu tout aussi bien remplacer « triangulaire supérieure » par « triangulaire inférieure ».

Par extension, une matrice carrée sera dite trigonalisable si et seulement si elle est semblable à une matrice triangulaire supérieure. Les valeurs propres d'une matrice trigonalisable sont ainsi données par les coefficients diagonaux d'une matrice triangulaire qui lui est semblable.

Théorème 1.29 (condition nécessaire et suffisante de trigonalisabilité) *Un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle est trigonalisable si et seulement si son polynôme caractéristique est scindé sur $\mathbb{K}$. En particulier, tout endomorphisme d'un $\mathbb{C}$-espace vectoriel est trigonalisable.*

DÉMONSTRATION. Considérons que l'espace est de dimension n et soit M une matrice de $M_n(\mathbb{K})$ représentant l'endomorphisme. Si l'endomorphisme est trigonalisable, il existe une matrice P de $GL_n(\mathbb{K})$ et une matrice T triangulaire supérieure telles que $M = PTP^{-1}$. En posant

$$T = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \dots & 0 & \lambda_n \end{pmatrix},$$

les coefficients λ_i, $i = 1, \dots, n$, appartenant à $\mathbb{K}$, on trouve alors que $\chi_M(X) = \chi_T(X) = \prod_{i=1}^{n}(X - \lambda_i)$. Le polynôme caractéristique de l'endomorphisme est donc scindé sur $\mathbb{K}$.

Montrons à présent la réciproque en raisonnant par récurrence sur l'ordre de la matrice. Si $n = 1$, tout élément de $M_n(\mathbb{K})$ est triangulaire et donc trigonalisable. Supposons à présent que l'entier n soit supérieur ou égal à 1 et que tout élément de $M_n(\mathbb{K})$ dont le polynôme caractéristique est scindé sur $\mathbb{K}$ soit trigonalisable. Soit une matrice M de $M_{n+1}(\mathbb{K})$, telle que χ_M est scindé sur $\mathbb{K}$. L'endomorphisme de $\mathbb{K}^{n+1}$ canoniquement associé à M admet au moins une valeur propre λ_1 appartenant à $\mathbb{K}$. Soit v_1 un vecteur propre associé à λ_1. La famille $\{v_1\}$ étant libre, on peut la compléter en une base de $\mathbb{K}^{n+1}$, la matrice de l'endomorphisme dans cette base s'écrivant alors sous la forme par blocs

$$M' = \begin{pmatrix} \lambda_1 & L \\ 0 & M_1 \end{pmatrix},$$

avec L appartenant à $M_{1,n}(\mathbb{K})$ et M_1 appartenant à $M_n(\mathbb{K})$. Les matrices M et M' étant semblables, il existe une matrice P' de $GL_{n+1}(\mathbb{K})$ telle que

$$M = P'M'(P')^{-1},$$

et un calcul par blocs du déterminant montre que $\chi_{M'}(X) = (X - \lambda_1)\det(X\,I_n - M_1) = (X - \lambda_1)\chi_{M_1}(X)$. D'autre part, χ_M étant scindé sur $\mathbb{K}$, on peut poser $\chi_M(X) = \prod_{i=1}^{n+1}(X - \lambda_i)$, dont on déduit que $\chi_{M_1}(X) = \prod_{i=2}^{n+1}(X - \lambda_i)$. L'hypothèse de récurrence assure alors l'existence d'une matrice P_1 de $GL_n(\mathbb{K})$ et d'une matrice triangulaire supérieure T_1 telles que $M_1 = P_1 T_1 {P_1}^{-1}$.

Il suffit ensuite de poser

$$P'' = \begin{pmatrix} 1 & 0 \\ 0 & P_1 \end{pmatrix}.$$

Par un calcul par blocs, on a $\det(P'') = \det(P_1)$, d'où P'' est inversible et, toujours en calculant par blocs, il vient

$$(P'')^{-1} = \begin{pmatrix} 1 & 0 \\ 0 & {P_1}^{-1} \end{pmatrix}$$

et

$$(P'')^{-1}M'P'' = \begin{pmatrix} \lambda_1 & LP_1 \\ 0 & T_1 \end{pmatrix} = T.$$

Posons finalement $P = P'P''$. Cette dernière matrice est inversible et l'on a

$$P^{-1}MP = (P'')^{-1}(P')^{-1}MP'P'' = (P'')^{-1}M'P'' = T.$$

L'endomorphisme représenté par M est donc bien trigonalisable. $\square$

Le dernier théorème est vrai pour une matrice carrée à coefficients réels pour peu que l'on factorise son polynôme caractéristique sur $\mathbb{C}$ et non sur $\mathbb{R}$. C'est en ce sens que l'on peut dire, comme on l'a écrit en début de section, que l'on peut toujours trouver une base dans laquelle la matrice d'un endomorphisme défini sur un $\mathbb{C}$-espace vectoriel est triangulaire.

La preuve du résultat suivant est laissée au lecteur.

Corollaire 1.30 *Soit E un $\mathbb{C}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$ pour lequel le polynôme caractéristique peut s'écrire*

$$\chi_u(X) = \prod_{i=1}^{p} (X - \lambda_i)^{m_{\lambda_i}}.$$

On a alors

$$\operatorname{tr}(u) = \sum_{i=1}^{p} m_{\lambda_i} \lambda_i \ \textit{et} \ \det(u) = \prod_{i=1}^{p} {\lambda_i}^{m_{\lambda_i}}.$$

La trigonalisation d'une matrice en pratique

Soit M une matrice d'ordre n à coefficients dans le corps $\mathbb{K}$.

1. Calculer le polynôme caractéristique χ_M, trouver ses racines et une factorisation associée du polynôme pour obtenir les valeurs propres. Si le polynôme caractéristique n'est pas scindé, la matrice n'est pas trigonalisable.
2. Si le polynôme caractéristique est scindé, il s'agit à présent de trouver une base dans laquelle la matrice de l'endomorphisme représenté par M sera triangulaire supérieure.

   On a intérêt à placer dans cette base le plus grand nombre possible de vecteurs propres de la matrice, en déterminant pour cela une base de chaque sous-espace propre. La réunion de ces bases constitue bien une partie de la base recherchée, puisque les sous-espaces propres sont en somme directe. Le choix de ces premiers vecteurs (ici au nombre de p) impose que la matrice triangulaire sera de la forme

$$T = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 & * & \dots & \dots & * \\ 0 & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & & \vdots \\ \vdots & & \ddots & \lambda_p & * & & & \vdots \\ \vdots & & & \ddots & \lambda_{p+1} & \ddots & & \vdots \\ \vdots & & & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & & & \ddots & \ddots & * \\ 0 & \dots & \dots & \dots & \dots & \dots & 0 & \lambda_n \end{pmatrix}.$$

   Il reste à compléter la famille obtenue pour arriver à une base. On peut pour cela s'y prendre de la façon suivante.

   Procédons itérativement en supposant que l'on a déjà déterminé les vecteurs $V_1, \dots, V_j$, avec $p \leq j \leq n-1$. Afin de choisir le vecteur suivant, on commence par compléter la famille $\{V_1, \dots, V_j\}$ en une base $\{V_1, \dots, V_j, U_{j+1}, \dots, U_n\}$ de $M_{n,1}(\mathbb{K})$, puis on cherche V_{j+1} sous la forme $V_{j+1} = \sum_{i=j+1}^{n} \alpha_i U_i$, la forme de la matrice T imposant que $MV_{j+1} = \sum_{i=1}^{j} t_{i\,j+1} V_i + \lambda_{j+1} V_{j+1}$. En explicitant cette relation, on obtient n équations linéaires dont les inconnues sont les n coefficients $t_{1\,j+1}, \dots, t_{j\,j+1}, \alpha_{j+1}, \dots, \alpha_n$:

$$\sum_{i=1}^{j} t_{i\,j+1} V_i + \sum_{i=j+1}^{n} \alpha_i (\lambda_{j+1} U_i - M U_i) = 0.$$

Toute solution non nulle de ce système fournit un vecteur V_{j+1} possible et les coefficients possiblement non nuls correspondants de la $j+1^{\text{e}}$ colonne de T.

1.6 Polynômes annulateurs

Soit $P(X) = a_0 + a_1 X + a_2 X^2 + \dots$ un polynôme de $\mathbb{K}[X]$. Pour tout endomorphisme u d'un $\mathbb{K}$-espace vectoriel E de dimension finie non nulle égale à n, on définit

$$P(u) = a_0\, id_E + a_1\, u + a_2\, u^2 + \dots, \tag{1.1}$$

où, pour tout entier naturel non nul k, on a posé $u^k = \underbrace{u \circ \cdots \circ u}_{k \text{ termes}}$.

La preuve du prochain résultat est laissée en exercice au lecteur.

Proposition 1.31 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. L'application de $\mathbb{K}[X]$ dans $\mathscr{L}(E)$ qui a tout polynôme P associe l'application linéaire $P(u)$ définie par (1.1) est un* ***morphisme d'algèbres****, c'est-à-dire qu'elle est linéaire, i.e.*

$$\forall \lambda \in \mathbb{K},\ \forall (P, Q) \in (\mathbb{K}[X])^2,\ (\lambda P + Q)(u) = \lambda P(u) + Q(u),$$

et vérifie

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

On remarquera en particulier que, pour tous polynômes P et Q et tout endomorphisme u, les endomorphismes $P(u)$ et $Q(u)$ commutent entre eux.

On peut de la même manière définir un morphisme d'algèbres sur l'ensemble des matrices carrées en posant, pour toute matrice M de $M_n(\mathbb{K})$,

$$P(M) = a_0 I_n + a_1 M + a_2 M^2 + \dots$$

Définition 1.32 (polynôme annulateur) *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$. On dit qu'un polynôme P est un* ***polynôme annulateur*** *de u si l'endomorphisme $P(u)$ est nul.*

Une définition similaire vaut pour le polynôme annulateur d'une matrice carrée.

Exemple 1.33 *On rappelle qu'une projection u est un endomorphisme de E tel que $u^2 = u \circ u = u$. Ainsi, $X^2 - X$ est un polynôme annulateur de u.*

L'ensemble des polynômes annulateurs d'un même endomorphisme forme un *idéal*[2].

Le résultat suivant établit un lien entre les valeurs propres d'un endomorphisme et les racines de tout polynôme annulateur de cet endomorphisme.

Proposition 1.34 *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et P un polynôme annulateur de u. On a l'inclusion*

$$\mathrm{Sp}(u) \subseteq \{\textit{racines de } P\}.$$

DÉMONSTRATION. Si x est un vecteur propre de u associé à une valeur propre λ, on a $u(x) = \lambda\, x$, ce qui implique que, pour tout entier naturel k, $u^k(x) = \lambda^k x$ et, plus généralement que, pour tout polynôme P de $\mathbb{K}[X]$, $P(u)(x) = P(\lambda)x$. En particulier, avoir $P(u)(x) = 0_E$ implique que $P(\lambda)x = 0_E$ et donc que $P(\lambda) = 0$, puisque le vecteur x est non nul. □

Attention : toute racine d'un polynôme annulateur n'est pas nécessairement une valeur propre.

Théorème 1.35 (« théorème de Cayley–Hamilton ») *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle. On a*

$$\chi_u(u) = 0_{\mathscr{L}(E)}.$$

2. Un idéal est un sous-ensemble remarquable d'un anneau : c'est un sous-groupe du groupe additif de l'anneau, qui est de plus stable par multiplication par les éléments de l'anneau.

DÉMONSTRATION. Notons E le sous-espace de l'énoncé et n la dimension de E. On va montrer que, pour tout vecteur x non nul de E, $\chi_u(u)(x) = 0_E$. Soit $\mathscr{E}$ le sous-ensemble de $\mathbb{N}$ défini par

$$\mathscr{E} = \left\{k \in \mathbb{N}^* \mid \{u^i(x)\}_{0\leq i\leq k-1} \text{ est une famille libre}\right\}.$$

C'est une partie non vide de $\mathbb{N}$ (puisque x est non nul), majorée par n. Elle admet donc un plus grand élément, que l'on note p. Par définition de p, la famille $\{x, u(x), \dots, u^{p-1}(x)\}$ est libre et on peut la compléter au besoin en une base $\mathscr{B} = \{x, u(x), \dots, u^{p-1}(x), e_{p+1}, \dots, e_n\}$ de E. Toujours par définition de p, la famille $\{x, u(x), \dots, u^p(x)\}$ est liée et l'on peut donc poser $u^p(x) = \sum_{i=0}^{p-1} a_i\, u^i(x)$. La matrice de u dans la base $\mathscr{B}$ s'écrit par blocs

$$\begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$$

avec

$$B = \begin{pmatrix} 0 & \dots & \dots & 0 & a_0 \\ 1 & \ddots & & \vdots & a_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \dots & 0 & 1 & a_{p-1} \end{pmatrix} \tag{1.2}$$

une matrice d'ordre p. Un calcul par blocs du déterminant définissant χ_u conduisant à $\chi_u = \chi_B\chi_D$, un développement par rapport à la dernière colonne du déterminant fournissant χ_B donne pour sa part

$$\chi_B(X) = X^{p-1}(X - a_{p-1}) + \sum_{i=0}^{p-2}(-1)^{p+i+1}(-a_i)\,\Delta_i,$$

où

$$\Delta_i = \begin{vmatrix} X & 0 & \dots & 0 & * & \dots & \dots & * \\ * & \ddots & \ddots & \vdots & \vdots & & & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & & & \vdots \\ * & \dots & * & X & * & \dots & \dots & * \\ 0 & \dots & \dots & 0 & -1 & * & \dots & * \\ \vdots & & & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & & & \vdots & \vdots & \ddots & \ddots & * \\ 0 & \dots & \dots & 0 & 0 & \dots & 0 & -1 \end{vmatrix} = X^i(-1)^{p-1-i}.$$

Par suite, on trouve que

$$\chi_B(X) = X^p - a_{p-1}X^{p-1} - \sum_{i=0}^{p-2} a_i X^i = X^p - \sum_{i=0}^{p-1} a_i X^i,$$

On a finalement

$$\chi_u(u) = (u^p - \sum_{i=0}^{p-1} a_i\, u^i) \circ \chi_D(u).$$

Puisque ces deux polynômes de u commutent entre eux, on a obtenu que

$$\chi_u(u)(x) = \chi_D(u)\left(u^p(x) - \sum_{i=0}^{p-1} a_i\, u^i(x)\right) = \chi_D(u)(0_E) = 0_E.$$

Cette égalité restant vraie pour $x = 0_E$, on a prouvé le résultat. □

Ce résultat montre que le polynôme caractéristique (et tous ses multiples) est un polynôme annulateur de l'endomorphisme auquel il est associé.

Remarque 1.36 *Dans la démonstration ci-dessus, on a cherché le plus petit sous-espace vectoriel stable par u contenant le vecteur x. Dans la base choisie pour ce sous-espace vectoriel, la matrice de l'endomorphisme induit par u a la forme d'une* matrice compagnon *(voir* (1.2)*). L'endomorphisme induit par u sur le sous-espace est alors dit* cyclique.

Définition 1.37 (polynôme minimal) *Un* ***polynôme minimal*** *d'un endomorphisme est un polynôme annulateur de cet endomorphisme, unitaire et de degré minimal.*

Proposition 1.38 *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle. Il existe un unique polynôme minimal de u, noté μ_u, et celui-ci divise tout polynôme annulateur de u*

DÉMONSTRATION. Notons E l'espace vectoriel de l'énoncé et n sa dimension. Considérons l'ensemble des degrés des polynômes annulateurs de u :

$$\{\deg(P)\,|\,P \in \mathbb{K}[X],\ P \neq 0_{\mathbb{K}[X]},\ P(u) = 0_{\mathscr{L}(E)}\}.$$

C'est une partie non vide de $\mathbb{N}$, car elle contient l'entier n en vertu du théorème de Cayley–Hamilton. Soit m son plus petit élément. Considérons un polynôme annulateur Π de u de degré m. On va montrer que tout polynôme annulateur P de u est un multiple de Π. Pour cela, on effectue la division euclidienne de P par Π :

$$P = \Pi Q + R,$$

où $\deg(R) \leq m-1$. Si $P(u) = 0_{\mathscr{L}(E)}$ et $\Pi(u) = 0_{\mathscr{L}(E)}$, alors $R(u) = 0_{\mathscr{L}(E)}$, ce qui implique que R est nul et le polynôme P est donc multiple de Π.

