Un (premier morceau de) livre (pas encore) autorisé pour le concours de l'agrégation

Thomas Dedieu & Stéphane Lamy

Version septembre 2023
compilée le 8 septembre 2023

Avertissement. Les chapitres qui suivent sont extraits d'un projet en commun avec Stéphane Lamy. Je (ThD) assume néanmoins toute responsabilité quant aux âneries que vous pourriez trouver dans cette version préliminaire.

Malgré mes efforts, ce texte n'est pas encore complètement rédigé. En particulier il comporte des trous et des passages en style télégraphique. Il a cependant vocation à ne pas comporter d'erreur, même si je ne suis pas inconscient au point d'espérer que c'est effectivement le cas. Si vous en repérez il m'intéresse que vous me les communiquiez, ainsi que toute remarque que vous jugerez judicieuse.

Détail pratique utile : il y a des liens invisibles un peu partout, vous pouvez naviguer en cliquant dessus.

Table des matières

Chapitre 1

Pivot de Gauss et applications

Cette partie est consacrée à l'étude du Pivot de Gauss pour des matrices à coefficients dans un corps : $\mathbf{k}$ désigne un corps arbitraire (par définition, un corps est commutatif).

1.0.1 notation. Dans cette section on note parfois $[n]$ l'ensemble des entiers compris entre 1 et n, noté $[\![1, n]\!]$ d'habitude.

1.1 – Produits de matrices

On commence par reprendre les formules bien connues du produit matriciel, sous une forme peut-être méconnue bien qu'instructive et souvent fort utile. En particulier elles sont parfaitement adaptées à l'interprétation des opérations élémentaires sur les lignes et colonnes comme multiplication à gauche et à droite respectivement par des matrices inversibles. On donne les formules sans preuve, estimant que le public auquel est destiné ce livre saura s'en débrouiller.

1.1.1. La première règle élémentaire est la suivante. Soit $A = (A_1, \dots, A_q) \in \mathcal{M}_{p,q}(\mathbf{k})$, cette notation signifiant que A est la matrice constituée des colonnes $A_1, \dots, A_r \in \mathbf{k}^p = \mathcal{M}_{p,1}(\mathbf{k})$. Pour tout $x_1, \dots, x_q \in \mathbf{k}$, on a

$$A \times \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} = \left(\begin{array}{c|c|c} & & \\ A_1 & \cdots & A_q \\ & & \end{array}\right) \times \begin{pmatrix} x_1 \\ \vdots \\ x_q \end{pmatrix} = x_1.A_1 + \cdots + x_q.A_q \in \mathcal{M}_{p,1}(\mathbf{k}).$$

Cette règle détermine le produit de deux matrices de tailles compatibles arbitraires par la seconde règle suivante. Pour tout $B = (B_1, \dots, B_r) \in \mathcal{M}_{q,r}(\mathbf{k})$, on a

$$A \times \left(\begin{array}{c|c|c} & & \\ B_1 & \cdots & B_r \\ & & \end{array}\right) = \left(\begin{array}{c|c|c} & & \\ AB_1 & \cdots & AB_r \\ & & \end{array}\right) \in \mathcal{M}_{p,r}(\mathbf{k}).$$

Bien entendu on a la version transposée de ces deux règles élémentaires. Soit $A = (A^1, \dots, A^p)^\mathsf{T} \in \mathcal{M}_{p,q}(\mathbf{k})$, cette notation signifiant que A est la matrice constituée des lignes $A^1, \dots, A^p \in (\mathbf{k}^q)^\mathsf{T} = \mathcal{M}_{1,q}(\mathbf{k})$. Pour tout $x_1, \dots, x_p \in \mathbf{k}$, on a

$$\begin{pmatrix} x_1 & \cdots & x_p \end{pmatrix} \times A = \begin{pmatrix} x_1 & \cdots & x_p \end{pmatrix} \times \left(\begin{array}{c} \hline A^1 \\ \hline \vdots \\ \hline A^p \\ \hline \end{array}\right) = x_1.A^1 + \cdots + x_p.A^p \in \mathcal{M}_{1,q}(\mathbf{k}).$$

Pour tout $B = (B^1, \dots, B^r)^\mathsf{T} \in \mathcal{M}_{r,p}(\mathbf{k})$, on a

$$\left(\begin{array}{c} B^1 \\ \hline \vdots \\ \hline B^r \end{array}\right) \times A = \left(\begin{array}{c} B^1 A \\ \hline \vdots \\ \hline B^r A \end{array}\right) \in \mathcal{M}_{r,q}(\mathbf{k}).$$

1.1.2 Exemple. En particulier, on a les identités suivantes :

$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} \times \left(\begin{array}{c} A^1 \\ \hline \vdots \\ \hline A_d \end{array}\right) = \left(\begin{array}{c} d_1 A^1 \\ \hline \vdots \\ \hline d_n A_n \end{array}\right)$$

et

$$\left(\begin{array}{c|c|c} A_1 & \cdots & A_n \end{array}\right) \times \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} = \left(\begin{array}{c|c|c} d_1 A_1 & \cdots & d_n A_n \end{array}\right).$$

En conjuguant ces deux identités, on peut obtenir une autre identité utile :

$$\begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix} A \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_n \end{pmatrix}^{-1} = \left(\tfrac{d_i}{d_j} a_{ij}\right)_{1 \leqslant i,j \leqslant n}.$$

1.1.3. On a aussi les règles suivantes, utiles en particulier dans le contexte des formes bilinéaires (voir notamment dans cette section le théorème 1.3.6 à propos de la décomposition de Cholesky). Pour $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbf{k}$, on a

$$(x_1 \cdots x_n) \times \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = x_1 y_1 + \cdots + x_n y_n.$$

Soit

$$A = \left(\begin{array}{c|c|c} A_1 & \cdots & A_q \end{array}\right) = \left(\begin{array}{c} A^1 \\ \hline \vdots \\ \hline A^p \end{array}\right) \quad \text{et} \quad B = \left(\begin{array}{c|c|c} B_1 & \cdots & B_p \end{array}\right) = \left(\begin{array}{c} B^1 \\ \hline \vdots \\ \hline B^q \end{array}\right)$$

deux matrices de tailles $p \times q$ et $q \times p$ respectivement. On a

$$A \times B = \begin{pmatrix} A^1 B_1 & \cdots & A^1 B_p \\ \vdots & & \vdots \\ A^p B_1 & \cdots & A^p B_p \end{pmatrix} \in \mathcal{M}_p(\mathbf{k}) \quad \text{et} \quad B \times A = \begin{pmatrix} B^1 A_1 & \cdots & B^1 A_q \\ \vdots & & \vdots \\ B^q A_1 & \cdots & B^q A_q \end{pmatrix} \in \mathcal{M}_q(\mathbf{k}).$$

Ceci permet de calculer

$$\operatorname{Tr}(AB) = \sum_{i=1}^{p} A^i B_i = \sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij} b_{ji} = \sum_{j=1}^{q} \sum_{i=1}^{p} b_{ji} a_{ij} = \sum_{j=1}^{q} B^j A_j = \operatorname{Tr}(BA).$$

Si en revanche A et B sont toutes les deux de taille $p \times q$, on a

$$\operatorname{Tr}(A^\mathsf{T} B) = \sum_{i=1}^{p} \sum_{j=1}^{q} a_{ij} b_{ij},$$

qui donne la forme quadratique canonique sur $\mathcal{M}_{p,q}(\mathbf{k})$.

1.2 – Transvections, dilatations, permutations

Ces règles de calcul permettent d'interpréter agréablement les opérations élémentaires sur les lignes (resp. les colonnes) du pivot de Gauss comme des multiplications à gauche (resp. à droite) par des matrices inversibles particulières.

1.2.1 Définition. *Soit $\lambda \in \mathbf{k}$. Une* transvection élémentaire *(ou simplement transvection, si le contexte est clair) de rapport λ est une matrice de la forme*

$$T_{ij}(\lambda) = \mathrm{Id}_n + \lambda.E_{ij} \in \mathrm{GL}_n(\mathbf{k}) \quad (i \neq j)$$

où $E_{ij} = (\delta_{ij})_{1\leqslant i,j\leqslant n}$ est la matrice avec des 0 partout, sauf un 1 à l'entrée sur la i-ème ligne et la j-ème colonne.

Soit $\lambda \in \mathbf{k}^*$. Une dilatation élémentaire *(ou simplement dilatation, si le contexte est clair) de rapport λ est une matrice de la forme*

$$D_i(\lambda) = \mathrm{diag}(1, \ldots, \lambda, \ldots, 1) \in \mathrm{GL}_n(\mathbf{k}),$$

matrice diagonale avec des 1 partout sur la diagonale, sauf un λ en i-ème position.

Soit $\sigma \in \mathfrak{S}_n$. La matrice de permutation *associée à σ est*

$$P(\sigma) = (\delta_{i,\sigma(j)})_{1\leqslant i,j\leqslant n} \in \mathrm{GL}_n(\mathbf{k}).$$

Autrement dit, pour chaque j, la j-ème colonne a des 0 partout sauf un 1 en $\sigma(j)$-ème position.

1.2.2 Multiplication à gauche par une transvection (resp. dilatation, permutation). Soit $A \in \mathcal{M}_{n,p}(\mathbf{k})$. Pour tout $\lambda \in \mathbf{k}$ et $i \neq j$ dans $[\![1, n]\!]$, la matrice $A' = T_{ij}(\lambda) \times A$ est obtenue en effectuant l'opération élémentaire

$$L'_i = L_i + \lambda L_j$$

sur les lignes de A (on remplace la i-ème ligne par la somme de la i-ème ligne et de λ fois la j-ème ligne). En effet, il résulte des règles de calcul énoncées ci-dessus que pour $i' \neq i$ la i'-ème ligne de A' est

$$\begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix} \times A = L_{i'},$$

tandis que la i-ème ligne est

$$\begin{pmatrix} 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \end{pmatrix} \times A = L_i + \lambda L_j.$$

De la même façon, pour tout $\lambda \in \mathbf{k}^*$ et $i \in [\![1, n]\!]$ la matrice $A' = D_i(\lambda) \times A$ est obtenue en effectuant l'opération élémentaire

$$L'_i = \lambda L_i.$$

Enfin pour tout $\sigma \in \mathfrak{S}_n$ la matrice $A' = P(\sigma) \times A$ est obtenue en effectuant l'opération élémentaire

$$\forall i \in [\![1, n]\!] : \quad L'_i = L_{\sigma^{-1}(i)} \ ;$$

en effet, la i-ème ligne de $P(\sigma)$ est

$$\begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}$$

avec le 1 en $\sigma^{-1}(j)$-ème position.

1.2.3 Multiplication à droite. En appliquant les règles de multiplication transposées, on a de la même façon que les matrices $A \times T_{ij}(\lambda)$, $A \times D_i(\lambda)$ et $A \times P(\sigma)$ sont obtenues en effectuant les opérations élémentaires sur les colonnes de A : $C'_j = C_j + \lambda C_i$, $C'_i = \lambda C_i$, et $\forall : C_j = C_{\sigma(j)}$ respectivement.

1.2.4 Sous-groupes de $\mathrm{GL}_n(\mathbf{k})$. Les opérations élémentaires sur les lignes ou les colonnes introduites ci-dessus sont manifestement inversibles. On en déduit que les transvections, dilatations, et permutations sont des matrices inversibles comme nous l'avions annoncé sans preuve dans la définition. Précisément, on a

$$T_{ij}(\lambda)^{-1} = T_{ij}(-\lambda), \quad D_i(\lambda)^{-1} = D_i(\lambda^{-1}), \quad \text{et} \quad P(\sigma)^{-1} = P(\sigma^{-1})$$

pour tout $i \neq j$ dans $[\![1, n]\!]$. On vérifie de la même façon que les applications

$$\lambda \in (\mathbf{k}, +) \mapsto T_{ij}(\lambda) \in \mathrm{GL}_n, \quad \lambda \in (\mathbf{k}^*, \times) \mapsto D_i(\lambda) \in \mathrm{GL}_n, \quad \text{et} \quad \sigma \in \mathfrak{S}_n \mapsto P(\sigma) \in \mathrm{GL}_n$$

sont des morphismes de groupes pour tout $i \neq j$.

Plus généralement, pour tout $j \in [\![1, n+1]\!]$ l'application

$$T_j : (a_1, \ldots, a_n) \in \mathbf{k}^n \longmapsto \begin{pmatrix} 1 & & & a_1 & & & \\ & \ddots & & \vdots & & & \\ & & 1 & a_{j-1} & & & \\ & & & 1 & & & \\ & & & a_j & 1 & & \\ & & & \vdots & & \ddots & \\ & & & a_n & & & 1 \end{pmatrix} \in \mathcal{M}_{n+1}(\mathbf{k})$$

est un morphisme de groupes de $(\mathbf{k}^n, +)$ dans $\mathrm{GL}_{n+1}(\mathbf{k})$, ou mieux dans $\mathrm{PGL}_{n+1}(\mathbf{k})$. Multiplier à droite par la matrice $T_j(a_1, \ldots, a_n)$ revient à effectuer l'opération sur les colonnes

$$C'_j = C_j + a_1 C_1 + \cdots + a_{j-1} C_{j-1} + a_j C_{j+1} + \cdots a_n C_{n+1},$$

qui se décompose en n opérations élémentaires

$$C'_j = C_j + a_i C_i \quad (i < j), \quad \text{et} \quad C'_j = C_j + a_i C_{i+1} \quad (i \geqslant j)$$

qu'on peut effectuer dans l'ordre qu'on veut. On a donc

$$T_j(a_1, \ldots, a_n) = T_{j1}(a_1) \cdots T_{j,j-1}(a_{j-1})\, T_{j,j+1}(a_j) \cdots T_{j,n+1}(a_n),$$

et les n matrices de transvection du membre de droite commutent deux à deux.

Encore plus généralement (après on arrête, promis), l'application

$$(A, B) \in \mathrm{GL}_n(\mathbf{k}) \times \mathbf{k}^n \longmapsto \begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix} \in \mathcal{M}_{n+1}(\mathbf{k})$$

réalise un plongement du groupe affine $\mathrm{Aff}_n \cong (\mathrm{GL}_n(\mathbf{k}), \times) \ltimes (\mathbf{k}^n, +)$ dans le groupe $\mathrm{GL}_{n+1}(\mathbf{k})$, ou plutôt dans le groupe $\mathrm{PGL}_{n+1}(\mathbf{k})$. Le groupe affine se comprend très bien sans utiliser le produit semi-direct : par définition $(A, B) \in \mathrm{GL}_n(\mathbf{k}) \times \mathbf{k}^n$ agit sur l'espace affine $\mathbf{k}^n$ par "application linéaire et translation" :

$$\forall X \in \mathbf{k}^n : \ (A, B).X = AX + B.$$

Pour (A, B) et $(A', B') \in \mathrm{GL}_n(\mathbf{k}) \times \mathbf{k}^n$, on a donc

$$\forall X \in \mathbf{k}^n : \ (A, B).(A', B').X = A(A'X + B') + B = AA'X + AB' + B,$$

d'où la loi de groupe dans le groupe affine :

$$(A, B) \cdot (A', B') = (AA', AB' + B).$$

Maintenant le produit de matrices par blocs

$$\begin{pmatrix} A & B \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} A' & B' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} AA' & AB' + B \\ 0 & 1 \end{pmatrix}$$

permet de vérifier que l'application ci-dessus est effectivement un morphisme de groupes.

1.2.5 Théorème. *Soit $A \in \mathcal{M}_{pq}(\mathbf{k})$. La matrice A est équivalente dans $\mathcal{M}_{pq}(\mathbf{k})$ à la matrice $J_r = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0)$, où r est le rang de A.*

Dans l'énoncé ci-dessus, on s'autorise à parler de matrice diagonale même pour une matrice qui n'est pas carrée ; les coefficients de la matrice J_r sont donnés par

$$\forall i \in [\![1, p]\!], \forall j \in [\![1, q]\!] : (J_r)_{i,j} = \begin{cases} \delta_{ij} & \text{si } i \leqslant r \\ 0 & \text{si } i > r. \end{cases}$$

On va démontrer ce théorème en utilisant le pivot de Gauss, ce qui nous permettra d'en tirer des corollaires intéressants. Toutefois avant de se lancer, il est bon de se souvenir que ce résultat se démontre très simplement en adoptant un point de vue géométrique.

Preuve géométrique. Considérons l'application linéaire $f : X \in \mathbf{k}^q \mapsto AX \in \mathbf{k}^p$. Soit $(Y_1, \ldots, Y_r)$ une base de l'image de f, $r = \mathrm{rg}(A)$. On la complète par $(Y_{r+1}, \ldots, Y_p)$ en une base de $\mathbf{k}^p$. Pour tout $i = 1, \ldots, r$, il existe $X_i \in \mathbf{k}^q$ tel que $AX_i = Y_i$. La famille $(X_1, \ldots, X_r)$ est libre, puisque son image par f l'est ; on peut donc la compléter par $(X_{r+1}, \ldots, X_q)$ en une base de $\mathbf{k}^q$. La matrice de f dans les bases $(X_1, \ldots, X_q)$ et $(Y_1, \ldots, Y_p)$ est égale la matrice $J_r \in \mathcal{M}_{pq}(\mathbf{k})$.