Enfin, si Π_1 et Π_2 sont deux polynômes annulateurs de u de degré m, ils sont multiples l'un de l'autre et ne diffèrent par conséquent que d'une constante multiplicative. On a donc unicité du polynôme minimal si l'on suppose que le coefficient de son terme de plus haut degré est fixé, ce que l'on a fait en posant qu'il est égal à 1. □

Proposition 1.39 *Soit un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle. Les racines du polynôme minimal de cet endomorphisme sont exactement les valeurs propres de cet endomorphisme.*

DÉMONSTRATION. Compte tenu de la proposition 1.34, on a juste à montrer que toute racine du polynôme minimal d'un endomorphisme est une valeur propre de cet endomorphisme. Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel E de dimension finie non nulle dont le scalaire λ est racine du polynôme minimal. Dans ce cas, on a $\mu_u(X) = (X-\lambda)Q(X)$, avec $\deg(Q) \leq \deg(\mu_u) - 1$ et on a donc

$$0_{\mathscr{L}(E)} = \mu_u(u) = (u - \lambda\, id_E) \circ Q(u).$$

Puisque $Q(u) \neq 0_{\mathscr{L}(E)}$ par minimalité de μ_u, l'endomorphisme $u - \lambda\, id_E$ n'est pas injectif et λ est une donc valeur propre de u. □

Le résultat suivant caractérise le polynôme minimal d'un endomorphisme dont le polynôme caractéristique est scindé.

Théorème 1.40 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $u \in \mathscr{L}(E)$ dont le polynôme caractéristique est scindé :*

$$\chi_u(X) = \prod_{i=1}^{p} (X - \lambda_i)^{m_{\lambda_i}},$$

où les entiers naturels m_{λ_i}, $i = 1, \ldots, p$, sont non nuls et les valeurs propres λ_i, $i = 1, \ldots, p$, sont deux à deux distinctes. Alors on a

$$\mu_u(X) = \prod_{i=1}^{p} (X - \lambda_i)^{l_{\lambda_i}},$$

où les entiers naturels l_{λ_i}, $i = 1, \ldots, p$, sont respectivement compris entre 1 *et m_{λ_i}.*

DÉMONSTRATION. On a précédemment montré que toute racine du polynôme minimal μ_u est une valeur propre de u. On a donc nécessairement que l_{λ_i} est supérieur ou égal à 1 pour tout entier i de $\{1, \ldots, p\}$. On sait par ailleurs que le polynôme minimal μ_u divise le polynôme caractéristique χ_u, d'où l_{λ_i} est inférieur ou égal à m_{λ_i} pour tout entier i de $\{1, \ldots, p\}$. □

Exemple 1.41 *Considérons une homothétie de rapport $\lambda \neq 0$, $u = \lambda\, id_E$. La matrice de u dans n'importe quelle base de E est diagonale, de coefficients diagonaux tous égaux à λ, d'où $\chi_u(X) = (X-\lambda)^n$. Cependant, le polynôme $X-\lambda$ est un polynôme annulateur. Comme il est de degré 1, ce polynôme est nécessairement le polynôme minimal : $\mu_u(X) = X-\lambda$.*

L'exemple précédent est particulièrement éclairant, car diagonaliser un endomorphisme revient à écrire l'espace comme une somme directe de sous-espaces propres de cet endomorphisme pour chacun desquels la restriction de l'endomorphisme est une homothétie. Cette observation conduit à une nouvelle caractérisation de la diagonalisation.

Théorème 1.42 *Un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle dont le polynôme caractéristique est scindé sur $\mathbb{K}$ est diagonalisable si et seulement si son polynôme minimal est scindé à racines simples.*

DÉMONSTRATION. Notons E l'espace vectoriel et u l'endomorphisme de l'énoncé. Considérons le polynôme $P(X) = \prod_{i=1}^{p}(X-\lambda_i)$ dont les racines sont les valeurs propres de u. Soit x un vecteur propre de u associé à une valeur propre λ_j. L'image de x par $P(u)$ est $P(u)(x) = \prod_{i=1}^{p}(\lambda_j-\lambda_i)x = 0_E$.

Supposons que u est diagonalisable. Il existe alors une base E formée de vecteurs propres de u et $P(u)$ s'annule sur chacun des vecteurs composant cette base. Le polynôme P est donc un polynôme annulateur de u, multiple du polynôme minimal μ_u. Or, par le théorème 1.40, on a que P divise μ_u, d'où $P = \mu_u$.

Réciproquement, supposons que le polynôme minimal de u soit

$$\mu_u(X) = \prod_{i=1}^{p}(X-\lambda_i).$$

Les sous-espaces propres étant en somme directe, on doit simplement montrer que tout vecteur de E s'écrit comme une combinaison linéaire de vecteurs propres. Par une décomposition en éléments simples, on a

$$\frac{1}{\prod_{k=1}^{p}(X-\lambda_k)} = \sum_{i=1}^{p}\left(\frac{1}{\prod_{j=1,j\neq i}^{p}(\lambda_i-\lambda_j)}\right)\frac{1}{X-\lambda_i}.$$

En multipliant cette égalité par le dénominateur du membre de gauche, il vient [3]

$$1 = \sum_{i=1}^{p}\left(\frac{1}{\prod_{j=1,j\neq i}^{p}(\lambda_i-\lambda_j)}\right)\prod_{k=1,k\neq i}^{p}(X-\lambda_k).$$

Posons alors $\alpha_i = \frac{1}{\prod_{j=1,j\neq i}^{p}(\lambda_i-\lambda_j)}$, $P_i(X) = \prod_{k=1,k\neq i}^{p}(X-\lambda_k)$. On a que

$$id_E = \sum_{i=1}^{p}\alpha_i\, P_i(u),$$

d'où

$$\forall x \in E,\ x = \sum_{i=1}^{p}\alpha_i\, P_i(u)(x).$$

Enfin, en utilisant que

$$\forall i \in \{1,\dots,p\},\ P_i(X)(X-\lambda_i) = \prod_{j=1}^{p}(X-\lambda_j) = \mu_u(X),$$

on trouve que

$$0_E = \mu_u(u)(x) = (u-\lambda_i\, id_E)(P_i(u)(x)),$$

d'où $P_i(u)(x)$ appartient à E_{λ_i}. □

Nous concluons cette section avec un résultat riche de conséquences. Avant de l'énoncer, rappelons tout d'abord que deux polynômes P et Q de $\mathbb{K}[X]$ sont dits *premiers entre eux* si et seulement si

$$D \text{ divise } P \text{ et } Q \text{ dans } \mathbb{K}[X] \implies D \text{ est un polynôme constant.}$$

3. C'est l'écriture du polynôme constant et égal à 1 dans la base des polynômes de Lagrange associés aux valeurs propres de u.

En particulier, deux polynômes à coefficients complexes sont premiers entre eux si et seulement s'ils n'ont pas de racines communes.

D'après le théorème de Bézout, le fait que les polynômes P et Q sont premiers entre eux équivaut à ce qu'il existe des polynômes R et S tels que $RP + SQ = 1$.

Proposition 1.43 (« lemme de décomposition des noyaux ») *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel E et P et Q des polynômes premiers entre eux. On a*

$$\ker((PQ)(u)) = \ker(P(u)) \oplus \ker(Q(u)).$$

DÉMONSTRATION. Soit x un vecteur appartenant à $\ker(P(u))$. On a alors

$$P(u)(x) = 0_E \implies Q(u)(P(u)(x)) = 0_E \implies (P(u) \circ Q(u))(x) = 0_E \implies (PQ)(u)(x) = 0_E,$$

d'où x appartient à $\ker((PQ)(u))$. Ainsi, on a prouvé que $\ker(P(u)) \subset \ker((PQ)(u))$. De la même manière, on a $\ker(Q(u)) \subset \ker((PQ)(u))$.

Les polynômes P et Q étant premiers entre eux, on a que

$$R(u) \circ P(u) + S(u) \circ Q(u) = id_E.$$

En évaluant cette dernière identité en un vecteur x de $\ker((PQ)(u))$, on obtient

$$(R(u) \circ P(u))(x) + (S(u) \circ Q(u))(x) = x.$$

Posons alors $y = (S(u) \circ Q(u))(x)$ et $z = (R(u) \circ P(u))(x)$. Il vient

$$P(u)(y) = P(u)((S(u) \circ Q(u))(x)) = S(u)((P(u) \circ Q(u))(x)) = S(u)(0_E) = 0_E,$$

et, de la même façon, $Q(u)(z) = 0_E$. On a ainsi que y appartient à $\ker(P(u))$ et que z appartient à $\ker(Q(u))$,d'où $\ker((PQ)(u)) = \ker(P(u)) + \ker(Q(u))$. Enfin, pour tout vecteur x de $\ker(P(u)) \cap \ker(Q(u))$, on a, toujours en utilisant l'identité, $x = 0_E + 0_E = 0_E$, d'où $\ker(P(u)) \cap \ker(Q(u)) = \{0_E\}$. □

Il est possible de généraliser ce résultat par récurrence.

Corollaire 1.44 (« lemme de décomposition des noyaux généralisé ») *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel, k un entier naturel strictement plus grand que 1 et $P_1, \dots, P_k$ des polynômes deux à deux premiers entre eux. On a*

$$\ker((P_1 \dots P_k)(u)) = \ker(P_1(u)) \oplus \cdots \oplus \ker(P_k(u)).$$

Une ultime caractérisation de la diagonalisabilité découle de ce dernier résultat.

Théorème 1.45 *Un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle est diagonalisable si et seulement s'il existe un polynôme non nul, scindé sur $\mathbb{K}$ et à racines simples, annulateur de cet endomorphisme.*

DÉMONSTRATION. Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel E de dimension finie non nulle.

L'implication directe du théorème découle du théorème 1.42. Pour montrer l'implication réciproque, on suppose qu'il existe un polynôme P annulateur de u, de la forme

$$P(X) = \prod_{i=1}^{p} (X - \mu_i),$$

les scalaires μ_i étant deux à deux distincts. Les monômes $X - \mu_1, \dots, X - \mu_p$ étant deux à deux premiers entre eux, le corollaire 1.44 permet d'écrire que

$$E = \bigoplus_{i=1}^{p} \ker(u - \mu_i \, id_E).$$

Une base de E adaptée [4] à cette décomposition en somme directe est une base de vecteurs propres de u et l'endomorphisme est donc diagonalisable. □

On peut à présent résumer les conditions de diagonalisabilité et de trigonalisation établies dans ce chapitre.

4. Si $\ker(u - \mu_i \, id_E)$ se trouve réduit au vecteur nul, on considère qu'une « base » de ce noyau est l'ensemble vide.

Conditions de diagonalisabilité et de trigonalisabilité d'un endomorphisme d'un $\mathbb{K}$-espace vectoriel de dimension finie non nulle

Soit E un espace vectoriel sur $\mathbb{K}$ de dimension finie non nulle égale à n et $u \in \mathscr{L}(E)$. Ci-dessous, on a noté m_{λ_i} l'ordre de multiplicité algébrique de la valeur propre λ_i de u et n_{λ_i} la dimension du sous-espace propre associé $\ker(u - \lambda_i\, id_E)$, $i = 1, \dots, p$, χ_u le polynôme caractéristique de u et μ_u le polynôme minimal de u.

u est diagonalisable $\iff$ il existe une base de E dans laquelle la matrice de u est diagonale,
$\iff$ il existe une base de E constituée de vecteurs propres de u,
$\iff$ E est la somme directe des sous-espaces propres de u,
$\iff$ $n = \sum_{i=1}^{p} n_{\lambda_i}$,
$\iff$ χ_u est scindé sur $\mathbb{K}$ et, $\forall i \in \{1, \dots, p\}$, $n_{\lambda_i} = m_{\lambda_i}$,
$\iff$ μ_u est scindé sur $\mathbb{K}$, à racines simples,
$\iff$ il existe un polynôme P non nul, scindé sur $\mathbb{K}$, à racines simples et tel que $P(u) = 0_{\mathscr{L}(E)}$,
$\Longleftarrow$ u possède n valeurs propres simples $\iff$ χ_u est scindé sur $\mathbb{K}$, à racines simples.

u est trigonalisable $\iff$ il existe une base de E dans laquelle la matrice de u est triangulaire supérieure,
$\iff$ χ_u est scindé sur $\mathbb{K}$,
$\iff$ E est la somme directe des sous-espaces caractéristiques de u.

1.7 Réduction de Jordan (hors programme)

On peut aller plus loin que la simple trigonalisation en construisant une base de l'espace dans laquelle la matrice de l'endomorphisme est diagonale par blocs, chacun des blocs étant une matrice triangulaire supérieure dont les coefficients diagonaux sont identiques.

1.7.1 Sous-espaces caractéristiques

Définition 1.46 (sous-espace caractéristique) *Soit E un $\mathbb{K}$-espace vectoriel, u un endomorphisme de E et λ une valeur propre de u. On appelle* ***sous-espace caractéristique*** *(ou* ***sous-espace spectral****, ou encore* ***sous-espace propre généralisé****) de u associé à λ le sous-espace vectoriel*

$$F_\lambda = \ker\left((u - \lambda\, id_E)^{l_\lambda}\right),$$

où l'entier l_λ est l'ordre de multiplicité de λ en tant que racine du polynôme minimal de u. Un vecteur non nul x de E appartenant à F_λ est un ***vecteur propre généralisé*** *de u associé à λ.*

Proposition 1.47 *Soit u un endomorphisme d'un $\mathbb{K}$-espace vectoriel E de dimension finie non nulle. Supposons que le polynôme minimal μ_u de u soit scindé :*

$$\mu_u(X) = \prod_{i=1}^{p} (X - \lambda_i)^{l_{\lambda_i}}.$$

Alors, chaque espace caractéristique F_{λ_i} est stable par u. De plus, $\dim(F_{\lambda_i}) = m_{\lambda_i}$, où m_{λ_i} est l'ordre de multiplicité algébrique de λ_i. Enfin,

$$E = \bigoplus_{i=1}^{p} F_{\lambda_i}.$$

DÉMONSTRATION. Le dernier point est une conséquence directe du lemme de décomposition des noyaux généralisé. De plus, pour n'importe quel polynôme P, l'espace $\ker P(u)$ est stable par u (le montrer). Ceci démontre le premier point. Il reste à montrer que $\dim(F_{\lambda_i}) = m_{\lambda_i}$. Comme chacun des sous-espaces caractéristiques F_{λ_i} est stable par u, on peut construire une base $\mathscr{B}$ de E en concaténant des bases $\mathscr{B}_i$ de F_{λ_i}. On a alors

$$[u]_{\mathscr{B}} = \begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_p \end{pmatrix},$$

où M_i est une matrice d'ordre $\dim(F_{\lambda_i})$ représentant u restreint à F_i dans la base $\mathscr{B}_i$. On calcule le polynôme caractéristique :

$$\chi_u(X) = \chi_{M_1}(X) \cdots \chi_{M_p}(X)$$

Notons que λ_i est l'unique racine de χ_{M_i} (pourquoi ?). Comme χ_u est scindé, on a donc forcément

$$\chi_u(X) = (X - \lambda_1)^{\dim(F_{\lambda_1})} \cdots (X - \lambda_p)^{\dim(F_{\lambda_p})}.$$

Par définition des multiplicités algébriques, on a donc $\dim(F_{\lambda_i}) = m_{\lambda_i}$. $\square$

Le dernier résultat de la proposition montre qu'un endomorphisme d'un espace vectoriel est trigonalisable si et seulement si l'espace est la somme (directe) des sous-espaces caractéristiques de l'endomorphisme, c'est-à-dire si et seulement s'il existe une base de l'espace formée de vecteurs propres généralisés de l'endomorphisme. Cette caractérisation, déjà donnée dans le tableau récapitulatif en fin de section précédente, rejoint celle basée sur le polynôme caractéristique, qui doit être scindé pour que l'endomorphisme soit trigonalisable.

1.7.2 Décomposition de Jordan–Chevalley

Théorème 1.48 (« décomposition de Jordan–Chevalley », forme faible) *Soit u un endomorphisme dont le polynôme caractéristique est scindé sur $\mathbb{K}$. Alors, il existe un couple (d, n) d'endomorphismes de E, avec d diagonalisable et n nilpotent, tels que $u = d + n$, $n \circ d = d \circ n$.*

DÉMONSTRATION. Comme le polynôme caractéristique, et donc le polynôme minimal, de u sont scindés, on peut appliquer la proposition précédente qui donne la décomposition

$$E = \bigoplus_{i=1}^{p} F_{\lambda_i}.$$

Soit d l'endomorphisme qui agit comme l'homothétie de facteur λ_i sur chaque espace caractéristique F_{λ_i}, et soit $n = u - d$. Notons que d est diagonalisable (en prenant par exemple la base $\mathscr{B}$ de la preuve précédente). De plus, d et n commutent. Pour vérifier cela, il suffit de vérifier la propriété $d \circ n(x) = n \circ d(x)$ pour $x \in F_{\lambda_i}$ pour un certain i (pourquoi ?). Or, pour un tel x, $n(x) \in F_{\lambda_i}$ puisque $n = d - u$ et que d et u laissent stables F_{λ_i}. On a donc $d(n(x)) = \lambda_i n(x)$ et $n(d(x)) = n(\lambda_i x) = \lambda_i n(x)$. Ainsi, d et n commutent.

Pour conclure, montrons que l'endomorphisme n est nilpotent. Encore une fois, il suffit de montrer que $n^k(x) = 0_E$ pour k assez grand lorsque $x \in F_{\lambda_i}$. Mais $n^k(x) = (u - \lambda_i \,\mathrm{id}_E)^k(x)$ lorsque $x \in F_{\lambda_i}$ (par définition de d et de n). Ainsi, comme $F_{\lambda_i} = \ker((u - \lambda_i \,\mathrm{id}_E)^{l_{\lambda_i}})$, en prenant k égal au maximum des multiplicité l_{λ_i}, on obtient que $n^k(x) = 0_E$ pour $x \in F_{\lambda_i}$, et ceci pour tout i. Ceci conclut la preuve. $\square$

Remarque 1.49 *Il existe une version plus forte de ce théorème, donnant l'unicité de la décomposition ainsi que le fait que d et n s'écrivent comme des polynômes en u.*

Exemple 1.50 *Soit $M \in M_{n,n}(\mathbb{C})$ une matrice dont toutes les valeurs propres sont de modules strictement plus petits que 1. Alors M^k tend vers 0 lorsque k tend vers l'infini. En effet, par la décomposition de Jordan-Chevalley, on peut écrire $M = D + N$ avec D diagonalisable, N nilpotente et $DN = ND$. On a alors :*

$$\begin{aligned} M^2 &= (D+N)(D+N) = D^2 + ND + DN + N^2 = D^2 + 2DN + N^2 \\ M^3 &= M^2(D+N) = D^3 + 2DND + N^2D + D^2N + 2DN^2 + N^3 = D^3 + 3D^2N + 3DN^2 + N^3 \end{aligned}$$

en utilisant que $DN = ND$. De manière plus générale, on trouve

$$M^k = \sum_{l=0}^{k} \binom{k}{l} D^{k-l} N^l.$$

Comme N est nilpotente, on a $N^n = 0$. Ainsi, lorsque $k \geq n$, on peut arrêter la somme précédente à n :

$$M^k = \sum_{l=0}^{n} \binom{k}{l} D^{k-l} N^l.$$

Pour montrer que M^k tend vers 0, il suffit de montrer que chaque terme $\binom{k}{l} D^{k-l} N^l$ tend vers 0 (c'est maintenant une somme de n termes avec n fixé). Pour conclure, on peut utiliser la borne $\binom{k}{l} \leq \frac{k^l}{l!}$ ainsi que le fait que $k^l \lambda^k$ tende vers 0 pour tout $\lambda \in \mathbb{C}$ avec $|\lambda| < 1$.

On est donc arrivé à la situation suivante : pour toute matrice $M \in M_{n,n}(\mathbb{C})$, on sait trouver une matrice semblable de la forme

$$\begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_p \end{pmatrix},$$

où $M_i = \lambda_i I_{m_{\lambda_i}} + N_i$ est une matrice d'ordre m_{λ_i}, qui s'écrit comme la somme d'une matrice diagonale et d'une matrice nilpotente N_i. Pour avoir une réduction encore plus précise de M, la dernière étape consiste à comprendre comment réduire les matrices nilpotentes. Soit J_k la matrice d'ordre k de la forme

$$J_k = \begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}.$$

On peut alors montrer que toute matrice nilpotente N est semblable à une matrice de la forme

$$\begin{pmatrix} J_{k_1} & & & \\ & J_{k_2} & & \\ & & \ddots & \\ & & & J_{k_L} \end{pmatrix}.$$

Ce résultat admis, on obtient alors la **décomposition de Jordan** d'une matrice. Soit $J_k(\lambda) = J_k + \lambda I_k$. Alors, toute matrice $M \in M_{n,n}(\mathbb{C})$ est semblable à une matrice de la forme

$$\begin{pmatrix} J_{k_1}(\lambda_1) & & & \\ & J_{k_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{k_L}(\lambda_L) \end{pmatrix},$$

où les $\lambda_1, \ldots, \lambda_L$ sont les valeurs propres de M, éventuellement répétées.

1.8 Applications (hors programme)

1.8.1 Modèles matriciels d'évolution de population

Un modèle matriciel de population opère en temps discret, en décrivant l'évolution d'une population d'un temps k à un temps $k+1$ (en années par exemple). Pour cela, on découpe une population dont on souhaite étudier l'évolution en sous-catégories (d'âge ou de taille par exemple). A chaque étape, la population des différentes catégories évoluent. Par exemple, si l'on parle de catégories d'âge, certains individus changent de catégories en devenant plus âgés, des nouveaux jeunes apparaissent par procréation tandis que d'autres individus meurent.

Considérons un exemple simple avec trois catégories d'âge : les jeunes (0 à 1 an), les jeunes adultes (1 à 2 ans) et les adultes. La population à l'an k de l'étude est ainsi décrit par un vecteur $X_k \in M_{3,1}(\mathbb{R})$, avec

$$X_k = \begin{pmatrix} j_k \\ s_k \\ a_k \end{pmatrix},$$

où j_k est la population de jeunes, s_k est la population de jeunes adultes et a_k est la population d'adultes. En supposant que seuls les adultes peuvent se reproduire, la population à l'année X_{k+1} est donnée par une relation du type

$$X_{k+1} = \begin{pmatrix} j_{k+1} \\ s_{k+1} \\ a_{k+1} \end{pmatrix} == \begin{pmatrix} 0 & 0 & \alpha \\ \beta & 0 & 0 \\ 0 & \gamma & \delta \end{pmatrix} \begin{pmatrix} j_k \\ s_k \\ a_k \end{pmatrix} = \begin{pmatrix} 0 & 0 & \alpha \\ \beta & 0 & 0 \\ 0 & \gamma & \delta \end{pmatrix} X_k,$$

où $M = \begin{pmatrix} 0 & 0 & \alpha \\ \beta & 0 & 0 \\ 0 & \gamma & \delta \end{pmatrix}$ est appelé la matrice de projection de population. Analysons le rôle de chaque paramètre :

— Le paramètre $\alpha \in [0,1]$ représente la proportion d'adultes qui enfantent à l'année k.
— Le paramètre $\beta \in [0,1]$ représente la proportion de jeunes qui survivent durant leur première année pour devenir de jeunes adultes.
— Le paramètre $\gamma \in [0,1]$ représente la proportion de jeunes adultes qui survivent durant leur première année pour devenir des adultes.
— Le paramètre $\delta \in [0,1]$ représente la proportion d'adultes qui survivent pendant l'année k.

Le rôle des biologistes est alors d'estimer les différents paramètres $\alpha, \beta, \gamma, \delta$ à partir d'observations faites sur le terrain. Par exemple, des biologistes voulant étudier le rôle de la déforestation sur l'évolution de la population de chouettes tachetées dans les forêts du nord-ouest Pacifique en Amérique du Nord ont estimé que pour cette population $\alpha = 0,33$, $\beta = 0,18$, $\gamma = 0,71$ et $\delta = 0,94$. Que peut-on en déduire sur l'évolution à long terme de la population de la chouette tachetée ?

Théorème 1.51 *Soit $M \in M_{n,n}(\mathbb{K})$ une matrice et $X_0 \in M_{n,1}(\mathbb{K})$ un vecteur.*

1. *La suite définie par $X_{k+1} = MX_k$ pour $k \geq 0$ satisfait $X_k = M^k X$ pour tout $k \geq 0$.*
2. *Supposons que M que toutes les valeurs propres (complexes) de M sont de modules strictement inférieurs à* 1. *Alors, le vecteur X_k tend aussi vers* 0 *lorsque k tend vers l'infini.*

Démonstration. Le premier point se démontre par récurrence. Pour le second, on utilise l'exemple 1.50. □

Un calcul (sur ordinateur) montre que les valeurs propres de M pour l'exemple des chouettes tachetées sont égales à $\lambda_1 \simeq 0,98$, $\lambda_2 \simeq -0,02 + 0,21i$ et $\lambda_3 \simeq -0,02 - 0,21i$. Toutes les valeurs propres sont ici de modules strictement inférieures à 1. D'après le théorème 1.51, on en conclut que le vecteur X_k tend vers 0. On en conclut que si les estimations des biologistes sont exactes et que si la modélisation proposée n'est pas trop éloignée de la réalité, cette population de chouettes tachetées va disparaître. **Quelles critiques peut-on faire à ce modèle ?**

1.8.2 Chaînes de Markov

Une chaîne de Markov est un modèle permettant de décrire un système dont l'état évolue aléatoirement au cours du temps. Le jeu de l'oie est un exemple très simple de chaîne de Markov : l'état d'un joueur est décrit par un nombre i entre 1 et n, où n est le nombre de cases du jeu de l'oie. Le joueur étant à la $i$ème case, il se retrouvera à la fin de son tour à différentes cases selon le résultat de son lancer de dé. Dans le cas le plus simple, le joueur se trouvera à l'étape suivante sur une des cases $i+1, \dots, i+6$, chacune avec chance $1/6$. Des cases peuvent aussi contenir des règles particulières (reculer de 4 cases, retourner à la case départ, relancer le dé), pouvant rendre les probabilités de se retrouver à une certaine case en fin de tour en partant de i plus compliqué à calculer.

Le jeu de l'oie est particulièrement simple dans le sens où la probabilité de se retrouver à la case j en fin de tour en partant de la case i dépend **uniquement** de i et j : notons-la Π_{ji}. On parle d'un système sans mémoire : seul le présent (et non les événements passés) influent sur les probabilités futures. On appelle **matrice de transition** de la chaîne de Markov la matrice Π de taille n par n formée avec les nombres Π_{ji}. Pour chaque colonne i, la somme des entrées correspondantes $\sum_{j=1}^{n} \Pi_{ji}$ est égale à 1 : c'est la somme des probabilités de tous les événments possibles pouvant arriver en partant de la case i. On appelle une matrice vérifiant une telle propriété une matrice stochastique.

Définition 1.52 *Un vecteur $X \in M_{n,1}(\mathbb{R})$ avec des entrées positives qui se somment à 1 est appelé un* ***vecteur de probabilités****. Une matrice $\Pi \in M_{n,n}(\mathbb{R})$ dont les colonnes sont des vecteurs de probabilités est appelé une* ***matrice stochastique****. Une* ***chaîne de Markov*** *est une suite de vecteurs de probabilités $X_0, X_1, X_2, \dots$ avec une matrice stochastique Π (appelée matrice de transition de la chaîne de Markov) telle que :*

$$X_1 = \Pi X_0, \quad X_2 = \Pi X_1, \quad X_3 = \Pi X_2, \quad \dots$$

Le vecteur de probabilités X_k représente les probabilités de se retrouver dans les différents états du système (représentées par les indices $1, \dots, n$) après k pas de la chaîne de Markov, en partant d'un état aléatoire donné par le vecteur de probabilités X_0. Par exemple, pour le jeu de l'oie, en prenant

$$X_0 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

le vecteur X_k représente les probabilités de se retrouver sur les différentes cases en partant de la case départ (numéro 1) après k tours de jeu.

Donnons un autre exemple. En 2005, l'économiste américain James D. Hamilton proposa de modéliser le taux de chômage selon des équations dépendant de l'« état » de l'économie américaine, décrite par trois valeurs : croissance normale, faible récession, forte récession. L'évolution d'un état à l'autre de l'économie américaine est elle-même modélisée à l'aide d'une chaîne de Markov avec matrice de transition

$$\Pi = \begin{pmatrix} 0.971 & 0.145 & 0 \\ 0.029 & 0.778 & 0.508 \\ 0 & 0.077 & 0.492 \end{pmatrix}, \tag{$\star$}$$

où l'échelle de temps entre k et $k+1$ représente un mois. Une question naturelle est alors le comportement en temps long du système : partant d'un vecteur de probabilité quelconque X_0, à quoi ressemblera le vecteur de probabilité X_k pour k grand ? Converge-t-il vers un certain vecteur ? Ce vecteur dépend-il de la condition initiale ?

On peut tout d'abord remarquer que si un vecteur X vérifie $\Pi X = X$, alors la chaîne de Markov partant de $X_0 = X$ est constante. On appelle un tel vecteur un **vecteur de probabilité stationnaire**.

Théorème 1.53 *Si Π est une matrice stochastique, alors 1 est valeur propre de Π. En particulier, toute chaîne de Markov admet au moins un vecteur de probabilité stationnaire.*

Démonstration. D'après la propriété 1.21, la matrice Π a les mêmes valeurs propres que la matrice transposée $\Pi^\top$. Soit $e \in M_{n,1}(\mathbb{R})$ le vecteur avec 1 à chacune de ses entrées. Alors, pour $i = 1, \dots, n$,

$$(\Pi^\top e)_i = \sum_{j=1}^{n} \Pi_{ij}^\top e_j = \sum_{j=1}^{n} \Pi_{ji} = 1$$

par définition d'une matrice stochastique. Ainsi, 1 est valeur propre de $\Pi^\top$, et donc valeur propre de Π. $\square$

Supposons qu'il existe une unique vecteur de probabilité stationnaire X_* et que toutes les valeurs propres de Π différentes de 1 sont de modules strictement inférieurs à 1. Alors, en complétant le vecteur X_* en une base $\mathscr{B}$ de vecteurs propres de Π, on peut observer que la matrice Π est semblable (via une matrice de passage P) à une matrice de la forme

$$\begin{pmatrix} 1 & 0 \\ 0 & \Pi' \end{pmatrix},$$

où toutes les valeurs propres de Π' sont de modules strictement plus petits que 1. Par un calcul matriciel par blocs, on peut voir que $\Pi^k = P\begin{pmatrix}1 & 0\\ 0 & \Pi'^k\end{pmatrix}P^{-1}$. D'après l'exemple 1.50, la suite de matrice Π'^k tend vers 0. Ainsi, la suite Π^k tend vers la matrice

$$\Pi_\infty = P\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}P^{-1}.$$

Réflechissons à comment la matrice Π_∞ agit en terme d'endomorphismes : la matrice Π_∞ agit dans la base $\mathscr{B} = (X_*, e_2, \dots, e_n)$ sur un vecteur X_0 s'écrivant de manière unique $X_0 = \lambda_1 X_* + \sum_{i=2}^n \lambda_i e_i$ via $\Pi_\infty X_0 = \lambda_1 X_*$. Lorsque X_0 est un vecteur de probabilité, on obtient que $X_k = \Pi^k X_0$ converge vers $\lambda_1 X_*$ (là encore, ce résultat est une conséquence de la continuité de la multiplication matricielle). Que peut valoir le scalaire λ_1 ? On peut noter que pour tout k, la somme des entrées de X_k est égale à 1. Par continuité de la somme, la somme des entrées de X_∞ doit aussi être égale à 1. Or, cette somme est aussi égale à $\sum_{i=1}^n \lambda_1 (X_*)_i = \lambda_1$. Ainsi, $\lambda_1 = 1$. On vient de montrer le théorème suivant.