Ainsi, si P et Q sont les matrices de passages des bases canoniques de $\mathbf{k}^p$ et $\mathbf{k}^q$ dans les bases $(Y_1, \ldots, Y_p)$ et $(X_1, \ldots, X_q)$ respectivement, on a $J_r = P^{-1}AQ$, ce qui prouve le résultat annoncé. □

Preuve par pivot de Gauss. On commence par démontrer par récurrence sur $\min(p, q)$ que la matrice A est équivalente à une matrice diagonale $\mathrm{diag}(d_1, \ldots, d_{\min(p,q)})$ égale à

$$\begin{pmatrix} d_1 & & & 0 & \ldots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & d_p & 0 & \ldots & 0 \end{pmatrix} \quad \text{ou} \quad \begin{pmatrix} d_1 & & \\ & \ddots & \\ & & d_q \\ 0 & \ldots & 0 \\ \vdots & & \vdots \\ 0 & \ldots & 0 \end{pmatrix}$$

selon que $p \leqslant q$ ou $q \leqslant p$.

Si $p = 0$ ou $q = 0$ le résultat est trivial. Supposons $\min(p, q) > 0$, et montrons que A est équivalente à une matrice

$$\begin{pmatrix} a & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & A' & \\ 0 & & & \end{pmatrix}$$

où $A' \in \mathcal{M}_{p-1,q-1}(\mathbf{k})$. Si tous les coefficients de A sur la première ligne et la première colonne sont nuls, A est elle-même du type cherché. Sinon, quitte à ajouter une ligne à la première ligne ou une colonne à la première colonne, (ce qui revient à remplacer A par $T_{1i}(1)A$ ou $AT_{j1}(1)$, respectivement), on peut supposer que le coefficient a_{11} est non nul. Alors on effectue les opérations élémentaires

$$\forall i > 1 : \ L'_i = L_i - \tfrac{a_{i1}}{a_{11}} L_1, \quad \text{puis} \quad \forall j > 1 : \ C'_j = C_j - \tfrac{a_{1j}}{a_{11}} C_1$$

pour obtenir une matrice du type annoncé ; autrement dit, la matrice

$$T_{1p}(-\tfrac{a_{p1}}{a_{11}}) \cdots T_{12}(-\tfrac{a_{21}}{a_{11}}) \times A \times T_{21}(-\tfrac{a_{12}}{a_{11}}) \cdots T_{1q}(-\tfrac{a_{1q}}{a_{11}})$$

est du type voulu. Elle est équivalente à A, puisque les matrices de transvection sont inversibles.

Par hypothèse de récurrence, la matrice A' est équivalente à une matrice diagonale : il existe donc P, Q inversibles de tailles $p-1$ et $q-1$ respectivement telles que $PA'Q$ soit diagonale. Alors

$$\begin{pmatrix} 1 & \\ & P \end{pmatrix} \begin{pmatrix} a & \\ & A' \end{pmatrix} \begin{pmatrix} 1 & \\ & Q \end{pmatrix}$$

est diagonale et équivalente à A, ce qui conclut notre récurrence.

Enfin pour arriver à la matrice J_r, on permute les lignes et les colonnes de notre matrice diagonale pour arriver à $\mathrm{diag}(d_1, \dots, d_{\min(p,q)})$ où $d_1, \dots, d_r$ sont non nuls et $d_{r+1}, \dots, d_{\min(p,q)}$ sont nuls ; comme précédemment, ces opérations ne font pas sortir de la classe d'équivalence de A. La matrice diagonale obtenue est de rang r, on a donc nécessairement $r = \mathrm{rg}(A)$ puisque le rang est constant sur les classes d'équivalence de matrices. On arrive ensuite à la matrice J_r en appliquant les bonnes dilatations :

$$D_1(d_1^{-1}) \cdots D_r(d_r^{-1}) \times \mathrm{diag}(d_1, \dots, d_r, 0, \dots, 0) = J_r.$$

□

1.2.6 Corollaire. *Le groupe* $\mathrm{GL}_n(\mathbf{k})$ *est engendré par les transvections élémentaires et les dilatations élémentaires. Le groupe* $\mathrm{SL}_n(\mathbf{k})$ *des matrices de déterminant* 1 *est engendré par les transvections élémentaires.*

Preuve. Soit $A \in \mathrm{GL}_n(\mathbf{k})$. En suivant la preuve par le pivot de Gauss du théorème 1.2.5, on voit qu'on a trouvé des matrices $T_1, \dots, T_N$ et $S_1, \dots, S_M$ qui sont toutes des transvections élémentaires, telles que

$$T_N \cdots T_1 \times A \times S_1 \cdots S_M = \mathrm{diag}(d_1, \dots, d_n).$$

Puisque A est inversible, les d_i sont nécessairement tous non nuls, donc

$$\mathrm{diag}(d_1, \dots, d_n) = D_1(d_1) \cdots D_n(d_n)$$

est un produit de dilatations élémentaires, et

$$A = T_1^{-1} \cdots T_N^{-1} \times D_1(d_1) \cdots D_n(d_n) \times S_M^{-1} \cdots S_1^{-1}$$

est bien un produit de transvections et dilatations élémentaires, puisque l'inverse d'une transvection élémentaire est une transvection élémentaire.

Si A est de déterminant 1, alors on a dans les mêmes notations $d_1 \cdots d_n = 1$. En appliquant le lemme 1.2.7 ci-dessous de manière répétée, d'abord aux lignes et colonnes numéros 1 et 2, puis aux lignes et colonnes numéros 2 et 3, et ainsi de suite jusqu'aux lignes et colonnes numéros $n-1$ et n (voir application 1.2.8), on trouve des matrices de transvection $T_{N+1}, \ldots, T_{N+2(n-1)}, S_{M+1}, \ldots, S_{M+2(n-2)}$ telles que

$$T_{N+2(n-1)} \cdots T_{N+1} \times \mathrm{diag}(d_1, \ldots, d_n) \times S_{M+1} \cdots S_{M+2(n-2)} = \mathrm{diag}(1, \ldots, 1, d_1 \cdots d_n) = \mathrm{Id}_n.$$

On a donc

$$T_{N+2(n-1)} \cdots T_{N+1} \times T_N \cdots T_1 \times A \times S_1 \cdots S_M \times S_{M+1} \cdots S_{M+2(n-2)} = \mathrm{Id}_n,$$

ce qui permet d'exprimer A comme un produit de transvections, puisque l'inverse d'une transvection est une transvection. □

1.2.7 Lemme. *Soit $a, b \in \mathbf{k}^*$. Il existe des matrices de transvections T_1, T_2, T_3, T_4 telles que*

$$T_4 \times T_2 \times \begin{pmatrix} a & \\ & b \end{pmatrix} \times T_1 \times T_3 = \begin{pmatrix} 1 & \\ & ab \end{pmatrix}.$$

Preuve. Il s'agit de trouver des opérations de transvection sur les lignes et les colonnes de $\mathrm{diag}(a, b)$ qui la transforment en $\mathrm{diag}(1, ab)$. On peut faire

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \xrightarrow{C_1' = C_1 + C_2} \begin{pmatrix} a & 0 \\ b & b \end{pmatrix} \xrightarrow{L_1' = L_1 + b^{-1}(1-a)L_2} \begin{pmatrix} 1 & 1-a \\ b & b \end{pmatrix} \xrightarrow{C_2' = C_2 + (a-1)C_1} \begin{pmatrix} 1 & 0 \\ b & ab \end{pmatrix}$$
$$\xrightarrow{L_2' = L_2 - bL_1} \begin{pmatrix} 1 & 0 \\ 0 & ab \end{pmatrix}.$$

On a donc

$$\begin{pmatrix} 1 & 0 \\ -b & 1 \end{pmatrix} \times \begin{pmatrix} 1 & b^{-1}(1-a) \\ 0 & 1 \end{pmatrix} \times \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \times \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \times \begin{pmatrix} 1 & a-1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & ab \end{pmatrix},$$

ce qui donne le résultat annoncé. □

1.2.8 Application. Dans la preuve du corollaire 1.2.6, on utilise le lemme pour transformer une matrice $\mathrm{diag}(d_1, \ldots, d_i, d_{i+1}, \ldots, d_n)$ en $\mathrm{diag}(d_1, \ldots, 1, d_i d_{i+1}, \ldots, d_n)$; on donne ici quelques détails sur cette opération. Le calcul ci-dessus nous dit qu'il existe quatre transvections $T_1, T_2, T_3, T_4 \in \mathrm{GL}_2(\mathbf{k})$ telles que

$$T_4 \times T_2 \times \mathrm{diag}(d_i, d_{i+1}) \times T_1 \times T_3 = \mathrm{diag}(1, d_i d_{i+1}).$$

Pour chaque $k = 1, \ldots, 4$, on considère la matrice diagonale par blocs

$$\tilde{T}_k = \begin{pmatrix} \mathrm{Id}_{i-1} & & \\ & T_k & \\ & & \mathrm{Id}_{n-i-1} \end{pmatrix},$$

qui est une transvection élémentaire de taille n. On vérifie alors qu'on a

$$\tilde{T}_4 \times \tilde{T}_2 \times \begin{pmatrix} * & & & \\ & d_i & & \\ & & d_{i+1} & \\ & & & * \end{pmatrix} \times \tilde{T}_1 \times \tilde{T}_3 = \begin{pmatrix} * & & & \\ & 1 & & \\ & & d_i d_{i+1} & \\ & & & * \end{pmatrix}.$$

1.2.9 Exemple. On donne ici une preuve indépendante du fait que les matrices de permutation sont des produits de transvections et dilatations élémentaires. Le groupe symétrique $\mathfrak{S}_n$ est engendré par les transpositions, donc toute matrice de permutation est produit de matrices de transposition, c'est-à-dire de matrices $P(\tau)$ où τ est une transposition (ij), $i \neq j$. Il suffit donc de montrer que toute matrice de transposition est produit de matrices de transvection et de dilatations.

Pour celà, on trouve un moyen astucieux (ou laborieux, selon les goûts) d'échanger deux lignes en ne faisant que des opérations de transvection et de dilatation. On peut faire

$$\begin{pmatrix} L_i \\ L_j \end{pmatrix} \to \begin{pmatrix} L_i + L_j \\ L_j \end{pmatrix} \to \begin{pmatrix} L_i + L_j \\ L_j - (L_i + L_j) \end{pmatrix} = \begin{pmatrix} L_i + L_j \\ -L_i \end{pmatrix} \to \begin{pmatrix} L_j \\ -L_i \end{pmatrix} \to \begin{pmatrix} L_j \\ L_i \end{pmatrix}$$

(pour $i \neq j$). Ceci prouve que

$$P(\tau) = D_j(-1) \times T_{ij}(1) \times T_{ji}(-1) \times T_{ij}(1),$$

pour $\tau = (ij)$, comme on voulait démontrer.

1.2.10 Une méthode magique pour inverser les matrices. Voici une recette permettant de calculer l'inverse d'une matrice $A \in \mathrm{GL}_n(\mathbf{k})$: effectuer les mêmes opérations élémentaires sur les lignes de A et de Id_n, jusqu'à avoir transformé A en l'identité ; la matrice obtenue en transformant Id_n est A^{-1}. Nous allons donner deux preuves du fait que cet algorithme fonctionne, confiants quant au fait que démonter les tours de magie de votre enfance ne diminuera pas votre émerveillement pour les mathématiques.

Preuve 1. Chaque opération élémentaire effectuant sur les lignes revient à une multiplication à gauche par une matrice de transvection, dilatation, ou permutation. On produit ainsi une suite de telles matrices $P_1, \ldots, P_N$ telles que

$$P_N \cdots P_1 \times A = \mathrm{Id}.$$

On a alors

$$A^{-1} = P_N \cdots P_1 = P_N \cdots P_1 \times \mathrm{Id},$$

cette dernière étant la matrice obtenue en effectuant la même suite d'opérations élémentaires sur les lignes de Id. □

Il y a une façon tristement plus élémentaire de comprendre cette méthode.

Preuve 2. L'inverse de A est la matrice B telle que $AX = Y \Leftrightarrow X = BY$ pour tout $X, Y \in \mathbf{k}^n$. On obtient donc l'inverse en résolvant le système

$$\begin{cases} a_{11}x_1 & + \cdots + & a_{1n}x_n & = & y_1 & & \\ \vdots & & \vdots & \vdots & & \ddots & \\ a_{n1}x_1 & + \cdots + & a_{nn}x_n & = & & & y_n \end{cases} \tag{1.2.10.1}$$

où $A = (a_{ij})_{1\leqslant i,j\leqslant n}$, les x_i sont les inconnues, et les y_j sont des variables. La solution de ce système donnera chaque x_i en fonction des variables y_j, sous la forme

$$\text{(1.2.10.2)} \qquad \begin{cases} x_1 & & = & b_{11}y_1 & +\cdots+ & b_{1n}y_n \\ \quad\ddots & & \vdots & \vdots & & \vdots \\ & x_n & = & b_{n1}y_1 & +\cdots+ & b_{nn}y_n, \end{cases}$$

et la matrice (b_{ij}) est l'inverse de A.

La méthode proposée consiste simplement à résoudre ce système en le présentant sous forme de tableau de nombre, en n'écrivant plus ni les inconnues ni les variables et en réservant une colonne pour chacune d'entre elles. Ainsi, (1.2.10.1) est représenté par le tableau

$$\text{(1.2.10.3)} \qquad \left(\begin{array}{ccc||ccc} a_{11} & \cdots & a_{1n} & 1 & \cdots & 0 \\ \vdots & & \vdots & & \ddots & \\ a_{n1} & \cdots & a_{nn} & 0 & \cdots & 1 \end{array}\right).$$

On résout le système par le pivot de Gauss, en opérant sur les lignes du système (1.2.10.1), ce qui se traduit par des opérations sur les lignes du tableau (1.2.10.3). À la fin de l'algorithme on arrive à la solution (1.2.10.2), dont la représentation sous forme d'un tableau de nombres est

$$\text{(1.2.10.4)} \qquad \left(\begin{array}{ccc||ccc} 1 & \cdots & 0 & b_{11} & \cdots & b_{1n} \\ & \ddots & & \vdots & & \vdots \\ 0 & \cdots & 1 & b_{n1} & \cdots & b_{nn} \end{array}\right).$$

La matrice (b_{ij}) a ainsi été obtenue en effectuant les opérations sur la partie droite du tableau, qui au départ est la matrice identité. □

1.3 – Décompositions LU

On appelle matrice *unitriangulaire* une matrice triangulaire dont tous les coefficients diagonaux sont égaux à 1. Pour toute matrice $A \in \mathcal{M}_n(\mathbf{k})$ et $p \in [\![1, n]\!]$, on note $A_{[p]}$ la sous-matrice de A obtenue en gardant les coefficients $a_{i,j}$ avec $i, j \in [\![1, p]\!]$. Les déterminants $\det A_{[p]}$ s'appellent les *mineurs principaux* de A.

1.3.1 Lemme. *Soit $L, U \in \mathcal{M}_n(\mathbf{k})$ des matrices unitriangulaires inférieure et supérieure respectivement, et $A \in \mathcal{M}_n(\mathbf{k})$. Alors les mineurs principaux de A et LAU sont égaux.*

Preuve. Il suffit de montrer l'assertion pour LA, la multiplication à droite par U se ramenant à ce cas par transposition. En terme d'opérations élémentaires, toute ligne de LA s'obtient comme somme de la ligne de A de même indice avec une combinaison linéaires des lignes d'indices inférieurs, ce qui donne le résultat. Formellement : Si $B = LA$, et $p \in [\![1, n]\!]$, nous avons

$$\begin{aligned} B_{[p]} &= L_{[1,p],[1,n]} \times A_{[1,n],[1,p]} \\ &= \begin{pmatrix} 1 & 0 & \dots & \dots & \dots & \dots & 0 \\ * & 1 & \ddots & & & & \vdots \\ \vdots & \ddots & \ddots & \ddots & & & \vdots \\ * & \dots & * & 1 & 0 & \dots & 0 \end{pmatrix} \times \begin{pmatrix} A^1_{[p]} \\ \hline \vdots \\ \hline A^n_{[p]} \end{pmatrix} \end{aligned}$$

où A^i désigne la i-ème ligne de A, et $A^i_{[p]}$ ses p premiers coefficients. En considérant le déterminant d'une matrice comme forme multilinéaire alternée sur ses lignes, on obtient

$$\begin{aligned}\det B_{[p]} &= \det\Big(A^1_{[p]}, \quad *A^1_{[p]} + A^2_{[p]}, \quad \dots, \quad *A^1_{[p]} + \cdots + *A^{p-1}_{[p]} + A^p_{[p]}\Big)\\ &= \det\Big(A^1_{[p]},\ A^2_{[p]},\ \dots,\ A^p_{[p]}\Big)\\ &= \det A_{[p]}.\end{aligned}$$

□

1.3.2 Théorème. *Soit $A \in \mathrm{GL}_n(\mathbf{k})$. Les deux assertions suivantes sont équivalentes :*

(i) *Pour tout $p = 1, \dots, n$, le mineur principal $\det A_{[p]}$ est non nul ;*

(ii) *Il existe une unique factorisation $A = LU$, avec L triangulaire supérieure, et U unitriangulaire supérieure.*

1.3.3 Remarque. Par un souci de symétrie on peut aussi énoncer l'assertion (ii) du théorème sous la forme équivalente suivante :

(ii′) Il existe une unique factorisation $A = LDU'$, avec L unitriangulaire supérieure, U' unitriangulaire supérieure, et D diagonale.