Théorème 1.54 (Convergence des chaînes de Markov) *Soit une chaîne de Markov donné par une matrice de transition* Π. *Supposons que la chaîne de Markov admet un unique vecteur de probabilité stationnaire* X_* *et que toutes les valeurs propres de* Π *différentes de* 1 *sont de modules strictement inférieurs à 1. Alors, pour tout vecteur de probabilité initiale* X_0*, la chaîne de Markov* $(X_k)_{k\geq 0}$ *converge vers* X_* *lorsque* k *tend vers l'infini.*

Le résultat précédent indique que, sous certaines conditions, peu importe la condition initiale, une chaîne de Markov convergera un temps long vers un comportement typique. Dans l'exemple donné par l'équation ($\star$), on trouve par un calcul (sur ordinateur) que les valeurs propres de Π sont $\lambda_1 = 1$, $\lambda_2 \simeq 0.39$ et $\lambda_3 \simeq 0.85$. Nous sommes donc dans le cadre du théorème précédent : il existe forcément un unique vecteur de probabilité stationnaire X_* puisque l'espace propre de la matrice associé à la valeur propre 1 est nécessairement de dimension 1 (il y a trois valeurs propres distinctes). On résoud l'équation $\Pi X = X$ pour trouver le vecteur propre associé

$$\begin{pmatrix}\frac{2540}{77}\\ \frac{508}{77}\\ 1\end{pmatrix}.$$

Attention : ce n'est pas le vecteur de probabilité stationnaire car la somme de ses entrées ne fait pas 1. Pour trouver X_*, il suffit de normaliser le vecteur précédent par la somme $\frac{2540}{77} + \frac{508}{77} + 1$ pour obtenir

$$X_* \simeq \begin{pmatrix}0,813\\ 0,162\\ 0,025\end{pmatrix}.$$

Ainsi, d'après cette modélisation, l'économie américaine passe environ 81,3% de son temps en croissance normale, 16,2% de son temps en récession faible et 2,5% de son temps en récession forte. **Quelles critiques peut-on faire à ce modèle ?**

Nous verrons en TD un ensemble de conditions qui permettent d'assurer qu'une chaîne de Markov satisfait les conditions du théorème 1.54. Cet ensemble de conditions est donné par le **théorème de Perron-Frobenius**.

Chapitre 2

Formes bilinéaires et quadratiques

On aborde dans ce chapitre les notion de *applications bilinéaires* et de *formes quadratiques*. En science des données, celles-ci interviennent dans deux exemples principaux : pour étudier des corrélations entre différentes variables en utilisant des matrices de covariances, et pour des problèmes d'optimisation à travers la matrice Hessienne. Nous verrons aussi à travers le cours l'utilisation de formes bilinéaires sur des algorithmes précis : l'analyse en composantes principales (ou PCA, pour *Principal Component Analysis*) et le partitionnement spectral (ou *spectral clustering*).

2.1 Généralités sur les applications bilinéaires

Une *application bilinéaire* est l'analogue à deux variables d'une application linéaire. Étant donné trois $\mathbb{K}$-espaces vectoriels E, F et G, une application b de $E \times F$ dans G est dite *bilinéaire* si et seulement si

$$\forall (x,x') \in E^2,\ \forall (y,y') \in F^2,\ \forall \lambda \in \mathbb{K},$$
$$b(x+x',y) = b(x,y) + b(x',y),\ b(x,y+y') = b(x,y) + b(x,y'),\ b(\lambda x,y) = b(x,\lambda y) = \lambda\, b(x,y).$$

On peut exprimer cette définition de manière un peu plus formelle en disant que l'application b est *linéaire par rapport à chacune de ses variables*. Ceci peut être rendu précis de la façon suivante.

Définition 2.1 *Soit E, F et G trois $\mathbb{K}$-espaces vectoriels. Une application b de $E \times F$ dans G est dite* ***bilinéaire*** *si et seulement si l'****application partielle à gauche de*** *b, notée $L(b)$, de E dans $\mathscr{L}(F;G)$ définie par*

$$\forall x \in E,\ L(b)(x) = (y \mapsto b(x,y)),$$

*et l'****application partielle à droite de*** *b, notée $R(b)$, de F dans $\mathscr{L}(E;G)$ définie par*

$$\forall y \in F,\ R(b)(y) = (x \mapsto b(x,y))$$

sont toutes deux linéaires.

Lorsque $G = \mathbb{K}$, on parle de *forme bilinéaire*.

L'ensemble des applications bilinéaires de $E \times F$ dans G, noté $\mathscr{L}(E,F;G)$, est un espace vectoriel sur $\mathbb{K}$. Il est en effet clair que la somme de deux applications bilinéaires est une application bilinéaire et que le produit d'une application bilinéaire par un scalaire est une application bilinéaire.

Définition 2.2 (application bilinéaire symétrique) *Soient E et G deux $\mathbb{K}$-espaces vectoriels et b une application bilinéaire de $E \times E$ dans G. On dit que l'application b est* ***symétrique*** *(resp.* ***antisymétrique****) si et seulement si*

$$\forall (x,y) \in E^2,\ b(x,y) = b(y,x) \text{ (resp. } b(x,y) = -b(y,x)\text{)}.$$

Exemple 2.3 — *Presque tout ce qui porte le nom de* produit *est bilinéaire. Ainsi, le produit dans $\mathbb{R}$,*

$$\begin{array}{ccc} \mathbb{R} \times \mathbb{R} & \to & \mathbb{R} \\ (x,y) & \mapsto & xy, \end{array}$$

est bilinéaire symétrique. Ceci résulte de l'associativité et de la commutativité de la multiplication et de la distributivité de la multiplication par rapport à l'addition.

— *Soit m, n et p des entiers naturels non nuls. Le produit matriciel de $M_{m,n}(\mathbb{R}) \times M_{n,p}(\mathbb{R})$ dans $M_{m,p}(\mathbb{R})$ est une application bilinéaire.*

— *Sur $\mathbb{R}^n$, le produit scalaire usuel,*

$$\begin{array}{ccc} \mathbb{R}^n \times \mathbb{R}^n & \to & \mathbb{R} \\ (x, y) & \mapsto & x \cdot y, \end{array}$$

défini par

$$x \cdot y = \sum_{i=1}^{n} x_i y_i,$$

où $x = (x_1, \ldots, x_n)$ et $y = (y_1, \ldots, y_n)$, est une forme bilinéaire symétrique.

2.2 Représentation matricielle d'une forme bilinéaire

Nous nous limitons à partir de maintenant aux *formes* bilinéaires, c'est-à-dire qu'on aura $G = \mathbb{K}$ dans la suite.

Définition 2.4 *Soient E et F deux $\mathbb{K}$-espaces vectoriels et $b : E \times F \to \mathbb{K}$ une forme bilinéaire. Soient $\mathscr{B} = \{e_1, \ldots, e_n\}$ une base de E et $\mathscr{C} = \{f_1, \ldots, f_m\}$ une base de F. On définit la matrice $M \in M_{n,m}(\mathbb{K})$ représentative de b dans les bases $\mathscr{B}$ et $\mathscr{C}$ par*

$$\forall i \in \{1, \ldots, n\}, \forall j \in \{1, \ldots, m\},\ M_{ij} = b(e_i, f_j).$$

On note $M = [b]_{\mathscr{B},\mathscr{C}}$.

Lorsque $E = F$ et que $\mathscr{B} = \mathscr{C}$, on note $[b]_{\mathscr{B}} = [b]_{\mathscr{B},\mathscr{B}}$.

Soient $x \in E$ et $y \in F$. Soient X la matrice colonne de $M_{n,1}(\mathbb{K})$ et Y la matrice colonne de $M_{m,1}(\mathbb{K})$ qui contiennent les coordonnées respectives des vecteurs x de E dans la base $\mathscr{B}$ et y de F dans la base $\mathscr{C}$. On a alors par les propriétés de bilinéarités que

$$b(x, y) = b(\sum_{i=1}^{n} x_i e_i, \sum_{j=1}^{m} y_j f_j) = \sum_{i=1}^{n} x_i b(e_i, \sum_{j=1}^{m} y_j f_j) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i y_j b(e_i, f_j).$$

Si on pose $M = [u]_{\mathscr{B},\mathscr{C}}$, on obtient

$$b(x, y) = X^\top M Y.$$

Proposition 2.5 *Soit E un espace vectoriel de dimension finie non nulle, $\mathscr{B}$ une base de E et b une forme bilinaire sur $E \times E$. La matrice de b relativement à la base $\mathscr{B}$ est symétrique (resp. antisymétrique) si et seulement si la forme b est symétrique (resp. antisymétrique).*

Changement de bases

Considérons une forme bilinéaire sur $E \times F$ et les bases $\mathscr{B}$ et $\mathscr{B}'$ de E, $\mathscr{C}$ et $\mathscr{C}'$ de F. On note P la matrice de passage de $\mathscr{B}$ à $\mathscr{B}'$, c'est-à-dire qu'on a $X = PX'$ avec $X = \mathrm{mat}_{\mathscr{B}}(x)$ et $X' = \mathrm{mat}_{\mathscr{B}'}(x)$ pour tout vecteur x de E, et Q la matrice de passage de $\mathscr{C}$ à $\mathscr{C}'$, c'est-à-dire qu'on a $Y = QY'$ avec $Y = \mathrm{mat}_{\mathscr{C}}(y)$ et $Y' = \mathrm{mat}_{\mathscr{C}'}(y)$ pour tout vecteur y de F. Comparons la matrice de b relativement aux bases $\mathscr{B}$ et $\mathscr{C}$ d'une part avec celle relativement aux bases $\mathscr{B}'$ et $\mathscr{C}'$ d'autre part. On a

$$(X')^\top M' Y' = b(x, y) = X^\top M Y = (PX')^\top M(QY') = (X')^\top P^\top M Q Y',$$

d'où, par identification,

$$M' = P^\top M Q. \tag{2.1}$$

Remarque 2.6 *On notera que la matrice transposée $P^\top$ est généralement différente de l'inverse P^{-1}. Par conséquent, même lorsque $F = E$ et $\mathscr{C} = \mathscr{B}$ (et donc $Q = P$), les matrices carrées M et M' ne sont généralement pas semblables (on dit qu'elles sont* congrues*). Ceci signifie en particulier que si la matrice d'un endomorphisme de E dans une base donnée est la même que celle d'une forme bilinéaire sur $E \times E$ relativement à cette base, il n'en sera pas nécessairement de même après changement de base.*

2.3 Non-dégénérescence d'une forme bilinéaire symétrique

Définition 2.7 (Noyau d'une forme bilinéaire symétrique) *Soit E un $\mathbb{K}$-espace vectoriel. Soit b une forme bilinéaire symétrique sur $E \times E$. On définit le noyau de b*

$$\ker(b) = \{x \in E : \ \forall y \in E, b(x,y) = 0\}.$$

On dit que b est non-dégénérée si $\ker(b) = \{0_E\}$.

Proposition 2.8 *Supposons que E est de dimension fini $n \geq 1$, et soit $\mathscr{B}$ une base de E. Soit $M = [b]_{\mathscr{B}}$ la matrice représentative de b dans la base $\mathscr{B}$. Soit $x \in E$ et soit $X \in M_{n,1}(\mathbb{K})$ la matrice colonne représentant x dans la base $\mathscr{B}$. Alors, $x \in \ker(b)$ si et seulement si $X \in \ker(M)$. En particulier, la non-dégénérescence de b est équivalent à l'inversibilité de la matrice M.*

DÉMONSTRATION. La condition $\forall y \in E,\ b(x,y) = 0$ est équivalente à la condition $\forall Y \in M_{n,1}(\mathbb{K}),\ X^\top MY = 0$. Cette condition est équivalente à l'égalité $X^\top M = 0$ (en choisissant Y parcourant la base canonique de $M_{n,1}(\mathbb{K})$). Or, par symétrie de M, on a $MX = M^\top X = (X^\top M)^\top = 0$. Ainsi, $x \in \ker(b)$ si et seulement si $X \in \ker(M)$. □

Remarque 2.9 *La restriction à un sous-espace vectoriel d'une forme bilinéaire non-dégénérée peut être dégénérée.*

Exemple 2.10 *La forme définie sur $\mathbb{R}^2 \times \mathbb{R}^2$ par*

$$b(x,y) = x_1y_1 - x_2y_2,$$

est représentée par la matrice $\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ relativement à la base canonique. Cette matrice est inversible : la forme b est donc non-dégénérée. Pourtant, sa restriction à la droite vectorielle engendrée par le vecteur $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ est dégénérée, car identiquement nulle.

Définition 2.11 (Rang d'une application bilinéaire) *Soit E un espace vectoriel de dimension finie et b une forme bilinéaire sur $E \times E$. Soient $\mathscr{B}$ et $\mathscr{C}$ deux bases de E. Le **rang** de la forme b est le rang de la matrice $[b]_{\mathscr{B},\mathscr{C}}$.*

Pour que cette définition soit valide, nous devons vérifier que le rang de la matrice $[b]_{\mathscr{B},\mathscr{C}}$ ne dépend pas du choix des bases $\mathscr{B}$ et $\mathscr{C}$. Cela est clair avec la formule (2.1), puisque le rang d'une matrice M est le même que celui de la matrice $P^\top MQ$ dès lors que les matrices P et Q sont inversibles.

2.4 Généralités sur les formes quadratiques

Définition 2.12 (Forme quadratique) *Soit E un $\mathbb{K}$-espace vectoriel. Une application q de E dans $\mathbb{K}$ est appelée une **forme quadratique** sur E s'il existe une forme bilinéaire b de $E \times E$ dans $\mathbb{K}$ telle que*

$$\forall x \in E,\ q(x) = b(x,x).$$

Toute forme bilinéaire b sur $E \times E$ donne naissance à une forme quadratique sur E en posant

$$\forall x \in E,\ q(x) = b(x,x).$$

On appelle alors q la *forme quadratique associée à b*. Une même forme quadratique peut ainsi être associée à plusieurs formes bilinéaires. On a toutefois le résultat suivant.

Proposition et définition 2.13 (Forme polaire d'une forme quadratique) *Sur un $\mathbb{K}$-espace vectoriel (avec $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$), toute forme quadratique est associée à une unique forme bilinéaire symétrique, cette dernière étant appelée la **forme polaire** de la forme quadratique.*

DÉMONSTRATION. Soit q une forme quadratique sur un $\mathbb{K}$-espace vectoriel E. Par définition, il existe une forme bilinéaire b, non nécessairement symétrique, telle que

$$\forall x \in E,\ b(x,x) = q(x).$$

Symétrisons la forme b en introduisant la forme bilinéaire c définie par

$$\forall (x,y)x \in E^2,\ c(x,y) = \frac{1}{2}\left(b(x,y) + b(y,x)\right).$$

On a alors

$$\forall x \in E,\ c(x,x) = b(x,x),$$

et q est donc aussi la forme quadratique associée à c. Par ailleurs, la forme c étant symétrique, on a

$$\begin{aligned}\forall (x,y) \in E^2,\ q(x+y) - q(x-y) &= c(x+y,x+y) - c(x-y,x-y)\\ &= c(x,x) + 2\,c(x,y) + c(y,y) - c(x,x) + 2\,c(x,y) - c(y,y)\\ &= 4\,c(x,y),\end{aligned}$$

ce qui donne

$$\forall (x,y) \in E^2,\ c(x,y) = \frac{1}{4}\left(q(x+y) - q(x-y)\right). \tag{2.2}$$

La forme c est donc définie de manière unique. □

La formule (2.2) obtenue dans la démonstration ci-dessus est appelée *identité de polarisation*. On notera que l'on a également

$$\begin{aligned}\forall (x,y) \in E^2,\ q(x+y) - q(x) - q(y) &= c(x+y,x+y) - c(x,x) - c(y,y)\\ &= c(x,x) + 2\,c(x,y) + c(y,y) - c(x,x) - c(y,y)\\ &= 2\,c(x,y),\end{aligned}$$

d'où la formule

$$\forall (x,y) \in E^2,\ c(x,y) = \frac{1}{2}\left(q(x+y) - q(x) - q(y)\right),$$

qui constitue une deuxième identité de polarisation. Une troisième identité de polarisation est donnée par

$$\forall (x,y) \in E^2,\ c(x,y) = \frac{1}{2}\left(q(x) + q(y) - q(x-y)\right).$$

En pratique, si l'on exhibe une forme bilinéaire symétrique ayant une forme quadratique q associée, c'est nécessairement la forme polaire de q. On n'a donc pas toujours besoin d'utiliser les identités de polarisation.

Un *polynôme homogène de degré deux* sur $\mathbb{K}$ en les variables $x_1, \dots, x_n$ est une expression de la forme

$$\sum_{i=1}^{n} \alpha_{ii}\, {x_i}^2 + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \alpha_{ij}\, x_i x_j,$$

où les coefficients α_{ij}, $1 \leq i \leq j \leq n$, appartiennent à $\mathbb{K}$.

Proposition 2.14 *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie non nulle et $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Les formes quadratiques sur E sont les applications q telles que, pour tout vecteur x de E, $q(x)$ est un polynôme homogène de degré deux en les coordonnées de x dans la base $\mathcal{B}$.*

DÉMONSTRATION. Soit b la forme polaire de q. Soit $M = [b]_{\mathcal{B}}$ et soit $x \in E$, avec $X \in M_{n,1}(\mathbb{K})$ le vecteur colonne représentant x dans la base $\mathcal{B}$. Alors,

$$\begin{aligned}q(x) = b(x,x) = X^\top M X &= \sum_{i=1}^{n}\sum_{j=1}^{n} M_{ij}\, x_i x_j\\ &= \sum_{i=1}^{n} M_{ii}\, {x_i}^2 + 2 \sum_{i=1}^{n} \sum_{j=i+1}^{n} M_{ij}\, x_i x_j,\end{aligned}$$

car on $M_{ij} = M_{ji}$. □

Remarque 2.15 *Une forme quadratique étant donnée par un polynôme homogène de degré deux, sa forme polaire peut être déterminée en* polarisant *chaque monôme de ce polynôme. Ainsi, un monôme de la forme* $a\,x_i{}^2$ *est polarisé en* $a\,x_i y_i$*, tandis qu'un monôme de la forme* $a\,x_i x_j$ *est polarisé en* $\frac{a}{2}\left(x_i y_j + x_j y_i\right)$.

Définitions 2.16 (Matrice, noyau et rang d'une forme quadratique) *Soit* q *une forme quadratique sur un* $\mathbb{K}$*-espace vectoriel, supposé de dimension finie. La matrice de* q *est celle de sa forme polaire. Le noyau et le rang de* q *sont aussi définis comme étant ceux de sa forme polaire. On note* $[q]_{\mathscr{B}}$ *la matrice représentative de* q *dans une base* $\mathscr{B}$ *de* E.

Il découle de ces définitions que l'on peut exprimer matriciellement cette dernière en utilisant ce qui a été vu dans le précédent chapitre. En notant $M = [q]_{\mathscr{B}}$ et en notant X la matrice colonne représentant un élément x de E dans la base $\mathscr{B}$, on a

$$q(x) = X^\top M X.$$

Une autre conséquence de cette définition est que l'espace vectoriel des formes quadratiques sur E, noté $\mathscr{Q}(E)$, est isomorphe à $S_n(\mathbb{K})$, l'ensemble des matrices symétriques à coefficients dans $\mathbb{K}$. En particulier, si l'espace E est de dimension n, alors l'espace $\mathscr{Q}(E)$ est de dimension $\frac{1}{2}n(n+1)$.

Définition 2.17 (Forme quadratique non dégénérée) *Une forme quadratique est dite* ***non dégénérée*** *si et seulement si son noyau est réduit au vecteur nul.*

Définition 2.18 (Formes quadratiques équivalentes) *Deux formes quadratiques* q *et* q' *sur un* $\mathbb{K}$*-espace vectoriel* E *sont dites* ***équivalentes*** *s'il existe un endomorphisme inversible* u *de* E *tel que*

$$\forall x \in E,\ q'(x) = q(u(x)).$$

Proposition 2.19 *Soient* E *un* $\mathbb{K}$*-espace vectoriel et* $\mathscr{B}$ *une base de* E*. Deux formes quadratiques* q *et* q' *sur un* $\mathbb{K}$*-espace vectoriel* E *sont équivalentes si et seulement si les matrices* $M = [q]_{\mathscr{B}}$ *et* $M' = [q']_{\mathscr{B}}$ *sont congrues, c'est à dire si il existe une matrice* $P \in M_{n,n}(\mathbb{K})$ *inversible avec* $M' = P^\top M P$.

DÉMONSTRATION. Montrons l'implication directe. Soit $P = [u]_{\mathscr{B}}$. Soient $x \in E$ et $X \in M_{n,1}(\mathbb{K})$ le vecteur colonne représentant x dans la base $\mathscr{B}$. Alors, le vecteur colonne représentant $u(x)$ dans la base $\mathscr{B}$ est égal à PX. Ainsi, on a $q(u(x)) = (PX)^\top M(PX) = X^\top(P^\top MP)X$. Ceci est égal à $q'(x) = X^\top M'X$. Comme cette égalité est valable pour tout $X \in M_{n,1}(\mathbb{K})$, on a $M' = P^\top MP$. □

2.5 Orthogonalité

Dans toute cette section, E désigne un $\mathbb{K}$-espace vectoriel, q une forme quadratique sur E et b la forme polaire de q.

Définition 2.20 (Cône isotrope) *Un vecteur* x *de* E *est dit* ***isotrope*** *pour* q *si et seulement si* $q(x) = 0$*. On appelle* ***cône isotrope*** *de* q *l'ensemble* $\mathscr{C}_q$ *des vecteurs isotropes pour* q*. On dit enfin que* q *est* ***définie*** *si et seulement si* $\mathscr{C}_q = \{0_E\}$.

Une forme bilinéaire b sur $E \times E$ est dite *alternée* si et seulement si tout vecteur de E est isotrope pour sa forme quadratique associée, c'est-à-dire si

$$\forall x \in E,\ b(x,x) = 0.$$

Remarque 2.21 *On a toujours l'inclusion* $\ker(q) \subset \mathscr{C}_q$*, mais le cône isotrope* $\mathscr{C}_q$ *n'est généralement pas un sous-espace vectoriel de* E*. Il en découle qu'une forme quadratique définie est non dégénérée, mais que la réciproque est généralement fausse.*

Exemple 2.22
- *Soient* $E = \mathbb{R}^2$ *et* $q(x) = x_1 x_2$ *pour* $x = (x_1, x_2) \in \mathbb{R}^2$*. Alors, on a* $\ker(q) = \{0_E\}$ *et* $\mathscr{C}_q = \{x_1 = 0\} \cup \{x_2 = 0\}$.
- *Soient* $E = \mathbb{R}^2$ *et* $q(x) = x_1{}^2 - x_2{}^2$ *pour* $x = (x_1, x_2) \in \mathbb{R}^2$*. Alors* q *est non dégénérée, mais pas définie.*

Définition 2.23 (Orthogonalité) *On dit que deux vecteurs x et y de E sont* ***orthogonaux*** *pour la forme q (ou pour sa forme polaire b), ou simplement q-orthogonaux, si et seulement si $b(x,y)=0$, ce que l'on note $x \perp y$.*

Si A est une partie de E, on appelle ***orthogonal*** *de A pour la forme q (ou pour sa forme polaire b), et on note $A^{\perp}$, la partie de E dont les éléments sont orthogonaux à ceux de A,*

$$A^{\perp} = \{y \in E \,|\, \forall x \in A,\ b(x,y)=0\}.$$

Enfin, deux sous-ensembles A et B de E sont dits orthogonaux pour la forme q (ou pour sa forme polaire b) si et seulement si

$$\forall x \in A,\ \forall y \in B,\ b(x,y)=0,$$

ce que l'on note encore $A \perp B$.

Il est immédiat que $A^{\perp}$ est un sous-espace vectoriel de E. Par ailleurs, pour toute partie A de E, l'orthogonal de A est égal à celui du sous-espace engendré par A, c'est-à-dire que

$$A^{\perp} = \mathrm{Vect}(A)^{\perp}.$$

Par conséquent, on considèrera dans la suite uniquement des orthogonaux de parties qui sont des sous-espaces vectoriels de E.

On remarquera par ailleurs qu'il découle des précédentes définitions que

$$\ker(q) = E^{\perp},$$

et que tout vecteur isotrope est orthogonal à lui-même.

Proposition 2.24 *Soit A et B des parties de E. On a*

(i) $A \subset (A^{\perp})^{\perp}$,

(ii) $A \subset B \implies B^{\perp} \subset A^{\perp}$.

DÉMONSTRATION.

(i) Soit x un vecteur de A. On a

$$\forall y \in A^{\perp},\ b(x,y)=0,$$

d'où x est orthogonal à tout vecteur de $A^{\perp}$ et appartient donc à $(A^{\perp})^{\perp}$.

(ii) Soit y un vecteur de $B^{\perp}$. On a

$$\forall x \in B,\ b(x,y)=0.$$

En particulier, cette égalité est vraie pour tout vecteur x de A, puisque A est inclus dans B. Par conséquent, le vecteur y appartient à $A^{\perp}$.

□

2.6 Formes quadratiques positives

Dans cette section, on suppose que E est un $\mathbb{R}$-espace vectoriel.

Définition 2.25 *Une forme quadratique sur E est dite* ***positive*** *(resp.* ***négative****) si et seulement si*

$$\forall x \in E,\ q(x) \geq 0 \text{ (resp. } q(x) \leq 0).$$

On dit qu'elle est ***définie positive*** *(resp.* ***définie négative****) si et seulement si*

$$\forall x \in E,\ x \neq 0_E,\ q(x) > 0 \text{ (resp. } q(x) < 0).$$

Par extension, une matrice symétrique est dite définie positive/positive/définie négative/négative si c'est la matrice d'une forme quadratique définie positive/positive/définie négative/négative sur un $\mathbb{R}$-espace vectoriel de dimension finie.

Attention : une matrice symétrique (définie) positive n'est pas nécessairement une matrice à coefficients (strictement) positifs, comme le montre le contre-exemple suivant :

$$M = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 2 \\ 0 & 2 & 3 \end{pmatrix},\ X = \begin{pmatrix} 2 \\ -2 \\ 1 \end{pmatrix} \text{ on a } X^\top M X = -1.$$

Nous donnons plusieurs inégalités importantes sur les formes quadratiques positives.

Théorème 2.26 (« inégalité de Cauchy–Schwarz ») *Soit q une forme quadratique positive sur E, de forme polaire b. On a*

$$\forall (x,y) \in E^2,\ |b(x,y)| \leq \sqrt{q(x)}\sqrt{q(y)}.$$

Si de plus la forme q est définie, il y a égalité dans l'inégalité si et seulement si les vecteurs x et y sont linéairement dépendants.

DÉMONSTRATION. La forme q étant positive, on a

$$\forall t \in \mathbb{R},\ \forall (x,y) \in E^2,\ q(t\,x + y) = t^2 q(x) + 2t\, b(x,y) + q(y) \geq 0.$$

Si $q(x) = 0$, l'inégalité ci-dessus devient

$$\forall t \in \mathbb{R},\ 2t\, b(x,y) + q(y) \geq 0,$$

ce qui entraîne que $b(x,y) = 0$. Sinon, le trinôme du second degré en t à gauche de l'inégalité possède un discriminant négatif, ce qui s'écrit encore

$$b(x,y)^2 - q(x)q(y) \leq 0,$$

conduisant à l'inégalité annoncée.

Si q est de plus définie, on suppose que le vecteur x est non nul (l'inégalité étant triviale lorsque ce n'est pas le cas). Alors $q(x)$ est non nul de sorte qu'on a égalité si et seulement le discriminant ci-dessus est nul, c'est-à-dire si et seulement s'il existe un réel t_0 tel que $q(t_0\, x + y) = 0$, ce qui équivaut à $t_0\, x + y = 0_E$. □

Corollaire 2.27 *Soit q une forme quadratique positive. Le cône isotrope de q est égal à son noyau.*

DÉMONSTRATION. On a toujours $\ker(q) \subset \mathscr{C}_q$ et il faut donc juste prouver l'inclusion réciproque. Soit x un vecteur appartenant à $\mathscr{C}_q$ et y un vecteur de E. En vertu de l'inégalité de Cauchy–Schwarz, on a

$$0 \leq |b(x,y)| \leq \sqrt{q(x)}\sqrt{q(y)} = 0,$$

ce qui implique que $b(x,y) = 0$, d'où x appartient à $\ker(q)$. □

Ce résultat vaut également pour une forme quadratique négative. Une forme positive (ou négative) est donc non dégénérée si et seulement si elle est définie.

Corollaire 2.28 (inégalité de Minkowski) *Soit q une forme quadratique positive. On a*

$$\forall (x,y) \in E^2,\ \sqrt{q(x+y)} \leq \sqrt{q(x)} + \sqrt{q(y)}.$$

DÉMONSTRATION. C'est une conséquence directe de l'inégalité de Cauchy–Schwarz. On a en effet

$$\forall (x,y) \in E^2,\ q(x+y) = q(x) + 2\, b(x,y) + q(y) \leq q(x) + 2\sqrt{q(x)}\sqrt{q(y)} + q(y) = \left(\sqrt{q(x)} + \sqrt{q(y)}\right)^2.$$

□

Chapitre 3

Espaces euclidiens

La notion fondamentale d'*espace euclidien* généralise de manière naturelle la géométrie « classique », introduite par Euclide dans les *Éléments*. Elle est définie par la donnée d'un espace vectoriel sur le corps des réels, de dimension finie, muni d'un *produit scalaire*, et permet, entre autres choses, de « mesurer » des distances et des angles.

3.1 Définitions

Définition 3.1 (produit scalaire) *Soit E un espace vectoriel sur $\mathbb{R}$. Une application $\langle \cdot, \cdot \rangle$ de $E \times E$ dans $\mathbb{R}$ est appelée un* ***produit scalaire*** *sur E si et seulement si elle est bilinéaire, symétrique, non dégénérée et positive.*

D'après un corollaire de l'inégalité de Cauchy–Schwarz (voir le corollaire 2.27), la forme polaire d'une forme quadratique q positive sur E est non dégénérée si et seulement si q est définie. À ce titre, on peut aussi définir un produit scalaire comme une forme bilinéaire symétrique définie positive.

Nombre de propriétés et de résultats vus dans ce chapitre ne dépendent pas du fait que l'espace vectoriel E soit de dimension finie. Néanmoins, on appelle *espace préhilbertien* tout espace vectoriel muni d'un produit scalaire. Si l'espace vectoriel est de dimension finie, on parle d'*espace euclidien*.