L'identité $U = DU'$ permet de faire le lien entre les deux versions.

Preuve. L'implication (ii′) ⇒ (i) découle directement du lemme 1.3.1, en remarquant que l'hypothèse A inversible implique que les coefficients diagonaux de D sont tous nous nuls, et donc également ses mineurs principaux.

Montrons à présent (i) ⇒ (ii), en commençant par la partie existence d'une factorisation LU. Le principe est d'appliquer le pivot de Gauss pour trouver L' unitriangulaire inférieure telle que $L'A$ soit triangulaire supérieure ; on commence par ajouter des multiples de la ligne A^1 aux lignes suivantes pour mettre des 0 sur la première colonne, puis on recommence avec la (nouvelle) ligne A^2 et la seconde colonne, etc. À chaque étape la condition sur les mineurs principaux nous assure qu'on a un coefficient non nul au bon endroit.

Pour formaliser cet argument, on procède par récurrence sur la taille de la matrice A, le cas d'une matrice de taille 1 étant clair. Dire que le premier mineur principal de $A = (a_{ij})$ est non nul revient à dire que le coefficient $a_{1,1}$ est non nul, on peut donc s'en servir comme pivot. A l'aide de $n-1$ opérations élémentaires sur les lignes on peut mettre à zéro tous les coefficients $a_{1,j}$, $j \geqslant 2$. Explicitement :

$$\begin{pmatrix} 1 & & & \\ -\frac{a_{21}}{a_{11}} & 1 & & \\ \vdots & & \ddots & \\ -\frac{a_{n1}}{a_{11}} & & & 1 \end{pmatrix} A = \begin{pmatrix} a_{1,1} & * & \dots & * \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{pmatrix} \tag{1.3.3.1}$$

pour une certaine matrice B de taille $(n-1) \times (n-1)$. Par le lemme 1.3.1, les mineurs principaux de B sont les mineurs principaux de A divisés par $a_{1,1}$, et sont donc tous

strictement positifs. Par hypothèse de récurrence on écrit $B = L_B U_B$, et on obtient

$$\begin{aligned} A &= \begin{pmatrix} 1 & & & \\ \frac{a_{21}}{a_{11}} & 1 & & \\ \vdots & & \ddots & \\ \frac{a_{n1}}{a_{11}} & & & 1 \end{pmatrix} \begin{pmatrix} a_{1,1} & * & \dots & * \\ 0 & & & \\ \vdots & & L_B U_B & \\ 0 & & & \end{pmatrix} \\ &= \begin{pmatrix} 1 & & & \\ \frac{a_{21}}{a_{11}} & 1 & & \\ \vdots & & \ddots & \\ \frac{a_{n1}}{a_{11}} & & & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & L_B & \\ 0 & & & \end{pmatrix} \begin{pmatrix} a_{1,1} & * & \dots & * \\ 0 & & & \\ \vdots & & U_B & \\ 0 & & & \end{pmatrix} \\ &= LU \end{aligned}$$

(voir 1.2.4 pour le calcul de l'inverse de la matrice la plus à gauche de (1.3.3.1)).

Finalement nous montrons l'unicité de la décomposition. Si $A = L_1U_1 = L_2U_2$, on écrit $J = L_2^{-1}L_1 = U_2U_1^{-1}$. Alors J est diagonale car à la fois triangulaire inférieure et supérieure, et égale à l'identité car $L_2^{-1}L_1$ est unipotente. On conclut $L_1 = L_2$ et $U_1 = U_2$. □

1.3.4 Remarque. La preuve précédente fournit un algorithme récursif pour calculer la décomposition LU : à chaque étape, il faut mémoriser la matrice

$$T = \begin{pmatrix} 1 & & & \\ -\frac{a_{21}}{a_{11}} & 1 & & \\ \vdots & & \ddots & \\ -\frac{a_{n1}}{a_{11}} & & & 1 \end{pmatrix},$$

calculer récursivement L_B, puis effectuer le produit $T^{-1} \times \mathrm{diag}(1, L_B)$ pour trouver L. En pratique il peut être plus judicieux de procéder de manière itérative, sans jamais rien avoir à garder en mémoire — à part bien sûr une matrice obtenue en transformant A dont la dernière version sera U, et une autre matrice qui peu à peu va devenir L. La preuve ci-dessus montre que $L = T_1^{-1}T_2^{-1}\cdots T_{n-1}^{-1}$, où chaque T_i est de la forme

$$T_i = \begin{pmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & & \\ & & \alpha_{i+1,i} & 1 & \\ & & \vdots & & \ddots \\ & & \alpha_{n,i} & & 1 \end{pmatrix},$$

et ainsi L s'obtient simplement en calculant successivement les produits $T_1^{-1}T_2^{-1}\cdots T_i^{-1}$ pour $i = 1, \dots, n-1$.

La version itérative de l'algorithme prend donc la forme suivante. On définit des suites (L_i) et (U_i) par récurrence, en posant

$$\begin{cases} U_0 = A \\ L_0 = \mathrm{Id}_n \end{cases} \quad \text{et} \quad \begin{cases} U_i = T_i U_{i-1} \\ L_i = L_{i-1} T_i^{-1} \end{cases} \quad \text{pour } i > 0.$$

Les matrices L et U qui nous intéressent sont respectivement L_{n-1} et U_{n-1}. Concrètement, la matrice U_{i-1} est échelonnée jusqu'à la $(i-1)$-ème colonne, c'est-à-dire de la forme

$$\begin{pmatrix} U & V \\ 0 & W \end{pmatrix}$$

avec U triangulaire supérieure de taille $i-1$; on obtient U_i en effectuant les opérations

$$R'_{i+s} = R_{i+s} - \tfrac{a_{i+s,i}}{a_{ii}} R_i, \qquad s = 1, \dots, n-i, \tag{1.3.4.1}$$

sur les lignes de U_{i-1} (on appelle ici R les lignes pour éviter les conflits de notation, et a_{st} les entrées de U_{i-1} pour éviter de surcharger les formules) pour mettre des 0 où il faut sur la i-ème colonne ; pour obtenir L_i il faut au fur et à mesure effectuer les opérations correspondantes

$$C'_i = C_i + \tfrac{a_{i+s,i}}{a_{ii}} C_{i+s} \qquad s = 1, \dots, n-i, \tag{1.3.4.2}$$

sur les colonnes de L_{i-1}, avec le même coefficient $\frac{a_{i+s,i}}{a_{ii}}$ dans (1.3.4.1) et (1.3.4.2) (mais '$-$' dans la première et '$+$' dans la seconde). On effectue des opérations sur les lignes dans un cas et sur les colonnes dans l'autre car on multiplie par T_i à gauche dans un cas, et par T_i^{-1} à droite dans l'autre.

1.3.5 Remarque. On peut donner une preuve brutale de l'unicité en donnant une formule pour chacun des coefficients des matrices L, D, U' en fonction de déterminants extraits de A. On a pour tout $i \geqslant j$:

$$d_i = \frac{|A_{[i]}|}{|A_{[i-1]}|} \; ; \quad \ell_{ij} = \frac{|A_{[j-1]\cup\{i\},[j]}|}{|A_{[j]}|} \text{ (pour } i \geqslant j) \; ; \quad u'_{ij} = \frac{|A_{[i],[i-1]\cup\{j\}}|}{|A_{[i]}|} \text{ (pour } j \geqslant i) \; ;$$

Pour obtenir ces formules, on applique la formule de Binet–Cauchy 2.2.5 à $A = LU$, avec $U = DU'$ comme plus haut. Les coefficients diagonaux de U sont alors ceux de D, qu'on note $d_1, \dots, d_n$. On a pour commencer

$$|A_{[i]}| = \sum_{Card(H)=i} |L_{[i],H}| \cdot |U_{H,[i]}| = |L_{[i]}| \cdot |U_{[i]}| = d_1 \cdots d_i,$$

qui donne aussitôt la formule pour d_i (dans la somme, seul le terme correspondant à $H = [p]$ est non nul, car L est triangulaire inférieure — ou car U est triangulaire supérieure). D'autre part,

$$|A_{[j-1]\cup\{i\},[j]}| = \sum_{Card(H)=j} |L_{[j-1]\cup\{i\},H}| \cdot |U_{H,[j]}| = |L_{[j-1]\cup\{i\},[j]}| \cdot |U_{[j]}| \qquad (\text{pour } i \geqslant j)$$

$$|A_{[i],[i-1]\cup\{j\}}| = \sum_{Card(H)=i} |L_{[i],H}| \cdot |U_{H,[i-1]\cup\{j\}}| = |L_{[i]}| \cdot |U_{[i],[i-1]\cup\{j\}}| \qquad (\text{pour } j \geqslant i)$$

(dans le premier cas on voit que la somme a un seul terme non nul car U est triangulaire supérieure, et dans le second car L est triangulaire inférieure).

Pour $i \geqslant j$, la matrice $L_{[j-1]\cup\{i\},[j]}$ est triangulaire inférieure avec coefficients diagonaux $1, \dots, 1, \ell_{ij}$ donc $|L_{[j-1]\cup\{i\},[j]}| = \ell_{ij}$; d'autre part la matrice $U_{[j]}$ est triangulaire supérieure de coefficients diagonaux $d_1, \dots, d_j$. On a donc finalement

$$|A_{[j-1]\cup\{i\},[j]}| = \ell_{ij} \cdot d_1 \cdots d_j = \ell_{ij} \cdot |A_{[i]}|$$

qui est la formule cherchée pour ℓ_{ij}.

De la même façon pour $j \geqslant i$, $U_{[i],[i-1]\cup\{j\}}$ est triangulaire supérieure de coefficients diagonaux $d_1, \dots, d_{i-1}, d_i \cdot u'_{ij}$, et $L_{[i]}$ est unitriangulaire inférieure, ce qui donne finalement

$$|A_{[i],[i-1]\cup\{j\}}| = d_1 \cdots d_{i-1} \cdot d_i \cdot u'_{ij} = |A_{[i]}|,$$

qui est la formule cherchée pour u'_{ij}.

Pour les matrices symétriques réelles, on a la spécialisation suivante de la décomposition *LDU* évoquée ci-dessus. L'équivalence (i) ⇔ (ii) s'appelle le critère de Sylvester.

1.3.6 Théorème (Décomposition de Cholesky)**.** *Soit S une matrique symétrique réelle de taille n. Les conditions suivantes sont équivalentes :*
(i) pour tout $p = 1, \dots, n$, le mineur principal $|S_{[1,p]}|$ est strictement positif ;
(ii) S est définie positive ;
(iii) il existe une unique factorisation $S = T^{\mathsf{T}} \times T$ avec T triangulaire supérieure à coefficients diagonaux strictement positifs.

Preuve. Montrons que (iii) ⇒ (ii). Pour tout $i = 1, \dots, n$, notons ℓ_i la forme linéaire donnée par la i-ème ligne de T : notant $T = (u_{ij})$, on a

$$\ell_i(x_1, \dots, x_n) = u_{ii}x_i + u_{i,i+1}x_{i+1} + \cdots + u_{in}x_n.$$

Puisque $u_{ii} \neq 0$ pour tout i, ces n formes linéaires sont linéairement indépendantes (de manière équivalente, T est inversible). Pour $X \in \mathbf{R}^n$, on a

$$X^{\mathsf{T}}SX = (TX)^{\mathsf{T}}(TX) = \sum\nolimits_{i=1}^{n} \ell_i(X)^2 \geqslant 0.$$

Si $X \neq 0$, puisque les ℓ_i forment une base de $\check{\mathbf{R}}^n$ il existe un i_0 tel que $\ell_{i_0}(X) \neq 0$, et ainsi $X^{\mathsf{T}}SX > 0$ comme il fallait démontrer.

Montrons réciproquement que (ii) ⇒ (iii). Il s'agit dans un premier temps de montrer que toute forme quadratique définie positive s'écrit comme somme de carrés de formes linéaires, ce que l'algorithme de Gauss permet de voir. On raisonne par récurrence sur n ; pour $n = 1$ l'énoncé est trivial. Notons $S = (a_{ij})$, et q la forme quadratique associée : pour $X = (x_1, \dots, x_n)^{\mathsf{T}} \in \mathbf{R}^n$ on a

$$q(x_1, \dots, x_n) = X^{\mathsf{T}}SX = \sum_{ij} a_{ij}x_ix_j = \sum_i a_iix_i^2 + 2\sum_{i<j} a_{ij}x_ix_j.$$

On a $a_{11} = q(1, 0, \dots, 0) > 0$ puisque q est définie positive par hypothèse ; on peut donc poser

$$\ell_1(x_1, \dots, x_n) = \frac{1}{\sqrt{a_{11}}}(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n),$$

de sorte que

$$\tilde{q}_1(x_1, \dots, x_n) := q(x_1, \dots, x_n) - \ell_1(x_1, \dots, x_n)^2$$

est une forme quadratique indépendante de x_1. Montrons que sa restriction q_1 à l'hyperplan $\{x_1 = 0\}$ est définie positive. Soit $(x_2, \dots, x_n) \in \mathbf{R}^{n-1}$ non nul. Puisque $a_{11} \neq 0$, il existe un unique $\tilde{x}_1 \in \mathbf{R}$ tel que $\ell_1(\tilde{x}_1, x_2, \dots, x_n) = 0$, et on a

$$q_1(x_2, \dots, x_n) = \tilde{q}_1(\tilde{x}_1, x_2, \dots, x_n) = q(\tilde{x}_1, x_2, \dots, x_n) > 0,$$

ce qui prouve bien que la restriction de q_1 est définie positive (la première égalité est donnée par le fait que q_1 est indépendante de x_1, et la seconde parce que $\ell_1(\tilde{x}_1, x_2, \dots, x_n) = 0$). On peut donc appliquer l'hypothèse de récurrence à q_1, ce qui donne l'existence d'une matrice triangulaire supérieure T_1 à coefficients diagonaux strictement positifs telle que

$$q_1(x_2, \dots, x_n) = \left[T_1 \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix} \right]^{\mathsf{T}} \times T_1 \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

Alors

$$q(x_1, x_2, \dots, x_n) = \ell_1(x_1, \dots, x_n)^2 + \left[T_1 \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix} \right]^{\mathsf{T}} \times T_1 \begin{pmatrix} x_2 \\ \vdots \\ x_n \end{pmatrix} = \left[T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \right]^{\mathsf{T}} \times T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

avec

$$T = \begin{pmatrix} \frac{1}{\sqrt{a_{11}}} & \frac{a_{12}}{\sqrt{a_{11}}} & \cdots & \frac{a_{1n}}{\sqrt{a_{11}}} \\ & & T_1 & \end{pmatrix}.$$

On a donc bien $S = T^{\mathsf{T}}T$ avec T du type demandé. Ainsi la proposition (ii) implique l'existence d'une décomposition $T^{\mathsf{T}}T$.

Montrons d'autre part qu'une telle écriture est toujours unique, indépendamment de l'hypothèse (ii). Soit T_1 et T_2 deux matrices triangulaires supérieures à coefficients diagonaux strictement positifs, telles que $S = T_1^{\mathsf{T}}T_1 = T_2^{\mathsf{T}}T_2$. Alors $(T_2^{\mathsf{T}})^{-1}T_1^{\mathsf{T}} = T_2(T_1)^{-1}$ est une matrice à la fois triangulaire supérieure et inférieure, donc diagonale. Notons $d_1, \dots, d_n$ et $d'_1, \dots, d'_n$ les coefficients diagonaux de T_1 et T_2 respectivement. Alors les coefficients diagonaux de $(T_2^{\mathsf{T}})^{-1}T_1^{\mathsf{T}}$ et $T_2(T_1)^{-1}$ sont respectivement $d_1/d'_1, \dots, d_n/d'_n$ et $d'_1/d_1, \dots, d'_n/d_n$, et on a donc

$$(T_2^{\mathsf{T}})^{-1}T_1^{\mathsf{T}} = \begin{pmatrix} d_1/d'_1 & & \\ & \ddots & \\ & & d_n/d'_n \end{pmatrix} = T_2(T_1)^{-1} = \begin{pmatrix} d'_1/d_1 & & \\ & \cdots & \\ & & d'_n/d_n \end{pmatrix}.$$

Ceci implique que $d_i^2 = (d'_i)^2$ pour tout i, et donc $d_i = d'_i$ puisque les coefficients diagonaux de T_1 et T_2 sont positifs. Finalement $T_2(T_1)^{-1} = \mathrm{Id}$ et $T_1 = T_2$ comme il fallait démontrer.

Montrons que (iii) $\Rightarrow$ (i). Soit T triangulaire supérieure inversible, et $p \in [\![1, n]\!]$. On s'intéresse aux mineurs principaux de $S = T^{\mathsf{T}} \times T$. En écrivant T par blocs

$$T = \begin{pmatrix} T_{[p]} & T' \\ 0 & T'' \end{pmatrix}$$

on voit que $S_{[p]} = T_{[p]}^{\mathsf{T}} \times T_{[p]}$, et donc $|S_{[p]}| = |T_{[p]}|^2 > 0$.