Exemple 3.2

- *$E = \mathbb{R}^n$, $\langle x, y \rangle = \sum_{i=1}^{N} x_i y_i$ est un produit scalaire sur E, qui munit $\mathbb{R}^n$ de sa structure euclidienne dite* canonique.
- *$E = C^0([a, b], \mathbb{R})$, $\langle f, g \rangle = \int_a^b f(t)g(t)\,\mathrm{d}t$ est un produit scalaire sur E*
- *$E = \ell^2 = \{(u_k)_{k\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \mid \sum_{k\in\mathbb{N}} (u_k)^2 < +\infty\}$, $\langle u, v \rangle = \sum_{k\in\mathbb{N}} u_k v_k$ est un produit scalaire sur E. C'est l'extension naturelle du premier exemple de produit scalaire en dimension infinie.*
- *$E = \mathbb{R}[X]$, $\langle P, Q \rangle = \int_{-1}^{1} P(t)Q(t)\,\mathrm{d}t$, $\langle P, Q \rangle = \int_0^{+\infty} P(t)Q(t)e^{-t}\,\mathrm{d}t$ et $\langle P, Q \rangle = \int_{-1}^{1} \frac{P(t)Q(t)}{\sqrt{1-t^2}}\,\mathrm{d}t$ sont trois produits scalaires sur E*

Théorème 3.3 *Soit E un espace préhilbertien réel, de produit scalaire noté $\langle \cdot, \cdot \rangle$. L'application de E dans $\mathbb{R}$ définie par*

$$\forall x \in E,\ \|x\| = \sqrt{\langle x, x \rangle},$$

est une ***norme*** *sur E, c'est-à-dire une application satisfaisant aux propriétés suivantes :*

- *séparation : $\forall x \in E$, $\|x\| = 0 \implies x = 0_E$,*
- *absolue homogénéité : $\forall (\lambda, x) \in \mathbb{R} \times E$, $\|\lambda x\| = |\lambda|\,\|x\|$,*
- *sous-additivité : $\forall (x, y) \in E^2$, $\|x + y\| \leq \|x\| + \|y\|$.*

On l'appelle ***norme associée au*** *(ou* ***dérivant du****)* ***produit scalaire****.*

DÉMONSTRATION. On a tout d'abord

$$\forall x \in E,\ \|x\| = 0 \iff \sqrt{\langle x, x \rangle} = 0 \iff \langle x, x \rangle = 0 \iff x = 0_E,$$

puisque la forme quadratique associée au produit scalaire est définie positive.

On a ensuite

$$\forall \lambda \in \mathbb{R},\ \forall x \in E,\ \|\lambda x\| = \sqrt{\langle \lambda x, \lambda x\rangle} = \sqrt{\lambda^2 \langle x, x\rangle} = |\lambda| \sqrt{\langle x, x\rangle} = |\lambda|\, \|x\|.$$

Enfin, la sous-additivité découle de l'inégalité de Minkowski. □

Une norme est toujours positive. En effet, on a

$$\forall x \in E,\ 0 = \|0_E\| = \|x - x\| \leq \|x\| + \|-x\| = \|x\| + \|x\| = 2\,\|x\|.$$

Une application de E dans $\mathbb{R}$ qui ne satisfait que les propriétés d'homogénéité et de sous-additivité énoncées plus haut est appelée une *semi-norme*.

L'inégalité de Cauchy–Schwarz dans un espace préhilbertien se réécrit :

$$\forall (x, y) \in E^2,\ |\langle x, y\rangle| \leq \|x\|\|y\|.$$

Un espace vectoriel muni d'une norme est appelé un *espace vectoriel normé*.

Toutes les normes sur un espace vectoriel ne dérivent pas nécessairement d'un produit scalaire. C'est le cas par exemple de la norme $\|(x_1, x_2)\| = \max(|x_1|, |x_2|)$ pour $x = (x_1, x_2) \in \mathbb{R}^2$. Cette application vérifie les axiomes d'une norme donnés dans le théorème 3.3, mais ne correspond à aucun produit scalaire.

Proposition 3.4 *Soit E un espace préhilbertien et $\|\cdot\|$ une norme euclidienne sur E. Pour tous vecteurs x et y de E, on a*

$$\|x + y\| = \|x\| + \|y\|$$

si et seulement si la famille $\{x, y\}$ est positivement liée, i.e. si $x = 0$ ou s'il existe un réel positif λ tel que $y = \lambda x$.

DÉMONSTRATION. Si le vecteur x est nul, l'égalité est évidente. Si x est non nul et que $y = \lambda x$, avec λ un réel positif, on a

$$\|x + y\| = \|(1 + \lambda)x\| = (1 + \lambda)\|x\| = \|x\| + \lambda\|x\| = \|x\| + \|\lambda x\| = \|x\| + \|y\|.$$

Réciproquement, si on a égalité alors

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

d'où $\langle x, y\rangle = \|x\|\|y\|$, ce qui correspond au cas d'égalité dans l'inégalité de Cauchy-Schwarz et implique que la famille $\{x, y\}$ est liée. Ainsi, soit x est nul, soit $y = \lambda x$ avec λ un réel. On a donc

$$\lambda\|x\|^2 = \lambda\langle x, y\rangle = \|x\|\|y\| \geq 0 \implies \lambda \geq 0.$$

□

Théorème 3.5 (identité du parallélogramme) *Soit E un espace préhilbertien réel. On a*

$$\forall (x, y) \in E^2,\ \|x + y\|^2 + \|x - y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right).$$

DÉMONSTRATION. On a d'une part

$$\forall (x, y) \in E^2,\ \|x + y\|^2 = \|x\|^2 + 2\langle x, y\rangle + \|y\|^2,$$

et d'autre part

$$\forall (x, y) \in E^2,\ \|x - y\|^2 = \|x\|^2 - 2\langle x, y\rangle + \|y\|^2.$$

L'égalité de l'énoncé s'obtient en sommant les deux précédentes. □

Remarque 3.6 *Il existe une réciproque à ce théorème. Si $(E, \|\cdot\|)$ est un espace vectoriel normé dans lequel l'identité est vérifiée, alors $\langle x, y\rangle = \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right)$ est un produit scalaire associé à la norme $\|\cdot\|$.*

L'identité s'interprète géométriquement de la manière suivante : la somme des carrés des longueurs des quatre côtés d'un parallélogramme est égale à la somme des carrés des longueurs de ses deux diagonales. Elle est par ailleurs équivalente au résultat suivant, connu sous le nom du *théorème de la médiane* ou *d'Apollonius* (de Perge) : *soit ABC un triangle du plan et I le milieu du segment $[BC]$, alors*

$$AB^2 + AC^2 = 2\left(BI^2 + AI^2\right).$$

La définition suivante est une conséquence de l'inégalité de Cauchy–Schwarz.

Définition 3.7 (écart angulaire) *Soit E un espace préhilbertien réel. Pour tous vecteurs x et y de E, il existe un unique réel θ appartenant à l'intervalle $[0, \pi]$ tel que*

$$\cos(\theta) = \frac{\langle x, y \rangle}{\|x\| \|y\|}.$$

*Ce réel est appelé **écart angulaire entre x et y** et correspond à une mesure d'angle non orienté.*

3.2 Orthogonalité dans les espaces euclidiens

La notion d'orthogonalité telle qu'on l'a introduite pour une forme quadratique subsiste bien évidemment avec un produit scalaire, avec l'avantage qu'une telle forme est définie positive.

3.2.1 Généralités

Dans la suite, on notera $\langle \cdot, \cdot \rangle$ le produit scalaire.

Définitions 3.8 *Soit E un espace préhilbertien. Deux vecteurs x et y de E sont dits **orthogonaux**, ce que l'on note $x \perp y$, si et seulement si $\langle x, y \rangle = 0$.*

*Soit A une partie non vide de E. On appelle **orthogonal de** A, et on note $A^\perp$, l'ensemble*

$$A^\perp = \{x \in E \mid \forall y \in A,\ \langle x, y \rangle = 0\}.$$

Le produit scalaire étant une forme symétrique, deux vecteurs x et y sont orthogonaux si et seulement si $\langle y, x \rangle = 0$.

On a vu précédemment que $A^\perp$ était un sous-espace vectoriel.

3.2.2 Bases orthonormées

Définitions 3.9 *Soit E un espace préhilbertien. Une famille de vecteurs non nuls $\{e_i\}_{i \in I}$ de E est dite **orthogonale** si elle vérifie*

$$\forall (i, j) \in I^2,\ i \neq j,\ \left\langle e_i, e_j \right\rangle = 0.$$

*Une telle famille est dite **orthonormale** (ou **orthonormée**) si l'on a de plus*

$$\forall i \in I,\ \|e_i\| = 1.$$

Proposition 3.10 (relation de Pythagore) *Soit E un espace euclidien, p un entier naturel non nul et $\{e_i\}_{i=1,\dots,p}$ une famille orthogonale de vecteurs de E. On a*

$$\|\sum_{i=1}^{p} e_i\|^2 = \sum_{i=1}^{p} \|e_i\|^2.$$

DÉMONSTRATION. laissée en exercice (récurrence) □

Proposition 3.11 *Une famille orthogonale est libre.*

DÉMONSTRATION. laissée en exercice □

Il découle de cette dernière propriété que l'on peut parler de *base orthogonale* (ou *orthonormale*) d'un espace préhilbertien.

Proposition 3.12 (coordonnées dans une base orthonormée) *Si E est un espace euclidien et si $\{e_1,\dots,e_n\}$ est une base orthonormée de E, on a*

$$\forall x\in E,\ x=\sum_{i=1}^{n}x_ie_i,\ \forall y\in E,\ y=\sum_{i=1}^{n}y_ie_i,\ \langle x,y\rangle=\sum_{i=1}^{n}x_iy_i.$$

De plus, les coordonnées d'un vecteur x de E dans cette base vérifient

$$\forall i\in\{1,\dots,n\},\ x_i=\langle x,e_i\rangle,$$

et l'on a

$$\|x\|^2=\sum_{i=1}^{n}\langle x,e_i\rangle^2.$$

DÉMONSTRATION. laissée en exercice □

On retiendra de cette proposition que l'évaluation du produit scalaire entre deux vecteurs est aisée lorsque l'on travaille dans une base orthonormée.

Remarque 3.13 *Cette façon de déterminer les coordonnées d'un vecteur s'applique également lorsque l'on veut déterminer la matrice d'un endomorphisme u de E dans une base orthonormée $\mathscr{B}=\{e_1,\dots,e_n\}$. On a en effet*

$$\forall(i,j)\in\{1,\dots,n\}^2,\ (mat_{\mathscr{B}}(u))_{ij}=\left\langle u(e_j),e_i\right\rangle.$$

Une manière de construire une base orthonormée à partir d'une base quelconque de l'espace est donnée par le *procédé d'orthonormalisation de Gram–Schmidt*, décrit dans la preuve du théorème suivant.

Théorème 3.14 *Soit E un espace euclidien de dimension n et $\{f_1,\dots,f_n\}$ une base de E. Alors, il existe une unique base orthonormée $\{e_1,\dots,e_n\}$ de E telle que*

$$\forall j\in\{1,\dots,n\},\ \text{Vect}(\{e_1,\dots,e_j\})=\text{Vect}(\{f_1,\dots,f_j\})\ \textit{et}\ \left\langle f_j,e_j\right\rangle>0.$$

DÉMONSTRATION. On va construire la base orthonormée par récurrence en procédant comme suit.

On pose tout d'abord $e_1=\frac{f_1}{\|f_1\|}$. On suppose ensuite avoir construit une famille $\{e_1,\dots,e_j\}$, avec j un entier de $\{1,\dots,n-1\}$. On définit alors

$$e_{j+1}=\frac{f_{j+1}-\sum_{k=1}^{j}\left\langle f_{j+1},e_k\right\rangle e_k}{\|f_{j+1}-\sum_{k=1}^{j}\left\langle f_{j+1},e_k\right\rangle e_k\|}.$$

Vérifions les différentes propriétés énoncées dans le théorème. Il est tout d'abord clair que chaque vecteur e_j, avec j appartenant à $\{1,\dots,n\}$, appartient à $\text{Vect}(\{f_1,\dots,f_j\})$ en utilisant les formules ci-dessus et en raisonnant par récurrence. Réciproquement, chaque vecteur f_j, avec j appartenant à $\{1,\dots,n\}$, appartient à $\text{Vect}(\{e_1,\dots,e_j\})$.

Le fait de pouvoir normaliser le vecteur à chaque étape en divisant $f_{j+1}-\sum_{k=1}^{j}\left\langle f_{j+1},e_k\right\rangle e_k$ par sa norme provient du fait que $\{f_1,\dots,f_n\}$ est une une famille libre et de la propriété précédente. Ceci permet en particulier d'assurer que le processus itératif qu'on a introduit est bien défini (on n'a pas de division par zéro).

Il faut ensuite montrer que la famille obtenue est orthogonale. Pour cela, il suffit de montrer que le vecteur e_j, avec j appartenant à $\{2,\dots,n\}$, est orthogonal à tout vecteur e_i, avec i appartenant à $\{1,\dots,j-1\}$, en raisonnant par récurrence. Pour $j=2$, on a

$$\langle e_2,e_1\rangle=\left\langle\frac{f_2-\langle f_2,e_1\rangle\ e_1}{\|f_2-\langle f_2,e_1\rangle\ e_1\|},e_1\right\rangle=\frac{1}{\|f_2-\langle f_2,e_1\rangle\ e_1\|}\left(\langle f_2,e_1\rangle-\langle f_2,e_1\rangle\|e_1\|^2\right)=0.$$

Pour j supérieur ou égal à 3, on fait l'hypothèse de récurrence

$$\forall(i,k)\in\{1,\dots,j-1\}^2,\langle e_i,e_k\rangle=\delta_{ik}.$$

On a alors

$$\begin{aligned}
\forall i \in \{1,\ldots,j-1\},\ \left\langle e_j, e_i\right\rangle &= \frac{1}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|}\left\langle f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k, e_i\right\rangle \\
&= \frac{1}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|}\left(\left\langle f_j, e_i\right\rangle - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle \langle e_k, e_i\rangle\right) \\
&= \frac{1}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|}\left(\left\langle f_j, e_i\right\rangle - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle \delta_{ik}\right) \\
&= \frac{1}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|}\left(\left\langle f_j, e_i\right\rangle - \left\langle f_j, e_i\right\rangle\right) \\
&= 0.
\end{aligned}$$

Enfin, on a, pour tout entier j de $\{1,\ldots,n\}$,

$$\|e_j\|^2 = \left\langle e_j, e_j\right\rangle = \frac{1}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|}\left\langle f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k, e_j\right\rangle = \frac{\left\langle f_j, e_j\right\rangle}{\|f_j - \sum_{k=1}^{j-1}\left\langle f_j, e_k\right\rangle e_k\|},$$

d'où $\left\langle f_j, e_j\right\rangle > 0$.

Il nous reste à montrer que la base est unique. Toute base $\{e_1,\ldots,e_n\}$ satisfaisant aux propriétés précédentes est telle que le vecteur e_1 appartient à $\mathrm{Vect}(\{f_1\})$, c'est-à-dire que $e_1 = \lambda f_1$. Puisque $\|e_1\| = 1$ et $\langle e_1, f_1\rangle > 0$, on trouve que $\lambda = \frac{1}{\|f_1\|}$.

Supposons ensuite que les vecteurs $e_1,\ldots,e_j$ soient de la forme donnée pour j superieur ou égal à 1. On sait que e_{j+1} appartient à $\mathrm{Vect}(\{f_1,\ldots,f_{j+1}\}) = \mathrm{Vect}(\{e_1,\ldots,e_j,f_{j+1}\})$, d'où

$$e_{j+1} = \sum_{k=1}^{j}\lambda_k e_k + \lambda f_{j+1}.$$

En considérant successivement les produits scalaires de ce vecteur avec les vecteurs e_i pour i appartenant à $\{1,\ldots,j\}$, on trouve que $\lambda_i = -\lambda\left\langle f_{j+1}, e_i\right\rangle$ et on a par conséquent

$$e_{j+1} = \lambda\left(f_{j+1} - \sum_{k=1}^{j}\left\langle f_{j+1}, e_k\right\rangle e_k\right).$$

Les conditions $\|e_{j+1}\| = 1$ et $\left\langle e_{j+1}, f_{j+1}\right\rangle > 0$ permettent alors de déterminer λ, conduisant à la forme donnée. □

Plus que l'existence de bases orthonormées, la preuve du théorème précédent donne un algorithme rapide pour obtenir une base orthonormée à partir d'une base quelconque.

Exemple 3.15 *Soient* $u = \begin{pmatrix}1\\0\\1\end{pmatrix}$, $v = \begin{pmatrix}1\\1\\1\end{pmatrix}$ *et* $w = \begin{pmatrix}-1\\-1\\0\end{pmatrix}$. *La famille* (u,v,w) *est libre. Appliquons lui le procédé d'orthonormalisation de Gram–Schmidt. Tout d'abord, on pose* $e_1 = \frac{u}{\|u\|}$. *Comme* $\|u\| = \sqrt{2}$, *on pose*

$$e_1 = \begin{pmatrix}1/\sqrt{2}\\0\\1/\sqrt{2}\end{pmatrix}.$$

On cherche ensuite e'_2 *de la forme* $e'_2 = v + \lambda e_1$ *de sorte que* $\langle e'_2, e_1\rangle = 0$. *On a*

$$\begin{aligned}
\langle e'_2, e_1\rangle &= \langle v, e_1\rangle + \lambda\langle e_1, e_1\rangle \\
&= \frac{2}{\sqrt{2}} + \lambda \\
&= \sqrt{2} + \lambda.
\end{aligned}$$

On pose donc $\lambda = -\sqrt{2}$, *ce qui donne*

$$e_2' = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}.$$

Ce vecteur est de norme 1, *donc on pose* $e_2' = e_2$. *Cherchons ensuite* e_3' *de la forme* $e_3' = w + \lambda e_1 + \mu e_2$, *avec les contraintes* $\langle e_3', e_1\rangle = 0$ *et* $\langle e_3', e_2\rangle = 0$. *On trouve*

$$\begin{aligned}\langle e_3', e_1\rangle &= \langle w, e_1\rangle + \lambda\langle e_1, e_1\rangle \\ &= -\frac{1}{\sqrt{2}} + \lambda,\end{aligned}$$

d'où $\lambda = \frac{1}{\sqrt{2}}$. *Ensuite, on a*

$$\begin{aligned}\langle e_3', e_2\rangle &= \langle w, e_2\rangle + \mu\langle e_2, e_2\rangle \\ &= -1 + \mu,\end{aligned}$$

d'où $\mu = 1$. *Ainsi, on pose*

$$e_3' = \begin{pmatrix} -\frac{1}{2} \\ 0 \\ \frac{1}{2} \end{pmatrix}.$$

On calcule $\|e_3'\| = 1/\sqrt{2}$, *et on pose* $e_3 = e_3'/\|e_3'\|$, *c'est-à-dire*

$$e_3 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

La base (e_1, e_2, e_3) *est la base orthonormale obtenue avec le procédé de Gram–Schmidt en partant de* (u, v, w).

Le procédé de Gram–Schmidt peut aussi s'appliquer en dimension infinie, quand on dispose d'une suite $(u_n)_{n\in\mathbb{N}}$ de vecteurs d'un espace préhilbertien E, telle que $\{u_1, \ldots, u_n\}$ soit libre pour tout entier naturel non nul n, pour former une suite de vecteurs orthonormés. Nous verrons des exemples importants en TD, où l'on obtiendra des familles de polynômes orthogonaux.

Dans la construction ci-dessus, on peut observer que la matrice de passage de la base $\{f_1, \ldots, f_n\}$ à la base $\{e_1, \ldots, e_n\}$ est triangulaire supérieure, à coefficients diagonaux strictement positifs (ce sont les inverses de normes de vecteurs).

Corollaire 3.16 *Un espace euclidien non réduit au vecteur nul possède une base orthonormée.*

Proposition 3.17 *Soit* E *un espace euclidien et* A *un sous-espace vectoriel de* E. *On a*

$$E = A \oplus A^{\perp} \text{ et } (A^{\perp})^{\perp} = A$$

DÉMONSTRATION. Supposons que l'espace E est de dimension n et que le sous-espace A est de dimension $p \leq n$. En vertu du théorème de la base incomplète, il est possible de compléter toute base de A en une base de E. En appliquant le procédé d'orthonormalisation de Gram–Schmidt à une telle base, on obtient une base orthonormée de E dont il n'est pas difficile de voir (il suffit pour cela d'utiliser les propriétés énoncées dans le théorème 3.14) que ses p premiers vecteurs forment une base orthonormée de A et les $n-p$ suivants une base orthonormée de $A^{\perp}$. Ceci permet de conclure que $E = A \oplus A^{\perp}$, et donc que $\dim(E) = \dim(A) + \dim(A^{\perp})$.

Pour prouver la seconde assertion, on sait déjà [1] que $A \subset (A^{\perp})^{\perp}$ et on utilise que $\dim(A) = \dim(E) - \dim(A^{\perp}) = \dim((A^{\perp})^{\perp})$. □

Remarque 3.18 *Cette proposition reste vraie dans un espace préhilbertien si le sous-espace* A *est de dimension* finie.

1. C'est un résultat de la proposition 2.24.

3.2.3 Projection orthogonale

Proposition et définition 3.19 (projection et symétrie orthogonales) *Soit E un espace euclidien et A un sous-espace vectoriel de E. Tout vecteur x de E s'écrit, de manière unique, comme la somme $x = u + v$ avec u appartenant à A et v appartenant à $A^\perp$. On appelle* ***projection orthogonale de x sur*** *A le vecteur u ainsi défini, et on note p_A l'application de E dans E telle que $x \mapsto u = p_A(x)$ ainsi construite. On appelle* ***symétrie orthogonale de x par rapport à*** *A le vecteur $u - v$ et on note s_A l'application de E dans E telle que $x \mapsto u - v = 2\,p_A(x) - x = s_A(x)$ ainsi construite.*

Rappel : étant donné un espace vectoriel E, A un sous-espace vectoriel de E et B un supplémentaire de A, on définit la *projection de E sur A parallèlement à B* comme l'endomorphisme $p_{A,B}$ de E défini par

$$\forall x \in E,\ p_{A,B}(x) = x_A,$$

où $x = x_A + x_B$ est l'unique décomposition du vecteur x comme somme d'un vecteur de A et d'un vecteur de B. La projection orthogonale p_A est alors la projection sur A parallèlement à $A^\perp$.

Note : une projection est idempotente, i.e. $p \circ p = p$.

Proposition 3.20 (propriétés d'une projection orthogonale) *Soit A un sous-espace vectoriel d'un espace euclidien E. On a les propriétés suivantes :*

1. *$\forall x \in E$, $x = p_A(x) + p_{A^\perp}(x)$, $x - p_A(x) \in A^\perp$ et $\|p_A(x)\| \leq \|x\|$.*
2. *Si $\{e_1, \dots, e_p\}$ est une base orthonormée de A, alors*
$$\forall x \in E,\ p_A(x) = \sum_{i=1}^{p} \langle x, e_i \rangle\ e_i.$$
3. *La projection orthogonale sur A se caractérise variationnellement par la propriété de minimalité suivante :*
$$\forall x \in E,\ \|x - p_A(x)\| = \inf_{y \in A} \|x - y\|.$$
4. *Si p est un endomorphisme de E idempotent et si $\mathrm{Im}\,(p) = \ker(p)^\perp$, alors p est la projection orthogonale sur $\mathrm{Im}\,(p)$.*

DÉMONSTRATION.

1. C'est une conséquence du fait que $E = A \oplus A^\perp$ et que $(A^\perp)^\perp = A$, l'inégalité découlant de la relation de Pythagore.
2. Il suffit d'écrire la projection de x comme une combinaison linéaire des vecteurs de la base orthonormée et on en détermine les coefficients en utilisant que $x - p_A(x)$ appartient à $A^\perp$, c'est-à-dire en écrivant que
$$\forall x \in E,\ \forall i \in \{1, \dots, p\},\ \langle x - p_A(x), e_i \rangle = 0.$$
3. On écrit que
$$\forall x \in E,\ \forall y \in A,\ \|x - y\|^2 = \|x - p_A(x) + p_A(x) - y\|^2 = \|x - p_A(x)\|^2 + \|p_A(x) - y\|^2,$$
la seconde égalité découlant de l'orthogonalité.
4. On montre que $\ker(p) \oplus \mathrm{Im}\,(p) = E$ en utilisant le théorème du rang et en servant de l'orthogonalité entre $\ker(p)$ et $\mathrm{Im}\,(p)$ pour en déduire que $\ker(p) \cap \mathrm{Im}\,(p) = \{0_E\}$. On conclut par la définition de la projection orthogonale.

□

Remarque 3.21 *La seconde propriété permet de réécrire la formule donnant les vecteurs de la base orthonormée obtenue par le procédé de Gram-Schmidt sous la forme*

$$\forall j \in \{1, \dots, n\},\ e_j = \frac{f_j - p_{\mathrm{Vect}(\{e_1, \dots, e_{j-1}\})}(f_j)}{\|f_j - p_{\mathrm{Vect}(\{e_1, \dots, e_{j-1}\})}(f_j)\|}.$$

Application à la résolution d'un problème de régression

Supposons que l'on dispose de m observations, représentées par les réels y_i, $i = 1, \dots, m$, associées à m conditions, représentées par les réels x_i, $i = 1, \dots, m$. On souhaite approcher le nuage de points (x_i, y_i) du plan par le graphe d'une fonction φ « simple » qui dépend linéairement de n coefficients c_i, $i = 1, \dots, n$. Le *problème de régression* consiste à déterminer « au mieux » les coefficients de la fonction φ, c'est-à-dire de manière à minimiser (en un sens à préciser) la distance entre le vecteur $(y_1, \dots, y_m)$ et le vecteur $(\varphi(x_1), \dots, \varphi(x_m))$.

D'un point de vue algébrique, le problème de régression se pose comme la résolution du système linéaire

$$AC = Y,$$

où $C = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix}$, $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix}$ et A est une matrice de $M_{m,n}(\mathbb{R})$ dépendant des conditions x_i, $i = 1, \dots, m$. Étant donné qu'on a $m > n$ dans la plupart des cas en pratique, ce système est sur-déterminé et il n'admet en général pas de solution. En effet, l'image de A est un sous-espace vectoriel de $M_{m,1}(\mathbb{R})$ de dimension au plus égale à n et le vecteur Y n'en fait généralement pas partie. L'introduction du *problème aux moindres carrés* vise à pallier ce problème en considérant la détermination d'un vecteur C minimisant la distance (euclidienne) entre Y et AC.

Définition 3.22 (problème aux moindres carrés) *Étant donné une matrice A de $M_{m,n}(\mathbb{R})$, avec $m > n$, et un vecteur Y de $M_{m,1}(\mathbb{R})$, le problème aux moindres carrés consiste à trouver un vecteur C de $M_{n,1}(\mathbb{R})$ minimisant la norme euclidienne de $Y - AC$.*

Cherchons le vecteur C réalisant $\inf_{C \in M_{n,1}(\mathbb{R})} \|Y - AC\|$. Par la caractérisation variationnelle de la projection orthogonale (voir la proposition 3.20), ce vecteur C sera tel que AC est la projection orthogonale de Y sur $\mathrm{Im}(A)$, ce qui signifie encore que le résidu $Y - AC$ doit être orthogonal $\mathrm{Im}(A)$, c'est-à-dire à AZ pour tout choix de vecteur Z de $M_{n,1}(\mathbb{R})$. Dans une base orthonormée, on aura ainsi

$$\forall Z \in M_{n,1}(\mathbb{R}),\ (AZ)^\top (Y - AC) = 0 \iff \forall Z \in M_{n,1}(\mathbb{R}),\ Z^\top A^\top (Y - AC) = 0,$$

ce qui équivaut à avoir

$$A^\top AC = A^\top Y,$$

ce dernier système portant le nom d'*équations normales*. On peut montrer (à titre d'exercice) que la matrice $A^\top A$ est inversible si et seulement si le rang de A est égal à n. Il en découle que le problème aux moindres carrés possède une unique solution si et seulement si le rang de la matrice A est égal au nombre de colonnes de cette dernière. Lorsque c'est le cas, la solution de $AC = Y$ au sens des moindres carrés est donnée par

$$\hat{C} = (A^\top A)^{-1} A^\top Y,$$

la matrice à n lignes et m colonnes $A^+ = (A^\top A)^{-1} A^\top$ étant appelée le *pseudo-inverse* (ou l'*inverse généralisé*) de A.

Mentionnons pour terminer cette section le théorème de Riesz, qui décrit l'ensemble des formes linéaires sur un espace euclidien.

Théorème 3.23 (théorème de représentation de Riesz) *Soit E un espace euclidien. Pour tout forme linéaire ℓ sur E, il existe un unique vecteur y de E tel que*

$$\forall x \in E,\ \ell(x) = \langle x, y \rangle.$$

DÉMONSTRATION. Soit $(e_1, \dots, e_n)$ une base orthonormée de E. Soit $y = \sum_{i=1}^n \ell(e_i) e_i$. On a alors pour $x = \sum_{i=1}^n x_i e_i$,

$$\ell(x) = \sum_{i=1}^n \ell(e_i) x_i = \langle x, y \rangle,$$

en utilisant l'expression du produit scalaire entre deux vecteurs décomposés dans une base orthonormée. □

3.3 Endomorphismes des espaces euclidiens

Dans toute cette section, sauf mention contraire, E désigne un espace euclidien de dimension n non nulle.

3.3.1 Adjoint d'un endomorphisme

Théorème et définition 3.24 (adjoint d'un endomorphisme) *Soit u un endomorphisme de E. Il existe un unique endomorphisme de E, noté u^* et appelé **adjoint de** u, tel que*

$$\forall (x, y) \in E^2,\ \langle u(x), y\rangle = \langle x, u^*(y)\rangle\,.$$

De plus, si $\mathscr{B}$ est une base orthonormée de E, on a

$$mat_{\mathscr{B}}(u^*) = mat_{\mathscr{B}}(u)^\top.$$

DÉMONSTRATION. Soit $\mathscr{B}$ une base orthonormée de E et M la matrice de l'endomorphisme u dans la base $\mathscr{B}$. On cherche un endomorphisme v de E qui soit l'adjoint de u, de matrice N dans la base $\mathscr{B}$. La relation de l'énoncé s'écrit matriciellement

$$\forall (X, Y) \in (M_{n,1}(\mathbb{R}))^2,\ (MX)^\top Y = X^\top N Y \iff \forall (X, Y) \in (M_{n,1}(\mathbb{R}))^2,\ X^\top M^\top Y = X^\top N Y.$$

La matrice N est donc entièrement et uniquement déterminée, et l'endomorphisme v également. □

3.3.2 Isométries vectorielles

Définition 3.25 (isométrie vectorielle) *Soit E un espace préhilbertien et u un endomorphisme de E. On dit que u est une **isométrie vectorielle** (ou encore un **endomorphisme orthogonal**) si*

$$\forall x \in E,\ \|u(x)\| = \|x\|.$$

On note $O(E)$ l'ensemble des isométries vectorielles de E.

Toute application qui conserve la norme n'est pas nécessairement linéaire, et n'est donc pas nécessairement une isométrie vectorielle. Un exemple est donné par l'application $x \mapsto \|x\|\, e$, avec e un vecteur unitaire. Une isométrie vectorielle est un *endomorphisme* qui conserve la norme.

On déduit immédiatement de la définition que les seules valeurs propres réelles possibles pour une isométrie vectorielle sont -1 et 1. Par exemple, les seules homothéties $x \mapsto \lambda x$ qui sont des isométries vectorielles sont $-id_E$ et id_E.