Enfin montrons que (i) $\Rightarrow$ (iii). Nous sommes dans les hypothèses du théorème 1.3.2, il existe donc L, D, U unitriangulaire inférieure, diagonale, et unitriangulaire supérieure respectivement telles que $S = LDU$. Puisque S est symétrique on a aussi $S = S^{\mathsf{T}} = U^{\mathsf{T}}DL^{\mathsf{T}}$, donc par unicité de la décomposition LDU on a $L = U^{\mathsf{T}}$. D'autre part les coefficients

diagonaux $d_1, \ldots, d_n$ de D sont les mineurs principaux de S, qui sont strictement positifs par hypothèse. On peut donc poser

$$T = \begin{pmatrix} \sqrt{d_1} & & \\ & \ddots & \\ & & \sqrt{d_n} \end{pmatrix} \times U,$$

qui est triangulaire supérieure, de coefficients diagonaux $\sqrt{d_1}, \ldots, \sqrt{d_n}$ strictement positifs, et telle que $S = T^{\mathsf{T}} \times T$. L'unicité d'une telle décomposition a déjà été démontrée plus haut. □

1.3.7 Remarque. On a donné ci-dessus une preuve indirecte du critère de Sylvester (i) ⇔ (ii), puisqu'on a montré indépendamment (i) ⇔ (iii) en utilisant la décomposition LU, et (ii) ⇔ (iii) en utilisant l'algorithme de Gauss pour les formes quadratiques. Voyons comment démontrer directement ce critère en supposant connue *a priori* l'orthogonalisation des formes quadratiques.

On commence par démontrer que le déterminant d'une matrice S définie positive est strictement positif. En orthogonalisant la forme quadratique associée à S, on obtient une matrice inversible P et une matrice diagonale D telles que $D = P^{\mathsf{T}}SP$. Puisque S est définie positive, les coefficients de D sont tous strictement positifs. Puisque $\det(S) \cdot \det(P)^2 = \det(D)$, on a donc que $\det(S) > 0$.

Montrons maintenant que (ii) ⇒ (i). Soit $p \in [\![1, n]\!]$. La matrice extraite $S_{[p]}$ est la matrice de la restriction au sous-espace engendré par les p premiers vecteurs de la base canonique de la forme quadratique définie par S ; elle donc elle aussi définie positive. En vertu de ce qui précède, on a donc $|S_{[p]}| > 0$ comme il fallait démontrer.

Réciproquement montrons que (i) ⇒ (ii), par récurrence sur n. Pour $n = 0$, le résultat est trivial. Pour $n > 0$, soit q la forme quadratique associée à S, et F_{n-1} le sous-espace de $\mathbf{R}^n$ engendré par les $n-1$ premiers vecteurs $e_1, \ldots, e_{n-1}$ de la base canonique. La restriction de q à F_{n-1} est donnée par la matrice extraite $S_{[n-1]}$, donc par hypothèse de récurrence elle est définie positive. Ceci prouve en particulier que F_{n-1} est totalement non-isotrope, ce qui implique que $F_{n-1} \oplus F_{n-1}^{\perp} = \mathbf{R}^n$. Soit e un vecteur non nul de $F_{n-1}^{\perp}$. La famille $(e_1, \ldots, e_{n-1}, e)$ est une base de $\mathbf{R}^n$, dans laquelle la matrice de q est

$$\begin{pmatrix} & & & 0 \\ & S_{[n-1]} & & \vdots \\ & & & 0 \\ 0 & \ldots & 0 & \alpha \end{pmatrix} = P^{\mathsf{T}}SP,$$

où P est la matrice de passage de la base canonique dans $(e_1, \ldots, e_{n-1}, e)$. On a donc

$$\alpha = \frac{\det(S) \cdot \det(P)^2}{\det(S_{[n-1]})} > 0,$$

et donc q est bien définie positive : pour $X \in \mathbf{R}^n$ non nul, de coordonnées $(x_1, \ldots, x_{n-1}, x)$ dans la base $(e_1, \ldots, e_{n-1}, e)$, on a

$$q(X) = q|_{F_{n-1}}(x_1, \ldots, x_{n-1}) + \alpha x^2 > 0$$

puisque $q|_{F_{n-1}}$ est définie positive et $\alpha > 0$.

Chapitre 2

Déterminant

2.1 – Polynôme déterminant, déterminant d'une matrice

2.1.1 Définition (polynôme déterminant)**.** *Soit $n \in \mathbf{N}^*$. On considère l'anneau $\mathbf{Z}[X_{ij}]$ des polynômes à coefficients entiers en les n^2 indéterminées X_{ij}, $1 \leqslant i, j \leqslant n$. Le* polynôme déterminant *de taille n est l'élément*

$$\det_n = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) X_{\sigma(1),1} \cdots X_{\sigma(n),n} \in \mathbf{Z}[X_{ij}].$$

On notera souvent simplement det au lieu de $\det_n$ par abus de notation.

2.1.2. Soit A un anneau commutatif unitaire. Le morphisme naturel $\mathbf{Z} \to A$ est défini par

$$n \in \mathbf{Z} \mapsto \underbrace{1_A + \cdots + 1_A}_{n \text{ fois}} \in A$$

(avec l'adaptation habituelle si $n < 0$). Ce morphisme s'étend en un morphisme $\mathbf{Z}[X_{ij}] \to A[X_{ij}]$.

2.1.3 Définition (déterminant d'une matrice)**.** *Soit A un anneau commutatif unitaire, et $M = (a_{ij}) \in \mathcal{M}_n(A)$ une matrice carrée de taille n à coefficients dans A. Le* déterminant de M *est l'élément* $\det(M) \in A$ *obtenu en évaluant le polynôme déterminant de taille n en les coefficients de M, autrement dit*

$$\det(M) := \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n} \in A.$$

2.2 – Interprétation comme forme multilinéaire alternée et applications

Le déterminant des matrices induit une forme sur l'espace des colonnes.

2.2.1 Proposition. *Soit A anneau commutatif. Le déterminant des matrices dans $\mathcal{M}_n(A)$ définit une forme n-linéaire alternée*

$$(A^n)^n \to A.$$

La bonne façon de calculer un déterminant c'est d'utiliser le pivot de Gauss. Explication par le caractère n-linéaire alterné de la formule pour le déterminant d'une matrice triangulaire : c'est le même calcul qu'en 2.6.6 !

2.2.2 action des opérations élémentaires.

2.2.3 déterminant d'une matrice triangulaire.

2.2.4 calcul du déterminant par pivot de Gauss. Ça fonctionne sur un anneau sans problème il me semble, mais à ce moment-là il faudra renvoyer au chapitre traitant le pivot de Gauss sur un anneau, qui est plus loin (et qui parle lui aussi de déterminants).

2.2.4.1 Exercice. Évaluer la complexité d'un tel calcul.

2.2.5 Proposition (Formule de Binet–Cauchy)**.** *Soit A et B des matrices à coefficients dans* $\mathbf{k}$ *de tailles $m \times n$ et $n \times p$ respectivement. Le produit AB est de taille $m \times p$. Soit I et J des sous-ensembles de même cardinal k de $[\![1, m]\!]$ et $[\![1, p]\!]$ respectivement. On a l'identité*

$$|(AB)_{IJ}| = \sum_{\substack{H \subseteq [\![1,n]\!] \\ \mathrm{Card}(H)=k}} |A_{IH}| \cdot |B_{HJ}|.$$

Preuve. En extrayant les lignes d'indices I de $A \times B$ de nos formules pour le produit de matrices, on obtient

$$(A \times B)_{I,[\![1,n]\!]} = A_{I,[\![1,n]\!]} \times B.$$

On ne regarde que les colonnes d'indices J, ce qui nous donne à nouveau grâce à nos formules pour le produit de matrices

$$(A \times B)_{IJ} = A_{I,[\![1,n]\!]} \times B_{[\![1,n]\!],J}.$$

Notons $X_1, \dots, X_n$ les colonnes (de taille k) de $A_{I,[\![1,n]\!]}$, et $Y_1, \dots, Y_k$ celles (de taille k aussi) de $(A \times B)_{IJ}$. D'après ce qui précède (et toujours nos formules pour le produit de matrices), on a pour tout $s = 1, \dots, k$

$$Y_s = b_{1j_s} X_1 + \cdots + b_{nj_s} X_n$$

où $J = \{j_1 < \cdots < j_k\}$.

En voyant le déterminant comme une forme sur les colonnes, on a donc

$$\begin{aligned}
|(A \times B)_{IJ}| &= \det(Y_1, \dots, Y_k) \\
&= \det\Big(\sum_{i=1}^{n} b_{ij_1} X_i, \dots, \sum_{i=1}^{n} b_{ij_k} X_i \Big) \\
&= \sum_{i_1, \dots, i_k} b_{i_1 j_1} \cdots b_{i_k j_k} \det(X_{i_1}, \dots, X_{i_k}) \\
&= \sum_{\substack{H \subseteq [\![1,n]\!] \\ H = \{i_1 < \cdots < i_k\}}} \sum_{\sigma \in \mathfrak{S}_k} b_{i_{\sigma(1)}, j_1} \cdots b_{i_{\sigma(k)}, j_k} \det(X_{i_{\sigma(1)}}, \dots, X_{i_{\sigma(k)}}) \\
&= \sum_{\substack{H \subseteq [\![1,n]\!] \\ H = \{i_1 < \cdots < i_k\}}} \sum_{\sigma \in \mathfrak{S}_k} b_{i_{\sigma(1)}, j_1} \cdots b_{i_{\sigma(k)}, j_k} \cdot \varepsilon(\sigma) \det(X_{i_1}, \dots, X_{i_k}) \\
&= \sum_{\substack{H \subseteq [\![1,n]\!] \\ H = \{i_1 < \cdots < i_k\}}} \det(X_{i_1}, \dots, X_{i_k}) \sum_{\sigma \in \mathfrak{S}_k} \varepsilon(\sigma) b_{i_{\sigma(1)}, j_1} \cdots b_{i_{\sigma(k)}, j_k}
\end{aligned}$$

qui est précisément l'identité recherchée. □

2.2.6 Corollaire. *Si A et B dont carrées de même taille, on a*

$$\det(AB) = \det(A)\det(B).$$

Nous proposons plusieurs preuves de cette formule : ici en conséquence de la formule de Binet–Cauchy, un peu plus loin en conséquence de la formule de Laplace ; ces deux preuves fonctionnent sans problème sur un anneau commutatif. Nous en donnons une troisième plus simple, qui fonctionne *a priori* seulement sur un corps ; il est toujours possible de s'en servir pour obtenir la formule en termes du polynôme déterminant , en utilisant le processus standard d'extension des identités algébriques.

2.3 – Développement de Laplace et applications

2.3.1 Théorème (Formule de Laplace)**.** *$A = (a_{ij})_{1\leqslant i,j\leqslant n}$, $J = \{j_1 < \cdots < j_p\}$ fixé :*

$$\det(A) = \sum_{I=\{i_1<\cdots<i_p\}} (-1)^{|I|+|J|} \det(A_{IJ}) \det(A_{\bar{I}\bar{J}}),$$

où $|I| = i_1 + \cdots + i_p$.

La formule de Laplace généralise la formule de développement par rapport à une colonne (celle-ci est le cas $p = 1$). Cette dernière peut se démontrer soit par un calcul direct sur le polynôme déterminant (c'est le plus agréable) ; sinon on peut la montrer directement, comme dans la preuve par récurrence sur n du théorème énoncé dans 2.6.1.

Je repousse la preuve de la formule de Laplace à la sous-section suivante 2.4, la lecture de la présente sous-section pouvant utilement préparer le lecteur à cette preuve. On peut là aussi procéder par un calcul direct sur le polynôme déterminant, la principale difficulté par rapport au cas $p = 1$ consistant alors en un calcul (un peu fin) de signature de permutation. Sinon on peut procéder par récurrence sur p, le cas de base étant alors le développement par rapport à une colonne. Une nouvelle fois il est possible de donner une preuve un peu plus simple en se limitant à des matrices à coefficients dans un corps, en tirant parti du théorème 2.6.1 ; voir 5.3.1.

2.3.2 Exemple. Développement d'une matrice 4×4 par rapport à deux colonnes.

2.3.3 Application 1 : déterminant des matrices triangulaires par blocs. *Soit $M \in \mathcal{M}_{p+q}(\mathrm{A})$ une matrice triangulaire supérieure par blocs,*

$$M = \begin{pmatrix} A & B' \\ 0 & B \end{pmatrix}$$

avec $A \in \mathcal{M}_p(\mathrm{A})$, $B \in \mathcal{M}_q(\mathrm{A})$, et $B' \in \mathcal{M}_{p,q}$. On a l'égalité $\det(M) = \det(A) \cdot \det(B)$.

2.3.3.1. Mise en garde dans le cas général = non-triangulaire (si on y réfléchit deux secondes, un tel résultat n'a pas lieu d'être car les blocs n'ont a priori aucune raison d'être carrés tous les quatre).

Preuve. On fait un développement de Laplace par rapport aux p premières colonnes de M (ou aux p premières lignes, cela revient au même). Dans les notations de 2.3.1, on choisit $J = [\![1, p]\!]$. Alors la matrice $p \times p$ extraite M_{IJ} possède une ligne nulle, sauf si $I = J = [\![1, p]\!]$. Il y a donc un seul terme non nul dans la somme de 2.3.1, qui est précisément $\det(A) \cdot \det(B)$. □

En fait la formule pour les déterminants triangulaires par blocs est un cas particulier plus simple de la formule de Laplace. On va en donner ci-dessous une preuve directe suivant les mêmes idées, dont la lecture pourra aider ensuite à suivre la preuve de la formule de Laplace.

Preuve de 2.3.3 sans formule de Laplace. On note $M = (m_{ij})$, $A = (a_{ij})$, $B = (b_{ij})$, $B' = (b'_{ij})$. Puisque $m_{ij} = 0$ si $j \leqslant p$ et $i > p$, on a

$$\det(M) = \sum_{\sigma([1,p])\subseteq[1,p]} \varepsilon(\sigma) \cdot m_{\sigma(1)1} \cdots m_{\sigma(p+q),p+q},$$

par quoi on veut dire qu'on somme sur toutes les permutations $\sigma \in \mathfrak{S}_n$ telles que $\sigma(\llbracket 1,p \rrbracket) \subseteq \llbracket 1,p \rrbracket$. Ces permutations sont exactement les

$$(\sigma', \sigma'') = \begin{pmatrix} 1 & \dots & p & p+1 & \dots & p+q \\ \sigma'(1) & \dots & \sigma'(p) & p+\sigma''(1) & \dots & p+\sigma''(q) \end{pmatrix},$$

où $\sigma' \in \mathfrak{S}_p$ et $\sigma'' \in \mathfrak{S}_q$. Ainsi on a

$$\det(M) = \sum_{\sigma'\in\mathfrak{S}_p} \sum_{\sigma''\in\mathfrak{S}_q} \varepsilon(\sigma', \sigma'') \cdot a_{\sigma'(1)1} \cdots a_{\sigma'(p)p} \cdot b_{\sigma''(1)1} \cdots b_{\sigma''(q)q}.$$

Il nous reste à démontrer que $\varepsilon(\sigma', \sigma'') = \varepsilon(\sigma') \cdot \varepsilon(\sigma'')$, ce qui donnera directement le résultat cherché. On écrit σ' et σ'' comme des produits de transpositions $\prod_{k=1}^{I'}(a'_k\ b'_k)$ et $\prod_{k=1}^{I''}(a''_k\ b''_k)$ respectivement ; alors $(\sigma', \sigma'') = \prod_{k=1}^{I'}(a'_k\ b'_k) \circ \prod_{k=1}^{I''}(p+a''_k\ p+b''_k)$, et donc

$$\varepsilon(\sigma', \sigma'') = (-1)^{I'+I''} = \varepsilon(\sigma') \cdot \varepsilon(\sigma'').$$

□

2.3.3.2 Variante. Pour calculer $\varepsilon(\sigma', \sigma'')$, on peut aussi calculer le cardinal de l'ensemble

$$\mathcal{I}(\sigma) = \{(i < j) \in \llbracket 1, p+q \rrbracket^2 : \sigma(i) > \sigma(j)\}$$

pour $\sigma = (\sigma', \sigma'')$. Pour $(i < j) \in \llbracket 1, p+q \rrbracket^2$ on a les trois configurations suivantes :
(a) $i, j \leqslant p$, et alors $(i < j) \in \mathcal{I}(\sigma', \sigma'')$ ssi $(i < j) \in \mathcal{I}(\sigma')$;
(b) $i \leqslant p$ et $j > p$, et alors on a toujours $(\sigma', \sigma'')(i) < (\sigma', \sigma'')(j)$;
(c) $i, j > p$, et alors $(i < j) \in \mathcal{I}(\sigma', \sigma'')$ ssi $(i < j) \in \mathcal{I}(\sigma'')$.
On a donc une bijection $\mathcal{I}(\sigma', \sigma'') \simeq \mathcal{I}(\sigma') \coprod \mathcal{I}(\sigma'')$, et par l'expression de la signature en termes de nombre d'inversion cette bijection nous donne

$$\varepsilon(\sigma', \sigma'') = (-1)^{\mathrm{Card}(\mathcal{I}(\sigma', \sigma''))} = (-1)^{\mathrm{Card}(\mathcal{I}(\sigma') + \mathrm{Card}(\mathcal{I}(\sigma''))} = \varepsilon(\sigma') \cdot \varepsilon(\sigma'').$$

2.3.4 Proposition. *Soit $M, N \in \mathcal{M}_n(A)$, A un anneau commutatif. On a* $\det(M \times N) = \det(M) \cdot \det(N)$.

Preuve. On note $M = (a_{ij})$ et $N = (b_{ij})$. Par la formule pour les déterminants triangulaires par blocs, on a

$$\det(M) \cdot \det(N) = \det \begin{pmatrix} M & 0 \\ -\mathrm{Id}_n & N \end{pmatrix},$$

où la matrice à droite est triangulaire par blocs avec tous ses blocs de taille n.

On effectue sur cette matrice par blocs les opérations élémentaires

$$L'_i = L_i + \sum_{1 \leqslant j \leqslant n} a_{ij} L_{n+j}$$

pour tout $i = 1, \ldots, n$. La matrice obtenue est

$$\begin{pmatrix} 0 & MN \\ -\mathrm{Id}_n & N \end{pmatrix}.$$

Elle aussi est triangulaire par blocs, et la formule pour les déterminants de ce type donne

$$\det \begin{pmatrix} 0 & MN \\ -\mathrm{Id}_n & N \end{pmatrix} = (-1)^n \det(-\mathrm{Id}_n) \cdot \det(MN) = \det(MN).$$

□

2.3.5 Différentielle du déterminant.

2.3.6 Irréductibilité du polynôme déterminant.

2.4 – Preuve de la formule de Laplace

2.4.1 Notations. Soit n un entier naturel. Pour tout $p \in [\![1, n]\!]$, on note $\mathcal{P}_p$ l'ensemble des parties à p éléments de $\{1, \ldots, n\}$. Pour $I = \{i_1 < \cdots < i_p\} \in \mathcal{P}_p$, on note $\bar{I}$ le complémentaire de I dans $[\![1, n]\!]$, $\bar{\imath}_1 < \cdots < \bar{\imath}_{n-p}$ les $n - p$ entiers tels que

$$\{i_1, \ldots, i_p, \bar{\imath}_1, \ldots, \bar{\imath}_{n-p}\} = \{1, \ldots, n\}.$$

Étant donné une matrice $A = (a_{ij})_{i \in I, j \in J}$, où I et J sont deux ensembles ordonnés finis, on note pour tout $I' \subseteq I$ et $J' \subseteq J$ $A_{I',J'}$ la matrice $(a_{ij})_{i \in I', j \in J'}$ extraite de A en retenant les lignes dont l'indice est dans I' et les colonnes dont l'indice est dans J'.

2.4.2 Proposition. *Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n}$, matrice carrée à coefficients dans* **k** *anneau commutatif (oui !). On fixe un ensemble de colonnes $J = \{j_1 < \cdots < j_p\}$. On a*

$$\det(A) = \sum_{I = \{i_1 < \cdots < i_p\}} (-1)^{|I| + |J|} \det(A_{IJ}) \det(A_{\bar{I}\bar{J}}),$$

où $|I| = i_1 + \cdots + i_p$.

On peut aussi noter $A^{I'J'}$ la matrice extraite de A obtenue en excluant les lignes dont l'indice est dans I' et les colonnes dont l'indice est dans J' ; autrement dit $A^{I'J'} = A_{\bar{I}\bar{J}}$.

2.4.3 Complément. On notera $\varepsilon(J, I) = (-1)^{|I| + |J|}$; c'est la signature de la permutation $\rho_{J,I}$ définie par

$$\begin{cases} \forall s \in [\![1, p]\!] : \rho_{J,I}(j_s) = i_s \\ \forall t \in [\![1, n-p]\!] : \rho_{J,I}(\bar{\jmath}_s) = \bar{\imath}_s. \end{cases}$$

La permutation $\rho_{J,I}$ est ainsi l'unique permutation induisant deux bijections croissantes de J sur I et de $\bar{J}$ sur $\bar{I}$ respectivement.

La preuve par calcul sur le polynôme déterminant est essentiellement semblable à celle du développement par rapport à une seule colonne, sauf qu'il y a un lemme sur la signature qui est un peu plus complexe.

2.4.4 Preuve par calcul sur le polynôme déterminant. On a

$$\det(A) = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma) a_{\sigma(1)1}\cdots a_{\sigma(n)n} = \sum_{\substack{I\subseteq[\![1,n]\!]\\ \mathrm{Card}(I)=p}} \sum_{\substack{\sigma\in\mathfrak{S}_n\\ \sigma(J)=I}} \varepsilon(\sigma) a_{\sigma(1)1}\cdots a_{\sigma(n)n}. \tag{2.4.4.1}$$

Soit I un sous-ensemble de $[\![1,n]\!]$ de cardinal p, et notons $[I]$ l'ensemble des permutations $\sigma\in\mathfrak{S}_n$ telles que $\sigma(I)=J$. Manifestement $[I]$ est en bijection avec le produit $\mathfrak{S}(J,I)\times\mathfrak{S}(\bar J,\bar I)$, où $\mathfrak{S}(\mathcal{E},\mathcal{E}')$ désigne l'ensemble des bijections entre deux ensembles $\mathcal{E}$ et $\mathcal{E}'$ de même cardinal. Puisque les ensembles J,I (resp. $\bar J,\bar I$) sont étiquetés de cardinal p (resp. $n-p$), on a donc une bijection Φ_J^I obtenue par la composition

$$\mathfrak{S}_p\times\mathfrak{S}_{n-p}\cong\mathfrak{S}(J,I)\times\mathfrak{S}(\bar J,\bar I)\ \overset{\Phi_J^I}{\cong}[I]\ ;$$

autrement dit pour tout $(\sigma',\sigma'')\in\mathfrak{S}_p\times\mathfrak{S}_{n-p}$, $\Phi_J^I(\sigma',\sigma'')$ est la permutation φ définie par les deux conditions

$$\forall s=1,\dots,p:\ \varphi(j_s)=i_{\sigma'(s)}\quad\text{et}\quad\forall t=1,\dots,n-p:\ \varphi(\bar\jmath_t)=\bar\imath_{\sigma''(t)}.$$

Avec cette description pour I, on a

$$\det(A)=\sum_I\sum_{\substack{\sigma'\in\mathfrak{S}_p\\ \sigma''\in\mathfrak{S}_{n-p}}} \varepsilon\big(\Phi_J^I(\sigma',\sigma'')\big)\cdot a_{\sigma'(j_1)j_1}\cdots a_{\sigma'(j_p)j_p}\cdot a_{\sigma''(\bar\jmath_1)\bar\jmath_1}\cdots a_{\sigma''(\bar\jmath_p)\bar\jmath_p}. \tag{2.4.4.2}$$

Tenant compte du Lemme 2.4.6 donnant la signature $\varepsilon\big(\Phi_J^I(\sigma,\sigma')\big)$, on obtient finalement

$$\begin{aligned}&\det(A)=\\ &\sum_I(-1)^{|I|+|J|}\Big(\sum_{\sigma'\in\mathfrak{S}_p}\varepsilon(\sigma')a_{i_{\sigma'(1)}j_1}\cdots a_{i_{\sigma'(p)}j_p}\Big)\Big(\sum_{\sigma''\in\mathfrak{S}_{n-p}}\varepsilon(\sigma'')a_{\bar\imath_{\sigma''(1)}\bar\jmath_1}\cdots a_{\bar\imath_{\sigma''(n-p)}\bar\jmath_{n-p}}\Big)\\ &=\sum_I(-1)^{|I|+|J|}\det(A_{IJ})\cdot\det(A_{\bar I\bar J}).\end{aligned}$$

comme il fallait démontrer. □

2.4.5 Remarque. Le point clef de la preuve est la partition $\mathfrak{S}_n=\coprod_I[I]$, qui n'est autre que la partition du groupe $\mathfrak{S}_n$ par ses classes de congruences modulo le sous-groupe

$$[J]=\{\sigma\in\mathfrak{S}_n:\sigma(J)=J\}.$$

Le quotient $\mathfrak{S}_n/[J]$ s'identifie à l'ensemble des sous-parties de cardinal p de $[\![1,n]\!]$; il n'est en général pas muni d'une structure canonique de groupe, puisque $[J]$ n'est pas un sous-groupe distingué. Nous recommandons de déterminer les cardinaux de $[J]$ et $\mathfrak{S}_n/[J]$, et de vérifier le théorème de Lagrange dans ce cas.

On prendra garde au fait que $[I]$ n'est pas un sous-groupe de $\mathfrak{S}_n$, et que Φ_J^I n'est pas un morphisme de groupes.

2.4.6 Lemme. *On a* $\varepsilon\big(\Phi_J^I(\sigma',\sigma'')\big)=(-1)^{|I|+|J|}\varepsilon(\sigma')\varepsilon(\sigma'')$.

Preuve. A tout $(\sigma', \sigma'') \in \mathfrak{S}_p \times \mathfrak{S}_{n-p}$, on associe une permutation $\Psi(\sigma', \sigma'') \in \mathfrak{S}_n$ définie par

$$\Psi(\sigma', \sigma'') = \begin{pmatrix} 1 & \cdots & p & p+1 & \cdots & p+(n-p) \\ \sigma'(1) & \cdots & \sigma'(p) & p+\sigma''(1) & \cdots & p+\sigma''(n-p) \end{pmatrix}.$$

Pour tout $I \subseteq [\![1, n]\!]$ de cardinal p, on définit une permutation

$$\rho_I = \begin{pmatrix} 1 & \cdots & p & p+1 & \cdots & p+(n-p) \\ i_1 & \cdots & i_p & \bar{\imath}_1 & \cdots & \bar{\imath}_{n-p} \end{pmatrix} \in \mathfrak{S}_n.$$

Je prétends que

$$\Phi_J^I(\sigma', \sigma'') = \rho_I \circ \Psi(\sigma', \sigma'') \circ \rho_J^{-1}, \tag{2.4.6.1}$$

et je laisse au lecteur le soin de le démontrer.

On va maintenant pouvoir montrer l'identité voulue sur les signatures en calculant des nombres d'inversions. D'une part, $\varepsilon(\Psi(\sigma', \sigma'')) = \varepsilon(\sigma')\varepsilon(\sigma'')$ car

$$\mathcal{I}(\Psi(\sigma', \sigma'')) = (\mathcal{I}(\sigma')) \coprod ((p, p) + \mathcal{I}(\sigma''))$$

($\mathcal{I}$ désigne l'ensemble des inversions, dont il faut calculer le cardinal pour définir le nombre d'inversions, voir [**?**]). D'autre part,

$$\begin{aligned} \mathcal{I}(\rho_I) &= \{(r, s') \in [\![1, p]\!] \times [\![p+1, n]\!] : i_r > \bar{\imath}_{s'-p}\} \\ &\simeq \coprod_{1 \leqslant r \leqslant p} \{s \in [\![1, n-p]\!] : \bar{\imath}_s < i_r\} \\ &\simeq \coprod_{1 \leqslant r \leqslant p} \bar{I} \cap [\![1, i_r - 1]\!] \\ &\simeq \coprod_{1 \leqslant r \leqslant p} [\![1, i_r - 1]\!] \setminus \{i_1, \ldots, i_{r-1}\} \end{aligned}$$

donc

$$I(\rho_I) = \sum_{1 \leqslant r \leqslant p} (i_r - 1 - (r-1)) = \sum_{1 \leqslant r \leqslant p} i_r - \sum_{1 \leqslant r \leqslant p} r = |I| - \frac{p(p+1)}{2}.$$

□

2.4.6.1 Variante. On peut aussi calculer ces signatures en termes de produits de transposition. Pour $\Psi(\sigma', \sigma'')$ c'est assez transparent, on le fait pour ρ_I. La permutation

$$\begin{pmatrix} 1 & 2 & \ldots & \ldots & \ldots & n \\ i_1 & 1 & \ldots & \hat{i}_1 & \ldots & n \end{pmatrix}$$

(qui place i_1 en tête puis laisse $[\![1, n]\!] - \{i_1\}$ dans l'ordre croissant ; comme d'habitude le chapeau signifie qu'il faut omettre i_1) est un produit de $i_1 - 1$ transpositions, puisque c'est le i_1-cycle $[i_1, i_1 - 1, \ldots, 2, 1]$. Ensuite pour "mettre i_2 après i_1" il faut $i_2 - 2$ nouvelles transpositions : la permutation

$$\begin{pmatrix} 1 & 2 & 3 & \ldots & \ldots & \ldots & \ldots & n \\ i_1 & i_2 & 1 & \ldots & \hat{i}_1 & \hat{i}_2 & \ldots & n \end{pmatrix}$$

s'obtient en composant la précédente avec le (i_2-1)-cycle $[i_2,\dots,\hat{i}_1,\dots,2,1]$. On raisonnant ainsi de proche en proche, on obtient l'égalité

$$\rho_I = [i_p,\dots,\hat{i}_{p-1},\dots,\hat{i}_1,\dots,2,1]\circ[i_2,\dots,\hat{i}_1,\dots,2,1]\circ[i_1,\dots,2,1],$$

qui permet d'écrire ρ_I comme un produit de $\sum_{k=1}^{p}(i_k-k)$ transpositions ; on retrouve bien la signature de ρ_I.

2.4.7. *Preuve du complément 2.4.3.* Il apparaît dans la preuve ci-dessus que $\varepsilon(J,I)$ est la signature de la permutation $\rho_I\circ\rho_J^{-1}$. Le résultat suit, en observant que $\rho_I\circ\rho_J^{-1}=\rho_{J,I}$. □

2.4.8 Preuve par récurrence sur p. On laisse l'initialisation de la récurrence au lecteur. Supposons par récurrence la formule connue pour $p'\leqslant p-1$ colonnes. On commence par développer par rapport à la colonne j_1 :

$$\det(A)=\sum_{1\leqslant i\leqslant n}(-1)^{i+j_1}a_{i,j_1}\det(A_{\hat{\imath},\hat{\jmath}_1}).$$

Ensuite on développe chacun des déterminants de taille $n-1$ par rapports aux colonnes correspondant aux colonnes $j_2,\dots,j_p$ de A. Attention, il y a un décalage de numérotation, elles sont devenues les colonnes numéros $j_2-1,\dots,j_p-1$ de la matrice $A_{\hat{\imath},\hat{\jmath}_1}$. On a un décalage du même type dans la numérotation des lignes ; pour $I'\in[\![1,n]\!]^{p-1}$ ne contenant pas i, on note $s(i,I')$ le nombre de $i'\in I$ qui sont $>i$ (c'est le nombre de lignes dont l'indice va être décalé par la suppression de la ligne i). On obtient

$$\det(A_{\hat{\imath},\hat{\jmath}_1})=\sum_{I'\not\ni i}(-1)^{|I'|-s(i,I')+|J'|-(p-1)}\det(A^{\hat{\imath},\hat{\jmath}_1}_{I'J'})\det(A_{\bar{I}\bar{J}}),$$

où on a posé $J'=\{j_2<\cdots j_p\}$, et $A^{\hat{\imath},\hat{\jmath}_1}_{I'J'}$ est la matrice de taille $p-1$ obtenue en extrayant de $A_{\hat{\imath},\hat{\jmath}_1}$ les lignes I' et les colonnes J' de A.

Reste à recombiner ces déterminants $p-1$ entre eux, à nouveau avec la formule de développement par rapport à une colonne, utilisée à l'envers. Le point sur lequel il faut se fixer est que chaque $I=\{i_1<\cdots<i_p\}$ apparaît sous les p formes (i_r,I'_r) avec $I'_r=I-\{i_r\}$, $r=1,\dots,p$; notons au passage que $s(i_r,I'_r)=p-r$.

$$\begin{aligned}\det(A)&=\sum_i\sum_{I'\not\ni i}(-1)^{i+j_1+|I'|-s(i,I')+|J'|-(p-1)}a_{i,j_1}\det(A^{\hat{\imath},\hat{\jmath}_1}_{I'J'})\det(A_{\bar{I}\bar{J}})\\&=\sum_I(-1)^{|I|+|J|}\left(\sum_{r=1}^{p}(-1)^{-s(i_r,I'_r)-p+1}a_{i_1,j_1}\det(A^{\hat{\imath}_1,\hat{\jmath}_1}_{I'J'})\right)\det(A_{\bar{I}\bar{J}}).\end{aligned}$$

Le terme entre parenthèses est le développement de $\det(A_{IJ})$ par rapport à sa première ligne, puisque $(-1)^{-s(i_r,I'_r)-p+1}=(-1)^{r+1}$ (voir ci-dessous). □

2.4.8.1 Remarque. On a utilisé la formule diabolique

$$\forall\varepsilon_1,\dots,\varepsilon_r=\pm1:\quad(-1)^{\varepsilon_1a_1+\cdots+\varepsilon_ra_r}=(-1)^{a_1+\cdots+a_r},$$

qui vaut en vertu de la non moins diabolique identité $(-1)^{-1}=-1$.

2.5 – Caractérisation de l'inversibilité et calcul du rang

2.5.1 Proposition (Formules de Cramer)**.** *Soit R un anneau commutatif, et $A=(a_{ij})\in\mathcal{M}_n(R)$. On considère sa comatrice* $\mathrm{Com}(A)=((-1)^{i+j}A_{ij})$. *On a*

$$A\times\mathrm{Com}(A)^{\mathsf{T}}=\mathrm{Com}(A)^{\mathsf{T}}\times A=\det(A).\mathrm{Id}.$$

Preuve. On utilise la formule pour développer un déterminant par rapport à une ligne (resp. colonne) pour calculer $A \times \mathrm{Com}(A)^{\mathsf{T}}$ (resp. $\mathrm{Com}(A)^{\mathsf{T}} \times A$).