Proposition 3.26 *Soit E un espace préhilbertien. Une application u de E dans E est une isométrie vectorielle si et seulement si*

$$\forall (x, y) \in E^2,\ \langle u(x), u(y)\rangle = \langle x, y\rangle\,.$$

DÉMONSTRATION. Supposons que u est une isométrie vectorielle de E. Par définition, c'est un endomorphisme de E vérifiant

$$\forall x \in E,\ \|u(x)\| = \|x\|.$$

On a alors, en vertu de l'identité de polarisation associée au produit scalaire sur E et au carré de la norme associée,

$$\begin{aligned}
\forall (x, y) \in E^2,\ \langle u(x), u(y)\rangle &= \frac{1}{4}\left(\|u(x) + u(y)\|^2 - \|u(x) - u(y)\|^2\right)\\
&= \frac{1}{4}\left(\|u(x + y)\|^2 - \|u(x - y)\|^2\right)\\
&= \frac{1}{4}\left(\|x + y\|^2 - \|x - y\|^2\right)\\
&= \langle x, y\rangle\,.
\end{aligned}$$

Réciproquement, en choisissant $x = y$ dans l'égalité de l'énoncé, on trouve que

$$\forall x \in E,\ \|u(x)\|^2 = \|x\|^2,$$

et l'application u préserve la norme. Il reste à montrer qu'elle est linéaire. On a, pour tout réel λ et tous vecteurs x et y de E,

$$\begin{aligned}
\|u(\lambda x + y) - \lambda u(x) - u(y)\|^2 &= \|u(\lambda x + y)\|^2 + \|\lambda u(x)\|^2 + \|u(y)\|^2 \\
&\quad - 2\langle u(\lambda x + y), \lambda u(x)\rangle - 2\langle u(\lambda x + y), u(y)\rangle + 2\langle \lambda u(x), u(y)\rangle \\
&= \|u(\lambda x + y)\|^2 + \lambda^2 \|u(x)\|^2 + \|u(y)\|^2 \\
&\quad - 2\lambda \langle u(\lambda x + y), u(x)\rangle - 2\langle u(\lambda x + y), u(y)\rangle + 2\lambda \langle u(x), u(y)\rangle \\
&= \|\lambda x + y\|^2 + \lambda^2 \|x\|^2 + \|y\|^2 - 2\lambda \langle \lambda x + y, x\rangle - 2\langle \lambda x + y, y\rangle + 2\lambda \langle x, y\rangle \\
&= \|\lambda x + y\|^2 + \|\lambda x\|^2 + \|y\|^2 - 2\langle \lambda x + y, \lambda x\rangle - 2\langle \lambda x + y, y\rangle + 2\langle \lambda x, y\rangle \\
&= \|\lambda x + y - \lambda x - y\|^2 \\
&= 0.
\end{aligned}$$

Par propriété de la norme, il vient alors que

$$\forall \lambda \in \mathbb{R},\ \forall (x,y) \in E^2,\ u(\lambda x + y) - \lambda u(x) - u(y) = 0_E,$$

d'où la conclusion. □

Dans la précédente proposition, on observera que l'application n'est pas supposée linéaire. Comme on peut le voir dans la preuve, la linéarité se déduit de la préservation du produit scalaire.

Il découle de cette caractérisation qu'une isométrie vectorielle conserve l'orthogonalité, mais toute application conservant l'orthogonalité n'est pas nécessairement une isométrie vectorielle, comme le montre l'exemple d'une homothétie de rapport différent de -1 ou 1.

Théorème 3.27 *Une isométrie vectorielle de E est un endomorphisme inversible de E dont l'inverse est donné par son adjoint.*

DÉMONSTRATION. Soit u une isométrie vectorielle de E. Pour tout vecteur x appartenant au noyau de u, on a

$$0 = \|u(x)\| = \|x\| \implies x = 0_E,$$

et l'endomorphisme est donc injectif. L'espace E étant de dimension finie, ceci équivaut à dire que u est bijectif.

Enfin, on a, d'après la précédente proposition,

$$\forall (x,y) \in E^2,\ \langle x, y\rangle = \langle u(x), u(y)\rangle = \langle x, u^*(u(y))\rangle,$$

d'où $u^* \circ u = id_E$. □

Proposition 3.28 *Soit u une isométrie vectorielle de E. Si A est un sous-espace vectoriel de E stable par u, alors son orthogonal $A^\perp$ est aussi stable par u.*

DÉMONSTRATION. L'endomorphisme u étant injectif, on a $\dim(u(A)) = \dim(A)$ et, par stabilité de A par u, on peut conclure que $u(A) = A$. On a alors

$$\forall x \in A^\perp,\ \forall y \in A,\ \langle u(x), u(y)\rangle = \langle x, y\rangle = 0,$$

d'où $u(x)$ appartient à $(u(A))^\perp = A^\perp$. □

Ce résultat est encore vrai si A est un sous-espace vectoriel de dimension finie d'un espace préhilbertien.

Proposition 3.29 *Un endomorphisme u de E est une isométrie vectorielle si et seulement si l'image d'une base orthonormale de E par u est une base orthonormale de E.*

DÉMONSTRATION. Soit $\mathscr{B} = \{e_1, \dots, e_n\}$ une base orthonormale de E. Supposons que l'endomorphisme u soit une isométrie vectorielle. On a

$$\forall (i,j) \in \{1, \dots, n\}^2,\ \left\langle u(e_i), u(e_j) \right\rangle = \left\langle e_i, e_j \right\rangle = \delta_{ij}.$$

La famille $\{u(e_1), \dots, u(e_n)\}$ est ainsi une famille libre de n vecteurs, c'est donc une base de E.

Réciproquement, supposons que l'endomorphisme u transforme la base $\mathscr{B}$ en une base orthonormale de E. On a alors

$$\forall x \in E,\ x = \sum_{i=1}^{n} x_i\, e_i,\ \|u(x)\|^2 = \|\sum_{i=1}^{n} x_i\, u(e_i)\|^2 = \sum_{i=1}^{n} {x_i}^2 \|u(e_i)\|^2 = \sum_{i=1}^{n} {x_i}^2 = \|x\|^2,$$

et u est une isométrie vectorielle. □

Proposition 3.30 (propriétés matricielles d'une isométrie vectorielle) *Soit $\mathscr{B}$ une base orthonormale de E. Un endomorphisme u de E est une isométrie vectorielle si et seulement si sa matrice M dans la base $\mathscr{B}$ est telle que*

$$M^\top M = M M^\top = I_n.$$

DÉMONSTRATION. Posons $\mathscr{B} = \{e_1, \dots, e_n\}$. Supposons que l'endomorphisme u soit une isométrie vectorielle. On a

$$\forall (i,j) \in \{1, \dots, n\}^2,\ (M^\top M)_{ij} = {C_i}^\top C_j,$$

où C_i et C_j sont respectivement les i^{e} et j^{e} colonnes de la matrice M, et par conséquent

$$\forall (i,j) \in \{1, \dots, n\}^2,\ (M^\top M)_{ij} = \left\langle u(e_i), u(e_j) \right\rangle = \left\langle e_i, e_j \right\rangle = \delta_{ij},$$

ce qui signifie encore que $M^\top M = I_n$. La matrice M est donc inversible, d'inverse $M^\top$, et on a donc aussi $M M^\top = I_n$.

Réciproquement, si la matrice M est telle que $M^\top M = M M^\top = I_n$, cela signifie que

$$\forall (i,j) \in \{1, \dots, n\}^2,\ \left\langle u(e_i), u(e_j) \right\rangle = \delta_{ij}.$$

L'image de la base $\mathscr{B}$ par l'endomorphisme u est donc une base orthonormée de E et on en déduit que u est une isométrie vectorielle en utilisant la proposition 3.29. □

Définition 3.31 (matrice orthogonale) *Soit n un entier naturel non nul. On appelle **matrice orthogonale** une matrice d'ordre n à coefficients réels telle que*

$$M^\top M = M M^\top = I_n.$$

On note $O(n)$ l'ensemble des matrices orthogonales d'ordre n.

La précédente proposition nous dit qu'un endomorphisme de E est une isométrie vectorielle si et seulement si sa matrice dans une base orthonormale quelconque de E est une matrice orthogonale. Ce résultat est *faux* si la base considérée n'est pas orthonormale.

Théorème 3.32 *Soit n un entier naturel non nul. Pour toute matrice M de $O(n)$, on a*

$$\det(M) = \pm 1.$$

*L'ensemble $O(n)$ est un sous-groupe de $GL(n, \mathbb{R})$, appelé **groupe orthogonal**.*

DÉMONSTRATION. Par propriété du déterminant, on a

$$\forall M \in M_n(\mathbb{R}),\ \det(M) = \det(M^\top),$$

et donc

$$\forall M \in O(n),\ (\det(M))^2 = \det(M M^\top) = \det(I_n) = 1.$$

Il en résulte que $O(n)$ est un sous-ensemble de $GL(n, \mathbb{R})$. Par ailleurs, la matrice I_n appartient à $O(n)$ et on a

$$\forall (M, n) \in (O(n))^2,\ (M^{-1})^{-1} = (M^\top)^{-1} = (M^{-1})^\top \text{ et } (MN)^{-1} = N^{-1} M^{-1} = N^\top M^\top = (MN)^\top,$$

ce qui permet de montrer que $O(n)$ est bien un sous-groupe de $GL(n, \mathbb{R})$. □

L'ensemble des matrices orthogonales de déterminant égal à 1 forme un sous-groupe du groupe orthogonal, appelé *groupe spécial orthogonal* et noté $SO(n)$.

3.3.3 Endomorphismes auto-adjoints

Définition 3.33 (endomorphisme auto-adjoint) *Un endomorphisme u de E est dit **auto-adjoint** (ou **symétrique**) si et seulement si $u^* = u$, ce qui équivaut encore à avoir*

$$\forall (x,y) \in E^2,\ \langle u(x), y\rangle = \langle x, u(y)\rangle .$$

*Par ailleurs, l'endomorphisme u est dit **anti-auto-adjoint** (ou **antisymétrique**) si et seulement si $u^* = -u$, c'est-à-dire si et seulement si*

$$\forall (x,y) \in E^2,\ \langle u(x), y\rangle = -\langle x, u(y)\rangle .$$

On peut montrer (exercice) que la matrice dans une base orthonormale d'un endomorphisme auto-adjoint est *symétrique*. Réciproquement, si la matrice dans une base orthonormale d'un endomorphisme est symétrique, alors l'endomorphisme est auto-adjoint.

Exemple 3.34 *Montrer qu'une projection orthogonale est un endomorphisme auto-adjoint.*

Proposition 3.35 *Soit E un espace euclidien et u un endomorphisme auto-adjoint ou anti-auto-adjoint de E. Si A est un sous-espace vectoriel de E stable par u, alors son orthogonal $A^\perp$ est aussi stable par u.*

DÉMONSTRATION. Il suffit de reprendre la preuve de la proposition 3.28, en utilisant que

$$\forall x \in A^\perp, \forall y \in A,\ \langle u(x), y\rangle = \pm \langle x, u(y)\rangle = 0,$$

selon que l'endomorphisme u est auto-adjoint ou anti-auto-adjoint. □

Théorème 3.36 (diagonalisation des endomorphismes auto-adjoints) *Soit un espace euclidien et un endomorphisme auto-adjoint de cet espace. Alors, l'endomorphisme est diagonalisable et ses sous-espaces propres sont deux à deux orthogonaux. Il existe donc une base orthonormale de l'espace formée de vecteurs propres. De plus, les valeurs propres de l'endomorphisme sont réelles.*

DÉMONSTRATION. On note E l'espace et u l'endomorphisme auto-adjoint de l'énoncé. On montre tout d'abord que le polynôme caractéristique de l'endomorphisme est scindé sur $\mathbb{R}$. On choisit pour cela une base orthonormée de E et on note M la matrice de l'endomorphisme u dans cette base. La matrice M est réelle symétrique (ou complexe hermitienne). Il existe une valeur propre λ de M, *a priori* complexe, et une matrice colonne non nulle Z, telles que $MZ = \lambda Z$. On a alors [2]

$$(MZ)^* Z = \overline{\lambda}\, Z^* Z = \overline{\lambda}\, \|Z\|^2 \text{ et } (MZ)^* Z = Z^* M^* Z = Z^* M Z = \lambda\, Z^* Z = \lambda\, \|Z\|^2,$$

d'où λ est réelle.

On montre ensuite que les sous-espaces propres sont deux à deux orthogonaux. Soit λ et μ deux valeurs propres distinctes de u, respectivement associées à des vecteurs propres x et y. On a alors

$$\lambda \langle x, y\rangle = \langle u(x), y\rangle = \langle x, u(y)\rangle = \mu \langle x, y\rangle ,$$

d'où $(\lambda - \mu)\langle x, y\rangle = 0$ et donc $\langle x, y\rangle = 0$.

On montre enfin que u est diagonalisable. Pour cela, on suppose que l'endomorphisme possède p valeurs propres distinctes $\lambda_1, \dots, \lambda_p$ et l'on considère la somme des sous-espaces propres

$$A = E_{\lambda_1} \oplus E_{\lambda_2} \oplus \cdots \oplus E_{\lambda_p}.$$

Il est clair que le sous-espace $u(A)$ est inclus dans A et, par la proposition 3.35, on a de plus que l'orthogonal $A^\perp$ est stable par u. La restriction de u à $A^\perp$, qui est un endomorphisme autoadjoint, possède alors une valeur propre si $A^\perp$ n'est pas réduit au vecteur nul. Ceci est impossible, puisque tous les vecteurs propres de u appartiennent à A par construction. Ainsi, le sous-espace A est égal à E et l'endomorphisme est par conséquent diagonalisable. □

On en déduit immédiatement le résultat suivant.

2. Lorsque E est un espace euclidien, et par conséquent réel, on doit se placer sur le *complexifié* de E, c'est-à-dire l'espace vectoriel sur $\mathbb{C}$ correspondant à $E \times E$ muni de l'addition usuelle et d'une multiplication externe par les nombre complexes, pour effectuer ces calculs.

Corollaire 3.37 (diagonalisation des matrices réelles symétriques) *Soit n un entier naturel non nul. Pour tout matrice M réelle symétrique d'ordre n, il existe une matrice orthogonale O d'ordre n telle que $O^\top MO$ est une matrice diagonale réelle.*

Ces deux derniers résultats sont très importants : ils affirment que tout endomorphisme auto-adjoint de $\mathbb{R}^n$ est représenté par une matrice diagonale, en remplaçant la base canonique de $\mathbb{R}^n$ par une autre base orthonormale. Ce résultat a des implications notamment pour l'étude des matrices hessiennes, des matrices de covariances, ou en science des données pour des méthodes de réduction de dimension comme l'*analyse en composantes principales*.

On rappelle qu'une matrice M symétrique est dite positive si la forme quadratique qui lui est canoniquement associée est positive. Elle est dite définie positive si la forme quadratique est définie positive. Le dernier corollaire montre par conséquent que M est positive si et seulement si toutes ses valeurs propres sont positives, et définie positive si et seulement si toutes ses valeurs propres sont strictement positives.

Corollaire 3.38 *Soit une forme quadratique sur un espace euclidien. Alors, il existe une base orthonormale de l'espace relativement à laquelle la matrice de la forme est diagonale.*

DÉMONSTRATION. On note E l'espace et q la forme quadratique de l'énoncé. Soit $\mathscr{B}$ une base orthonormée de E et M la matrice de la forme q relativement à la base $\mathscr{B}$. Cette dernière est symétrique et, d'après le précédent corollaire, il existe une matrice orthogonale O telle que la matrice $O^\top MO$ est diagonale réelle. La matrice O définit alors un changement de base faisant passer de $\mathscr{B}$ à une autre base orthonormée de E, qui est orthogonale pour la forme q (puisque la matrice de cette dernière relativement à cette nouvelle base est diagonale). □

Une application importante du résultat de diagonalisation d'un endomorphisme auto-adjoint est donnée par le corollaire suivant.

Corollaire 3.39 (décomposition en valeurs singulières) *Soit m et n des entiers naturels non nuls. Toute matrice M de $M_{m,n}(\mathbb{R})$ admet une **décomposition en valeurs singulières**, c'est-à-dire qu'il existe des matrices U de $M_m(\mathbb{R})$ et V de $M_n(\mathbb{R})$, toutes deux orthogonales, et Σ de $M_{m,n}(\mathbb{R})$, « diagonale » à coefficients positifs, telles que*

$$M = U\Sigma V^\top,$$

*les coefficients non nuls de Σ, appelés **valeurs singulières** de M et notés σ_i, $i = 1, \dots, r$, étant déterminés de manière unique. Si ces valeurs singulières sont distinctes, les colonnes des matrices U et V, respectivement appelées **vecteurs singuliers à gauche** et **à droite** de M, sont également déterminées de façon unique à un facteur multiplicatif unitaire près.*

Les valeurs singulières sont les valeurs propres non nulles des matrices $M^\top M$ et $MM^\top$. Les colonnes de U sont les vecteurs propres de $MM^\top$, celles de V les vecteurs propres de $M^\top M$, et l'on a

$$\forall k \in \{1, \dots, r\},\ {U_k}^\top M = \sigma_k V_k \text{ et } MV_k = \sigma_k U_k.$$