Le coefficients d'indice i, j de $A \times \mathrm{Com}(A)^{\mathsf{T}}$ est

$$\sum_{k=1}^{n} a_{ik}(-1)^{k+j} A_{jk} = \det \left(\begin{array}{c} L_1 \\ \hline \vdots \\ L_{j-1} \\ \hline L_i \\ \hline L_{j+1} \\ \vdots \\ L_n \end{array} \right) = \delta_{ij} \det(A)$$

(ici L_i est la i-ème ligne de A, et la matrice ci-dessus est obtenue à partir de A en mettant la i-ème ligne à la place de la j-ème). De la même façon, Le coefficients d'indice i, j de $\mathrm{Com}(A)^{\mathsf{T}} \times A$ est

$$\sum_{k=1}^{n} (-1)^{k+i} A_{ki} a_{kj} = \det \left(\begin{array}{c|c|c|c|c|c|c} C_1 & \cdots & C_{i-1} & C_j & C_{i+1} & \cdots & C_n \end{array} \right) = \delta_{ij}. \det(A).$$

□

2.5.2 Corollaire. *Soit A un anneau commutatif, et $M \in \mathcal{M}_n(A)$ une matrice carrée de taille n. La matrice M est inversible si et seulement si son déterminant est un élément inversible de A.*

Preuve. Supposons $MN = \mathrm{Id}$. Alors par 2.6.13 on a $\det(M) \cdot \det(N) = 1$ donc $\det(M)$ est inversible. Réciproquement, si $\det(M)$ est inversible on pose $M' = \det(M)^{-1}.\mathrm{Com}(M)$. On a

$$M \times M' = M' \times M = \det(M)^{-1} \det(M).\mathrm{Id} = \mathrm{Id}$$

par les formules de Cramer, donc M est inversible. □

2.5.3 Exemple. Une matrice $M \in \mathcal{M}_n(\mathbf{Z})$ (resp. $\mathcal{M}_n(\mathbf{k}[X])$) est inversible si et seulement si $\det(M) = \pm 1$ (resp. $\det(M) \in \mathbf{k}^*$).

2.5.4 Corollaire. *On considère R un anneau commutatif arbitraire. Soit $A, B \in \mathcal{M}_n(R)$. Les propositions suivantes sont équivalentes :*

(i) $AB = BA = \mathrm{Id}$ *;*
(ii) $AB = \mathrm{Id}$ *;*
(iii) $BA = \mathrm{Id}$.

Si R est un corps, on peut utiliser la théorie de la dimension pour démontrer cet énoncé. Nous encourageons le lecteur à tâcher de se remémorer comment faire. Voici une preuve directe basée sur les formules de Cramer.

Preuve. Supposons $AB = \mathrm{Id}$. Alors $\det(A) \cdot \det(B) = 1$, donc en posant $A' = \det(B).\mathrm{Com}(A)$, on a

$$A \times A' = A' \times A = \det(B) \det(A).\mathrm{Id} = \mathrm{Id}$$

par les formules de Cramer. Ainsi A est inversible. Ensuite $AB = \mathrm{Id}$ implique (en multipliant à gauche par A^{-1}) que $B = A^{-1}$. Ceci prouve (ii) ⇒ (i). L'implication (iii) ⇒ (i) se démontre de manière analogue. □

2.5.5 Exercice. Déterminer le rang de la comatrice $\mathrm{Com}(A)$ en fonction du rang de A.

2.5.6 Proposition. *Le rang d'une matrice est la taille de son plus grand mineur non-nul.*

Preuve. On va démontrer l'énoncé équivalent :

$$\forall r \in \mathbf{N} : \quad \text{rang} < r \Leftrightarrow \text{tous les mineurs de taille } r \text{ sont nuls.}$$

Montrons $\Rightarrow$. Soit $A \in \mathcal{M}_{m,n}$ une matrice de rang $< r$. Pour tout $1 \leqslant j_1 < \cdots < j_r \leqslant n$, il existe une relation de dépendance linéaire entre les colonnes de A d'indices $j_1, \ldots, j_r$. Celle-ci induit pour tout $1 \leqslant i_1 < \cdots < i_r \leqslant n$ une relation de dépendance linéaire entre les colonnes de la matrice extraite A_{IJ}, et on a donc $\det(A_{IJ}) = 0$, comme il fallait.

Montrons maintenant $\Leftarrow$ par contraposée. Soit donc $A \in \mathcal{M}_{m,n}$ une matrice de rang $\geqslant r$, et montrons qu'elle possède un mineur de taille r non-nul. D'après le théorème de la base extraite il existe r colonnes de A qui sont linéairement indépendantes. Je choisis r telles colonnes, et je les complète en une base de $\mathbf{k}^m$, en invoquant le théorème de la base incomplète. J'ai alors sous les yeux m colonnes dans $\mathbf{k}^m$, et je considère la matrice $\tilde{A} \in \mathcal{M}_m(\mathbf{k})$ obtenue en les concaténant. Par construction, on a $\det(A) \neq 0$. D'après la formule de Laplace, ce déterminant est une combinaison linéaire de déterminants de taille r extraits des r premières colonnes de $\tilde{A}$, et ainsi une combinaison linéaire de mineurs de taille r de A. Puisque $\det(A) \neq 0$, il est nécessaire que ces mineurs ne soient pas tous nuls. □

2.5.7 Remarque. Si on sait déjà que le rang d'une matrice est égal au rang de sa transposée (voir 4.4.4 pour deux arguments n'utilisant pas le déterminant), on peut argumenter de la façon suivante. Si les colonnes de A forment une famille de rang r, on peut trouver r colonnes $C_{j_1}, \ldots, C_{j_r}$ linéairement indépendantes ; en ne gardant que celles-ci, on obtient une matrice extraite de taille $m \times r$ et de rang r. Le rang de cette matrice étant égal au rang de sa transposée, ses lignes forment une famille de rang r, donc on peut trouver r d'entre elles $L_{i_1}, \ldots, L_{i_r}$ qui sont linéairement indépendantes. En ne gardant que celles-ci, on trouve la matrice extraite A_{IJ} de A ($I = \{i_1, \ldots, i_r\}$, $J = \{j_1, \ldots, j_r\}$), de taille $r \times r$ et de rang r. On a donc $\det(A_{IJ}) \neq 0$ est qui un mineur non nul de taille r.

D'autre part si A est de rang r, pour tout $r' > r$, toute famille de r' colonnes de A est liée, et donc tout mineur de taille r' de A est nul.

On peut au contraire faire le choix d'utiliser la caractérisation du rang par les mineurs pour démontrer l'invariance du rang par l'opération de transposition.

2.5.8 Application : invariance du rang par extension de corps. Soit $A \in \mathcal{M}_{mn}(\mathbf{k})$. Pour toute extension de corps $\mathbf{k}'/\mathbf{k}$, on peut voir A comme une matrice à coefficients dans $\mathbf{k}'$; notons ici A' cette matérialisation de A. Le rang de A est la dimension du sev de $\mathbf{k}^m$ engendré par les colonnes de A, celui de A' la dimension du sev de $(\mathbf{k}')^m$ engendré par ces mêmes colonnes. *A priori* il n'y a rien d'évident à ce que ces deux rangs coïncident.

Il est cependant bien vrai que $\mathrm{rg}(A) = \mathrm{rg}(A')$, et ceci se voit bien avec l'énoncé 2.5.6.

2.6 – Formes multilinéaires alternées et volumes

2.6.1 Théorème. *Les formes n-linéaires alternées sur E constituent un $\mathbf{k}$-ev de dimension* 1. *Soit $\mathcal{B}$ une base. Définition :* $\det_{\mathcal{B}}$ *est l'unique forme n-linéaire alternée f telle que $f(\mathcal{B}) = 1$.*

Preuve. Il y a plusieurs façons de démontrer le fait que toutes les formes n-linéaires alternées sur E sont proportionnelles.

Une première voie est de montrer que toute forme n-linéaire alternée est proportionnelle au polynôme déterminant. Ça se fait par un calcul assez naturel : soit $(e_1, \dots, e_n)$ une base, f n-linéaire alternée. On a

$$\begin{aligned} f\Big(\sum_i a_{i,1}e_i,\ \dots,\ \sum_i a_{i,n}e_i\Big) &= \sum_{(i_1,\dots,i_n)} a_{i_1,1}\cdots a_{i_n,n}\cdot f(e_{i_1},\dots,e_{i_n}) \\ &= \sum_{\sigma\in\mathfrak{S}_n} a_{\sigma(1),1}\cdots a_{\sigma(n),n}\cdot f(e_{\sigma(1)},\dots,e_{\sigma(n)}) \\ &= \sum_{\sigma\in\mathfrak{S}_n} a_{\sigma(1),1}\cdots a_{\sigma(n),n}\cdot \varepsilon(\sigma)\cdot f(e_1,\dots,e_n) \end{aligned}$$

(la première égalité est un développement par n-linéarité ; pour la seconde, on observe que par le caractère alterné seuls les n-uplets $(i_1, \dots, i_n)$ avec les i_s deux à deux distincts contribuent non trivialement, et ces n-uplets sont en bijection avec $\mathfrak{S}_n$; pour la troisième on utilise à nouveau le caractère alterné pour réordonner les arguments de f).

Une autre façon de faire est de procéder par récurrence sur n, en démontrant le caractère nécessaire de la formule de développement par rapport à une colonne. C'est techniquement moins agréable que la première approche. □

2.6.2 Proposition (). *Soit* $\mathbf{k}$ *un corps, et* $u\colon \mathbf{k}^n \to \mathbf{k}$ *une forme* n*-linéaire alternée. Alors* $u = \lambda \det$ *pour un certain* $\lambda \in \mathbf{k}$.

Preuve. Soit (e_i) la base canonique de K^n. On va montrer que $\lambda = u(e_1, \dots, e_n)$ convient. Par n-linéarité, pour tout choix de n vecteurs $v_j = \sum_i a_{ij}e_i$, on a

$$u(v_1, \dots, v_n) = \sum_{(j_i)} a_{1j_1} \dots a_{nj_n} u(e_{j_1}, \dots, e_{j_n}).$$

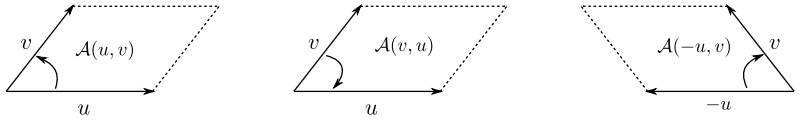
Si l'application $i \mapsto j_i$ est non injective, alors $u(e_{j_1}, \dots, e_{j_n}) = 0$. Donc seules les bijections contribuent à la somme ci-dessus, et on peut donc la reécrire en indiçant par les éléments du groupe symétrique S_n, ce qui fait apparaître la formule du déterminant :

$$\begin{aligned} u(v_1, \dots, v_n) &= \sum_{\sigma\in S_n} a_{1\sigma(1)} \dots a_{n\sigma(n)} u(e_{\sigma(1)}, \dots, e_{\sigma(n)}) \\ &= \sum_{\sigma\in S_n} (-1)^n a_{1\sigma(1)} \dots a_{n\sigma(n)} u(e_1, \dots, e_n) \\ &= u(e_1, \dots, e_n) \det(v_1, \dots, v_n) \end{aligned}$$

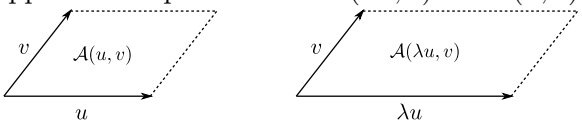
comme attendu. □

2.6.3 Définition (déterminant d'une famille de vecteurs)**.** *E de dimension n.* $\det_{\mathcal{B}}(u_1, \dots, u_n)$ *défini relativement à la base* $\mathcal{B}$ *par la condition* $\det(\mathcal{B}) = 1$.

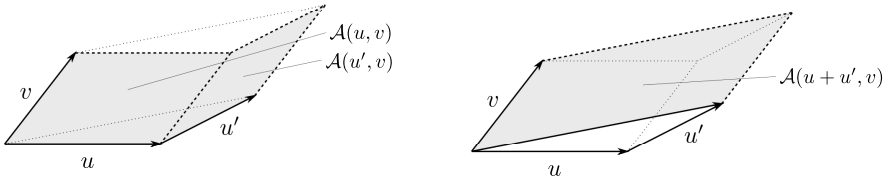
2.6.4 Interprétation géométrique. Slogan : « un déterminant c'est un volume (orienté) ». Aire *orientée* : $\mathcal{A}(v, u) = \mathcal{A}(-u, v) = -\mathcal{A}(u, v)$.



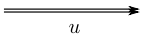
Axiomes d'un volume orienté :
— homogénéité par rapport à chaque facteur : $\mathcal{A}(\lambda u, v) = \lambda \mathcal{A}(u, v)$;



— additivité par rapport à chaque facteur : $\mathcal{A}(u + u', v) = \mathcal{A}(u, v) + \mathcal{A}(u', v)$;



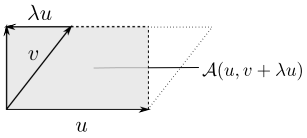
— caractère alterné : $\mathcal{A}(u, u) = 0$.



2.6.4.1. En profiter pour distinguer déterminants d'une famille de vecteurs, d'un endomorphisme, d'une matrice.

2.6.5 Cas euclidien. Si E est un espace euclidien, on a un choix canonique pour l'unité de volume, à savoir le volume orienté défini par une base orthnormée directe. Le déterminant dans une telle base ne dépend pas du choix de ladite base, et s'appelle le *produit mixte*.

2.6.6 Exemple. Aire d'un parallèlogramme = base × hauteur.



2.6.7 Proposition (Calcul du déterminant d'une famille de vecteurs comme déterminant d'une matrice)**.**

$$\det_{\mathcal{B}}(u_1, \ldots, u_n) = \det\big(\mathrm{Mat}_{\mathcal{B}}(u_1, \ldots, u_n)\big)$$

En particulier : Soit $M = (a_{ij}) \in \mathcal{M}_n(\mathbf{k})$ une matrice carrée de taille n à coefficients dans un corps $\mathbf{k}$. Notons $M_1, \ldots, M_n \in \mathbf{k}^n$ les colonnes de M. Le déterminant de M est le déterminant de la famille de vecteurs $(M_1, \ldots, M_n)$ dans la base canonique de $\mathbf{k}^n$.

2.6.8 Proposition. *Pour $f \in \mathcal{L}(E)$, l'application $f_* : \bigwedge^n E^\star \to \bigwedge^n E^\star$ définie par*

$$f_*\Delta(u_1, \ldots, u_n) = \Delta\big(f(u_1), \ldots, f(u_n)\big)$$

pour tout $\Delta \in \bigwedge^n E^\star$ est linéaire.

Il existe un unique scalaire $\lambda \in \mathbf{k}$ tel que

$$\mathrm{Mat}_{(\Delta_1)}(f_*) = (\lambda) \in \mathcal{M}_1(\mathbf{k})$$

pour tout choix d'une base (Δ_1) de $\bigwedge^n E^\star$.

2.6.9 Définition (déterminant d'un endomorphisme)**.** *Le scalaire λ comme dans la proposition 2.6.8 est le déterminant de f.*

2.6.10 Remarque. Le déterminant d'un endomorphisme ne dépend pas du choix d'une base.

2.6.11 Interprétation en termes de volume. C'est un rapport de volumes, donc il est sans unité, ce qui explique qu'il ne dépende pas du choix d'une base.

2.6.12 formule de changement de variables dans les intégrales.

2.6.13 Proposition. *Soit A un anneau. Pour toute matrice $M, N \in \mathcal{M}_n(A)$, on a* $\det(MN) = \det(M)\det(N)$.

Preuve. Si A est un corps, on considère l'application $M \mapsto \det(MN)$. C'est une forme n-linéaire alternée, donc de la forme $M \mapsto \lambda \cdot \det(M)$ par 2.6.2. En évaluant en $M = \mathrm{id}$, on trouve $\lambda = \det(N)$ ce qui donne la formule attendue.

□

2.6.14 Proposition. *Pour $f \in \mathcal{L}(E)$,*

$$\det(f) = \det\big(\mathrm{Mat}_{\mathcal{B}}(f)\big)$$

pour n'importe quel choix d'une base $\mathcal{B}$.

Les deux propositions ci-dessus sont étroitement liées l'une à l'autre.

Chapitre 3

Matrices sur un anneau

3.1 – Classes d'équivalence de matrices à coefficients dans un anneau principal, facteurs invariants

Rappelons que deux matrices $M, N \in \mathcal{M}_{mn}(A)$ sont *équivalentes* s'il existe $P \in \mathcal{M}_m(A)$, $Q \in \mathcal{M}_n(A)$ toutes deux inversibles telles que

$$N = PMQ.$$

Le résultat principal de cette sous-section donne un représentant canonique pour toute classe d'équivalence de matrices à coefficients dans un anneau principal A. Ceci permet en particulier de lister toutes les classes d'équivalence dans $\mathcal{M}_{mn}(A)$. Si A est un corps, on se souvient que deux matrices dans $\mathcal{M}_{mn}(A)$ sont équivalentes si et seulement si elles ont le même rang ; lorsque A est un anneau principal ce n'est plus suffisant, et il faut considérer ce qu'on appelle les facteurs invariants de la matrice. Comme le rang, ceux-ci peuvent être calculés par le pivot de Gauss.

Il sera utile pour la lecture de ce chapitre qu'une matrice carrée à coefficients dans un anneau A est inversible si et seulement si son déterminant est inversible dans A, voir 2.5.2.

3.1.1 Théorème (Théorème des facteurs invariants)**.** *Soit A un anneau principal, $M \in \mathcal{M}_{n,m}(A)$ ($n \leqslant m$ disons). La matrice M est équivalente dans $\mathcal{M}_{n,m}(A)$ à une matrice*

$$\begin{pmatrix} a_1 & & & 0 & \cdots & 0 \\ & \ddots & & \vdots & & \vdots \\ & & a_m & 0 & \cdots & 0 \end{pmatrix} \tag{3.1.1.1}$$

où $a_1 | \cdots | a_m$. Les a_i sont uniquement déterminés à multiplication par un inversible de A près.

En général, les premiers a_i sont égaux à 1 et les derniers à 0. Les a_i sont appelés les *facteurs invariants* de la matrice M. On appellera par abus de langage matrice *diagonale* une matrice comme en (3.1.1.1) (sans condition sur les facteurs diagonaux), et on la notera $\mathrm{diag}(a_1, \ldots, a_m)$ — ou éventuellement $\mathrm{diag}_{mn}(a_1, \ldots, a_m)$.

Deux mots sur la démonstration. L'existence d'une forme normale (3.1.1.1) s'obtient par un avatar du pivot de Gauss : c'est le pivot de Gauss authentique si A est euclidien, sinon il faut rajouter les opérations élémentaires de type Bezout. Tout ceci est discuté

dans la sous-section 3.2. Le pivot de Gauss ne donne aucunement l'unicité. Celle-ci est démontrée à part ci-dessous.

Les preuves qu'on trouve en général dans la littérature utilisent des techniques plus sophistiquées de théorie des modules, notamment la décomposition des modules de torsion selon les composantes primaires.

3.1.2. Pour $M \in \mathcal{M}_{mn}(A)$, on introduit pour $k = 0, \dots, \min(m,n)$ le pgcd $\delta_k(M)$ de tous les mineurs $k \times k$ de M,

$$\delta_k(M) := \mathrm{pgcd}_{I \in \mathcal{P}_k(m), J \in \mathcal{P}_k(n)} \Big(\det(M_{IJ}) \Big)$$

où $\mathcal{P}_k(n)$ désigne l'ensemble des parties de cardinal k de $[\![1,n]\!]$, et M_{IJ} la sous-matrice de M obtenue en extrayant les lignes dont l'indice est dans I et les colonnes dont l'indice est dans J.

3.1.3 Proposition. *Les quantités δ_k introduites ci-dessus sont constantes le long des classes d'équivalence de $\mathcal{M}_{mn}(A)$.*

Autrement dit, les δ_k sont des "invariants d'équivalence" de M. On en déduit immédiatement l'unicité des facteurs invariants : puisque $\delta_k(\mathrm{diag}(a_1, \dots, a_m)) = a_1 \cdots a_k$, si M est équivalente à $\mathrm{diag}(a_1, \dots, a_m)$, on a nécessairement

$$a_k = \begin{cases} \delta_k(M)/\delta_{k-1}(M) & \text{si } k \leqslant \mathrm{rg}(M) \\ 0 & \text{sinon.} \end{cases}$$

En principe cette formule permet le calcul des facteurs invariants sans avoir à appliquer le pivot de Gauss, mais il devrait être clair qu'il est déraisonnable d'espérer mettre ce calcul en œuvre en pratique.

Preuve de la Proposition 3.1.3. Soit $k \in [\![0,n]\!]$. Montrons que pour toute matrice $P \in \mathcal{M}_m(A)$, $\delta_k(M)$ divise $\delta_k(PM)$. Les lignes de la matrice PM sont des combinaisons linéaires des lignes de M, donc par multi-linéarité du déterminant, chaque mineur $k \times k$ de PM est une combinaison linéaire de mineurs $k \times k$ de M, et à ce titre est divisible par $\delta_k(M)$. Le résultat annoncé suit.

Lorsque P est inversible, on a aussi $M = P^{-1}(PM)$, et donc de la même façon $\delta_k(PM)$ divise $\delta_k(P^{-1}PM) = \delta_k(M)$. On a ainsi l'égalité $\delta_k(M) = \delta_k(PM)$.

En raisonnant de la même façon sur les colonnes plutôt que sur les lignes, on montre que pour tout $Q \in \mathcal{M}_n(A)$, $\delta_k(M)$ divise $\delta_k(MQ)$, puis que $\delta_k(M) = \delta_k(MQ)$ si Q est inversible. □

3.2 – Algorithme de Gauss

Cette sous-section est consacrée à la partie existence du Théorème 3.1.1, basée sur l'algorithme de Gauss (on se limitera au cas où A est euclidien, qui est le seul cas dont nous aurons besoin en pratique).

Comme dans le cas classique, il s'agit d'appliquer une suite d'opérations élémentaires aux lignes et aux colonnes de notre matrice pour la transformer en une matrice diagonale (on rappelle qu'on utilise ce terme même pour une matrice qui n'est pas carrée). La principale différence sur un anneau est que l'opération $L_i \leftarrow aL_i + bL_j$ n'est autorisée que si a est inversible dans A (si $A = \mathbf{Z}$, seulement si $a = \pm 1$, et si $A = \mathbf{k}[X]$, seulement si a est constant non-nul).

3.2.1 Principe de l'algorithme. On part d'une matrice $M = M_0 \in \mathcal{M}_{mn}(A)$. On va produire une suite de matrices $M_1, \dots, M_N \in \mathcal{M}_{mn}(A)$ telle que M_N est diagonale, et pour tout $i = 0, \dots, N-1$ il existe ou bien $P_i \in \mathrm{GL}_m(A)$ telle que

$$M_{i+1} = P_i M_i$$

ou bien $Q_i \in \mathrm{GL}_n(A)$ telle que $M_{i+1} = M_i Q_i$. Dans les deux cas M_{i+1} est équivalente à M_i, et puisque l'équivalence des matrices est une relation d'équivalence (!) on en déduit que M_N est équivalente à M_0.

Si $M_{i+1} = P_i M_i$, chaque ligne de M_{i+1} est la combinaison linéaire des lignes de M_i dont les coefficients sont donnés par la ligne correspondante de P_i. On passe donc de M_i à M_{i+1} « en effectuant des opérations sur les lignes ». De manière analogue, si $M_{i+1} = M_i Q_i$, on passe de M_i à M_{i+1} « en effectuant des opérations sur les colonnes ».

3.2.2 Opérations élémentaires. En pratique on se limite à prendre des matrices P_i et Q_i de certains types, qui correspondent à des *opérations élémentaires* sur les lignes et les colonnes de la matrice. Ici (sans doute par léger abus de langage vis à vis de la terminologie officielle), on appelle opération élémentaire sur les lignes (resp. sur les colonnes) la multiplication à gauche (resp. à droite) par une matrice du type

$$P = \begin{pmatrix} \lambda_1 & & \mu_1 & & \\ & \ddots & \vdots & & \\ & & \lambda_{i_0} & & \\ & & \vdots & \ddots & \\ & & \mu_m & & \lambda_m \end{pmatrix}, \quad \text{resp.} \quad Q = \begin{pmatrix} \lambda_1 & & & & \\ & \ddots & & & \\ \mu_1 & \cdots & \lambda_{i_0} & \cdots & \mu_n \\ & & & \ddots & \\ & & & & \lambda_n \end{pmatrix},$$

avec λ_i inversible pour tout i.

La multiplication à gauche par P revient à effectuer les opérations suivantes sur les lignes :

$$\begin{aligned} &L_{i_0} \leftarrow \lambda_{i_0} L_{i_0}, \\ &\forall i \neq i_0,\ L_i \leftarrow \lambda_i L_i + \mu_i L_{i_0}. \end{aligned}$$

On laisse au lecteur le soin d'écrire les opérations sur les colonnes correspondant à la multiplication à droite par Q.

On considérera aussi des *permutations* des lignes ou des colonnes. On peut les obtenir formellement comme combinaisons d'opérations élémentaires comme-ci-dessus mais en pratique il est plus agréable de les effectuer directement.

3.2.3 Opération de Bezout. En pratique il est commode de considérer des opérations supplémentaires sur les lignes et colonnes.

Soit $a, b \in A$, $d = \mathrm{pgcd}(a, b)$, et considérons une relation de Bezout : soit $u, v \in A$ tels que

$$au + bv = d \iff a'u + b'v = 1 \tag{3.2.3.1}$$

où l'on a écrit $a = a'd$ et $b = b'd$, avec $a', b' \in A$. La matrice

$$\begin{pmatrix} u & v \\ -b' & a' \end{pmatrix}$$

est inversible dans $\mathcal{M}_2(A)$ car de déterminant 1.

On en déduit qu'effectuer simultanément les deux opérations

$$\begin{cases} L_i \leftarrow uL_i + vL_j \\ L_j \leftarrow -b'L_i + a'L_j \end{cases} \quad \left(\text{resp.} \quad \begin{cases} C_i \leftarrow uC_i + vC_j \\ C_j \leftarrow -b'C_i + a'C_j \end{cases}\right)$$

(avec $i \neq j$) ne fait pas sortir de la classe d'équivalence : en effet cela revient à multiplier la matrice modifiée à gauche (resp. à droite) par une matrice inversible dans $\mathcal{M}_m(A)$ (resp. $\mathcal{M}_n(A)$) qu'on laisse au lecteur le soin d'écrire. Ces opérations transforment les lignes

$$\begin{matrix} L_i = \begin{pmatrix} a & * & \cdots & * \end{pmatrix} \\ L_j = \begin{pmatrix} b & * & \cdots & * \end{pmatrix} \end{matrix} \quad \text{en} \quad \begin{matrix} L_i' = \begin{pmatrix} d & *' & \cdots & *' \end{pmatrix} \\ L_j' = \begin{pmatrix} 0 & *' & \cdots & *' \end{pmatrix}. \end{matrix}$$

On les appelle *opérations de Bezout*.

3.2.3.1 Remarque. Dans le cas A euclidien — qui est celui considéré dans ce texte —, les opérations de Bezout sont inutiles en théorie car elles sont équivalentes à un nombre fini d'opérations élémentaires, grâce au fait que la relation (3.2.3.1) s'obtient par l'algorithme d'Euclide (voir Exemple 3.2.4 ci-dessous).

Le Théorème 3.1.1 est valable pour tout anneau A principal. Pour le démontrer dans ce cas, il est nécessaire de considérer les opérations de Bezout en plus des opérations élémentaires.

3.2.4 Exemple..

$$\begin{aligned} \begin{pmatrix} -2 & -11 & -7 \\ 5 & 18 & 11 \\ -6 & -20 & -12 \end{pmatrix} &\sim \begin{matrix} L_2 + 2L_1 \\ L_1 \\ L_3 - 3L_1 \end{matrix} \begin{pmatrix} 1 & -4 & -3 \\ -2 & -11 & -7 \\ 0 & 13 & 9 \end{pmatrix} \\ &\sim \begin{matrix} L_2 + 2L_1 \end{matrix} \begin{pmatrix} 1 & -4 & -3 \\ 0 & -19 & -13 \\ 0 & 13 & 9 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 19 & 13 \\ 0 & 13 & 9 \end{pmatrix} \\ &\sim \begin{matrix} -2L_2 + 3L_3 \\ 13L_2 - 19L_3 \end{matrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & -2 \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \end{aligned}$$

C'est plus intéressant comme ça qu'en étant malin pour fabriquer un 1 au premier coup avec L_2 et L_3, car on voit ainsi un vrai algorithme d'Euclide en cours de route :

$$\begin{aligned} 19 &= 13 \times 1 + 6 & 6 &= 19 - 13 \\ 13 &= 6 \times 2 + 1 & 1 &= 13 - 2 \cdot 6 \\ & & &= 13 - 2(19 - 13) = -2 \cdot 19 + 3 \cdot 13. \end{aligned}$$

3.2.5 Résolution pratique d'un système à coefficients dans un anneau. Plus haut je prétends qu'on arrive à la forme normale (3.1.1.1) par un pivot de Gauss. C'est vrai, mais il faut bien préciser qu'il faut autoriser les opérations élémentaires aussi bien sur les lignes (multiplication à gauche par $P \in \mathrm{GL}_m$) que sur les colonnes (multiplication à droite par $Q \in \mathrm{GL}_n$).

Si on se trouve concrètement en face d'un système d'équations linéaires, les opérations sur les lignes sont parfaitement banales, mais celles sur les colonnes reviennent à opérer

des changements de variables sur les inconnues. En théorie ces changements de variables ne sont en aucun cas une obstruction à la résolution du système, mais en pratique leur accumulation s'avère rapidement douloureuse. Il existe d'autres méthodes de résolution plus habiles en pratique, pour lesquelles on renvoie à la section 3.3.

3.2.6 Stathme euclidien. On note note

$$\varphi : A - \{0\} \to \mathbf{N}$$

le stathme euclidien, qu'on étend à A tout entier avec la convention $\varphi(0) = -\infty$. L'existence d'une division euclidienne assure que pour $a, b \in A$, si $b \neq 0$ alors il existe $q, r \in A$ tels que

$$a = bq + r \quad \text{et} \quad \varphi(r) < \varphi(b).$$

Nous considérerons seulement les deux anneaux euclidiens emblématiques $\mathbf{Z}$ et $\mathbf{k}[X]$.

3.2.6.1 Exemples. L'anneau $\mathbf{Z}$ muni du stathme φ défini par

$$\forall a \in \mathbf{Z}^*, \quad \varphi(a) = |a|$$

est euclidien.

L'anneau $\mathbf{k}[X]$, avec $\mathbf{k}$ un corps, muni du stathme φ défini par

$$\forall P \in \mathbf{k}[X], \quad \varphi(P) = \deg(P)$$

est euclidien.

3.2.7 Lemme. *Soit $m, n \in \mathbf{Z}_{\geqslant 2}$ et $M \in \mathcal{M}_{mn}(A)$. Il existe une matrice M' équivalente à M de la forme*

$$\begin{pmatrix} a & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & N & \\ 0 & & & \end{pmatrix}. \tag{3.2.7.1}$$

Preuve. Pour $M = (a_{ij})$, on pose

$$\varphi(M) := \min_{a_{ij} \neq 0} \varphi(a_{ij})\,;$$

on a donc $\varphi(M) = -\infty$ si et seulement si $M = 0$, et dans le cas contraire seuls les coefficients non nuls de M entrent en compte dans le calcul de $\varphi(M)$.

Montrons alors le résultat par récurrence sur $\varphi(M) \in \mathbf{N} \cup \{-\infty\}$. Si $\varphi(M) = -\infty$ alors $M = 0$ et le résultat est vrai. Supposons donc $\varphi(M) \in \mathbf{N}$. Dans ce cas $\varphi(M)$ est calculé par un coefficient non nul de M, et quitte à permuter les lignes et les colonnes on peut supposer que $\varphi(M) = a_{11}$.

Pour tout $i = 2, \ldots, m$ on écrit une division euclidienne $a_{i1} = a_{11}q_i + r_i$ de a_{i1} par a_{11}, et on effectue l'opération élémentaire $L_i \leftarrow L_i - q_i L_1$. De même pour tout $j = 2, \ldots, n$ on écrit une division euclidienne $a_{1j} = a_{11}p_i + s_i$ de a_{1i} par a_{11}, et on effectue l'opération élémentaire $C_j \leftarrow C_j - p_j C_1$. On obtient ainsi une matrice M' équivalente à M de la forme

$$M' = \begin{pmatrix} a_{11} & s_2 & \cdots & s_n \\ r_2 & & & \\ \vdots & & N & \\ r_m & & & \end{pmatrix}.$$

Si tous les r_i et s_j sont nuls, c'est une matrice M' de la forme voulue. Sinon on a

$$\varphi(M') \leqslant \min\Big(\min_{r_i\neq 0}\varphi(r_i), \min_{s_i\neq 0}\varphi(s_i)\Big) < \varphi(a_{11}) = \varphi(M)$$

et on conclut en appliquant l'hypothèse de récurrence à la matrice M'. □

3.2.8 Lemme. *Soit $a, b \in A$ euclidien, $d = \mathrm{pgcd}(a,b)$ et $m = \mathrm{ppcm}(a,b)$. Les deux matrices*

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \textit{et} \quad \begin{pmatrix} d & 0 \\ 0 & m \end{pmatrix}$$

sont équivalentes dans $\mathcal{M}_2(A)$.

Preuve. On commence par effectuer l'opération $C_1 \leftarrow C_1 + C_2$, qui donne

$$\begin{pmatrix} a & 0 \\ b & b \end{pmatrix}.$$

Ensuite on effectue une opération de Bezout comme en 3.2.3 (dont on reprend les notations), on obtient

$$\begin{pmatrix} d & bv \\ 0 & a'b \end{pmatrix}.$$

Il reste alors à faire $C_2 \leftarrow C_2 - b'vC_1$ pour arriver à

$$\begin{pmatrix} d & 0 \\ 0 & a'b \end{pmatrix}.$$

qui est la matrice cherchée puisque $a'b = da'b' = \mathrm{ppcm}(a,b)$. □

3.2.9 Preuve de l'existence des facteurs invariants dans le cas euclidien. Montrons par récurrence sur $m+n \in \mathbf{N}$ que toute matrice $M \in \mathcal{M}_{mn}(A)$ possède une matrice de la forme (3.1.1.1) dans sa classe d'équivalence. Quitte à raisonner sur la transposée, on peut supposer $m \leqslant n$. Si $m = 1$, on obtient directement le résultat en raisonnant comme dans la preuve du Lemme 3.2.7.

Supposons donc $m \geqslant 2$. Alors il existe une matrice diagonale par blocs comme en (3.2.7.1) équivalente à M. Par hypothèse de récurrence appliquée au bloc N, on obtient une matrice $M_0 = \mathrm{diag}(a, a_2, \ldots, a_m)$ équivalente à M avec $a_2|\cdots|a_m$.

On se ramène alors à une matrice du type voulu par application répétée du Lemme 3.2.8. On commence par l'appliquer aux deux premières lignes : on obtient une matrice équivalente $M_1 = \mathrm{diag}(a'_1, \tilde{a}_2, a_3, \ldots, a_m)$ avec $a'_1 = \mathrm{pgcd}(a, a_2)$ et $\tilde{a}_2 = \mathrm{ppcm}(a, a_2)$. Ensuite on définit par récurrence $M_i = \mathrm{diag}(a'_1, \ldots, a'_i, \tilde{a}_{i+1}, a_{i+2}, \ldots)$ pour tout $i = 2, \ldots, m-1$ en appliquant le lemme aux lignes d'indices i et $i+1$ de M_{i-1} ; ainsi

$$a'_i = \mathrm{pgcd}(\tilde{a}_i, a_{i+1}) \quad \textit{et} \quad \tilde{a}_{i+1} = \mathrm{ppcm}(\tilde{a}_i, a_{i+1}).$$

Montrons par récurrence sur $i = 1, \ldots, m-1$ que $a'_1|\cdots|a'_i|\tilde{a}_{i+1}$. Pour $i = 1$ il s'agit de démontrer que $a'_1|\tilde{a}_2$, c'est-à-dire $\mathrm{pgcd}(a,a_2)|\mathrm{ppcm}(a,a_2)$, ce qui est bien vrai. Soit $i > 1$ et supposons établi que $a'_1|\cdots|a'_{i-1}|\tilde{a}_i$. Par définition a'_{i-1} divise a_i, qui lui-même divise a_{i+1} par construction de M_0. Ainsi a'_{i-1} divise $\tilde{a}_i$ et a_{i+1}, et donc *a fortiori* aussi $\mathrm{pgcd}(\tilde{a}_i, a_{i+1}) = a'_i$. Enfin a'_i divise $\tilde{a}_{i+1}$ car $\mathrm{pgcd}(\tilde{a}_i, a_{i+1})|\mathrm{ppcm}(\tilde{a}_i, a_{i+1})$. Ceci conclut notre récurrence.

À l'arrivée, on a la matrice $M_{m-1} = \mathrm{diag}(a'_1, \ldots, a'_{m-1}, \tilde{a}_m)$ qui est équivalente à M et telle que $a'_1|\cdots, a'_{m-1}|\tilde{a}_m$: elle satisfait ainsi à toutes les propriétés requises. □

Pour finir, on donne un résultat utile sans lien direct avec ce qui précède.

3.2.10 Complétion d'un vecteur primitif en une matrice inversible. On démontre ici le résultat suivant : soit $a \in A^n$; il existe une matrice $A \in \mathrm{GL}_n(A)$ dont la première colonne est a si et seulement si les coordonnées de a sont premières entre elles dans leur ensemble. Voir aussi 3.3.12 et **??**. La preuve présentée ici est tirée de [**?**, II.1.39].

Un sens s'obtient rapidement : si

$$A = \begin{pmatrix} a_1 & * & * \\ \vdots & \vdots & \vdots \\ a_n & * & * \end{pmatrix}$$

est inversible, alors en développant le déterminant de A par rapport à la première colonne on trouve $v_1, \ldots, v_n \in A$ tels que

$$v_1 a_1 + \cdots + v_n a_n \in A^\times,$$

et donc $a_1, \ldots, a_n$ sont premiers entre eux dans leur ensemble par le théorème de Bezout.

Réciproquement, montrons par récurrence sur $n \in \mathbf{N}^*$ que si $a_1, \ldots, a_n$ sont premiers entre eux dans leur ensemble, alors il existe une matrice A comme ci-dessus qui est inversible. Pour $n = 1$ le résultat est trivial. Supposons donc $n > 1$ et le résultat démontré pour tout $n' < n$. On considère $a_1, \ldots, a_{n-1}$: on note $d = \mathrm{pgcd}(a_1, \ldots, a_{n-1})$, et on pose $a'_i = a_i/d$ pour $i = 1, \ldots, n-1$. Par hypothèse de récurrence, il existe une matrice

$$B' = \left(\begin{array}{c|c} a'_1 & \\ \vdots & \quad B_0 \quad \\ a'_{n-1} & \end{array}\right) \in \mathrm{GL}_{n-1}(A).$$

On pose $u = \det(B') \in A^\times$. On considère la matrice

$$B = \left(\begin{array}{c|c} a_1 & \\ \vdots & \quad B_0 \quad \\ a_{n-1} & \end{array}\right)$$

dont le déterminant est ud. Puisque $a_1, \ldots, a_{n-1}, a_n$ sont premiers entre eux, d et a_n sont premiers entre eux, et donc (puisque u est inversible) il existe $v_0, v_1 \in A$ tels que

$$v_0 \det(B) + v_1 a_n = 1.$$

La matrice

$$A = \left(\begin{array}{c|ccc|c} a_1 & & & & -u^{-1}a'_1 v_1 \\ \vdots & & B_0 & & \vdots \\ a_{n-1} & & & & -u^{-1}a'_{n-1} v_1 \\ \hline a_n & 0 & \cdots & 0 & v_0 \end{array}\right)$$

répond à notre problème. En effet, en développant le déterminant de A par rapport à la

dernière ligne on obtient

$$\begin{aligned}\det(A) &= (-1)^{n+1}a_n \det\left(\begin{array}{c|c} & -u^{-1}a'_1v_1 \\ B_0 & \vdots \\ & -u^{-1}a'_{n-1}v_1 \end{array}\right) + v_0\det(B)\\ &= (-1)^{n+1}(-1)^{n-2}(-u^{-1}v_1)a_n \det\left(\begin{array}{c|c} a'_1 & \\ \vdots & B_0 \\ a'_{n-1} & \end{array}\right) + v_0\det(B)\\ &= v_1a_n + v_0\det(B) = 1,\end{aligned}$$

donc A est bien inversible.

Le lecteur consciencieux vérifiera qu'en appliquant la construction ci-dessus à un vecteur primitif $a \in A^2$ on obtient une matrice de Bezout similaire à celle de 3.2.3.

3.3 – Forme normale de Hermite

3.3.1 Lemme. *Soit $a, b \in \mathbf{Z}$ non tous deux nuls, et d un PGCD de a, b. Alors il existe une matrice $U \in \mathrm{SL}_2(\mathbf{Z})$ tel que $\begin{pmatrix} a & b \end{pmatrix} U = \begin{pmatrix} 0 & d \end{pmatrix}$.*

Preuve. On trouve une relation de Bézout $au + bv = d$ grâce à l'algorithme d'Euclide étendu **??**. En posant $a = a'd$, $b = b'd$, on a donc $a'u + b'v = 1$, et

$$U = \begin{pmatrix} b' & u \\ -a' & v \end{pmatrix} \in \mathrm{SL}_2(\mathbf{Z})$$

convient. □

3.3.2 Remarque. L'argument marche également quand a ou b est nul. Si $a = 0$ on prend $U = \pm I_2$, et si $b = 0$ on prend $U = \pm \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

3.3.3 Proposition. *Soit $n \geqslant 2$, $a_1, \ldots, a_n \in \mathbf{Z}$, et d un PGCD des a_i. Alors il existe une matrice $U \in \mathrm{SL}_n(\mathbf{Z})$ tel que*

$$\begin{pmatrix} a_1 & \ldots & a_n \end{pmatrix} U = \begin{pmatrix} 0 & \ldots & 0 & d \end{pmatrix}$$

Preuve. Par récurrence, en utilisant le lemme 3.3.1. □

3.3.4 Exemple. Considérons le vecteur (2 3 5). En utilisant le procédé du lemme on obtient :

$$(2\ 3)\begin{pmatrix} 3 & -1 \\ -2 & 1 \end{pmatrix} = (0\ 1), \qquad (1\ 5)\begin{pmatrix} 5 & 1 \\ -1 & 0 \end{pmatrix} = (0\ 1).$$

D'où

$$(2\ 3\ 5)\begin{pmatrix} 3 & -1 & 0 \\ -2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & -1 & 0 \end{pmatrix} = (0\ 1\ 5)\begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & -1 & 0 \end{pmatrix} = (0\ 0\ 1).$$

Noter que la matrice U obtenue n'est pas unique, par exemple on a aussi

$$(2\ 3\ 5)\begin{pmatrix}-4 & -3 & -1\\ 1 & 2 & 1\\ 1 & 0 & 0\end{pmatrix} = (0\ 0\ 1).$$

Noter aussi que ce procédé de multiplications successives par matrices avec un bloc 2×2 n'est certainement pas l'algorithme le plus efficace .

Voici un analogue sur $\mathbf{Z}$ de la notion de matrice co-échelonnée en colonnes réduite :

3.3.5 Définition. *Une matrice rectangulaire $H = (h_{ij})$ à m lignes et n colonnes et à coefficients dans $\mathbf{Z}$ est sous* forme normale de Hermite *s'il existe $s \geqslant 0$ et une fonction strictement croissante $f : [\![s+1, n]\!] \to [\![1, m]\!]$ tels que*

(i) *les s premières colonnes de H sont nulles,*

(ii) *pour tout $j \in [\![s+1, n]\!]$ on a $h_{f(j),j} \geqslant 1$,*

(iii) *$h_{i,j} = 0$ pour $i > f(j)$,*

(iv) *$0 \leqslant h_{f(j),k} < h_{f(j),j}$ pour tout $k > j$.*

Les coefficients $h_{f(j),j} \geqslant 1$ sont appelés les pivots *de la matrice, chaque colonne non nulle contient un tel pivot par définition. La définition signifie que tous les coefficients sous un pivot sont nuls, et ceux à droite d'un pivot sont positifs ou nuls et inférieurs au pivot. Une matrice vérifiant seulement les conditions (i) à (iii) de la définition sera dite* co-échelonnée en colonnes.

3.3.6 Exemple. Les matrices

$$\begin{pmatrix}6 & -8 & -9\\ \boxed{2} & 1 & 0\\ 0 & \boxed{4} & 3\\ 0 & 0 & \boxed{3}\end{pmatrix}, \qquad \begin{pmatrix}0 & \boxed{2} & 1\\ 0 & 0 & -5\\ 0 & 0 & \boxed{1}\end{pmatrix}, \qquad \begin{pmatrix}0 & \boxed{1}\\ 0 & 0\end{pmatrix}$$

sont sous forme normale de Hermite, avec les pivots encadrés.

3.3.7 Théorème. *Soit A une matrice rectangulaire $m \times n$ à coefficients dans $\mathbf{Z}$. Alors il existe une matrice H de taille $m \times n$ et une matrice $U \in \mathrm{GL}_n(\mathbf{Z})$ tel que $H = AU$ et H soit sous forme normale de Hermite. De plus H est unique (mais pas U).*

Preuve. On commence par regarder la dernière ligne de la matrice. Si elle n'est pas nulle, par la proposition 3.3.3 on peut multiplier à droite par une matrice dans $\mathrm{SL}_n(\mathbf{Z})$ (ou dans $\mathrm{GL}_1(\mathbf{Z})$ quand $n = 1$) pour avoir un seul coefficient positif non nul en dernière position.

Supposons maintenant que les $k \geqslant 1$ dernières lignes de A forme une matrice $k \times n$ sous forme normale de Hermite, avec premier pivot en position j. Autrement dit A s'écrit par bloc de la façon suivante, où H' est sous forme normale de Hermite avec k lignes et première colonne non nulle, et les a_i sont des coefficients que l'on va maintenant chercher à simplifier :

$$A = \begin{pmatrix} & * & & *\\ a_1 & \dots & a_{j-1} & *\\ & 0 & & H'\end{pmatrix}$$

Si $j = 1$ ou $k = m$, la matrice est déjà sous forme normale de Hermite. Si $a_1 = \cdots = a_{j-1} = 0$, les $k+1$ dernières lignes de A forment déjà une matrice sous forme normale

de Hermite, avec indice j du premier pivot inchangé. Si $j > 1$, $k < m$, et que les a_i ne sont pas tous nuls, par la proposition 3.3.3 il existe une matrice $U' \in \mathrm{SL}_{j-1}(\mathbf{k})$ (ou dans $\mathrm{GL}_1(\mathbf{Z})$ si $j = 2$) telle que $\begin{pmatrix} a_1 & \dots & a_{j-1} \end{pmatrix} U' = \begin{pmatrix} 0 & \dots & 0 & d \end{pmatrix}$, et donc le produit

$$A \begin{pmatrix} U' & 0 \\ 0 & I_k \end{pmatrix} = \left(\begin{array}{cccc|ccc} & & * & & & * & \\ 0 & \dots & 0 & \boxed{\mathrm{d}} & b_j & \dots & b_n \\ \hline & & 0 & & & H' & \end{array} \right)$$

nous donne une matrice dont les $k+1$ dernières lignes sont co-échelonnées en colonnes, avec pivot d en colonne $j-1$. Reste à normaliser les coefficients b_i à droite du pivot d pour obtenir une forme normale de Hermite sur les $k+1$ dernières lignes. Pour chaque $l \geqslant j$, on écrit la division euclidienne $b_l = q_l d + r_l$ avec $0 \leqslant r_l < d$. En multipliant à droite par une matrice élémentaire, on remplace la colonne C_l par $C_l - q_l C_{j-1}$, ce qui remplace b_l par r_l, sans affecter la matrice H'.

Unicité : Si H et H' sont deux formes normales de Hermite, on montre leur égalité colonne par colonne, en partant de la colonne de droite (celle avec le pivot le plus bas), en exprimant chaque colonne de H' comme une combinaison linéaire des colonnes de H. □

3.3.8 Exemple. Plus de détails sur la preuve de l'unicité, sur l'exemple de la matrice

$$H = \begin{pmatrix} 6 & -8 & -9 \\ \boxed{2} & 1 & 0 \\ 0 & \boxed{4} & 3 \\ 0 & 0 & \boxed{3} \end{pmatrix}$$

Supposons qu'on ait une autre forme normale

$$H' = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \end{pmatrix} = HU$$

avec $U \in \mathrm{GL}_3(\mathbf{Z})$. Le pivot de la troisième colonne de H' est forcément à la dernière ligne, sinon la quatrième ligne de H' serait identiquement nulle et la dernière colonne de H ne pourrait être une combinaison entière des colonnes de H'. On a donc

$$H' = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \\ 0 & 0 & \boxed{*} \end{pmatrix} = HU$$

Mais les deux zéros à gauche du pivot impliquent que U est triangulaire par bloc :

$$H' = \begin{pmatrix} * & * & * \\ * & * & * \\ * & * & * \\ 0 & 0 & \boxed{*} \end{pmatrix} = HU = \begin{pmatrix} 6 & -8 & -9 \\ \boxed{2} & 1 & 0 \\ 0 & \boxed{4} & 3 \\ 0 & 0 & \boxed{3} \end{pmatrix} \begin{pmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ u_{2,1} & u_{2,2} & u_{2,3} \\ 0 & 0 & \pm 1 \end{pmatrix}$$

De plus comme un pivot est positif par définition, on doit avoir le coefficient $u_{3,3}$ de U égal à 1. En procédant de même pour les autres colonnes on obtient déjà que U est triangulaire

supérieure avec des 1 sur la diagonale :

$$H' = \begin{pmatrix} * & * & * \\ \boxed{2} & 1+2u_{1,2} & \cdots \\ 0 & \boxed{4} & 3+4u_{2,3} \\ 0 & 0 & \boxed{3} \end{pmatrix} = HU = \begin{pmatrix} 6 & -8 & -9 \\ \boxed{2} & 1 & 0 \\ 0 & \boxed{4} & 3 \\ 0 & 0 & \boxed{3} \end{pmatrix} \begin{pmatrix} 1 & u_{1,2} & u_{1,3} \\ 0 & 1 & u_{2,3} \\ 0 & 0 & 1 \end{pmatrix}$$

Comme par définition le coefficient $3+4u_{2,3}$ doit être positif et plus petit que le pivot 4, on en déduit $u_{2,3} = 0$. De même on obtient $u_{1,2} = 0$, puis $u_{1,3} = 0$ avec le coefficient ... que l'on laisse au lecteur le soin d'expliciter.

3.3.9 Application (Trouver une base)**.** Avec les notations du théorème, les colonnes non nulles de la matrice H forment une base du groupe abélien libre engendré par les colonnes de A. On pourra méditer l'exemple simple donné par

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 0 \end{pmatrix}, \qquad\qquad H = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},$$

où l'on peut constater qu'aucune des colonnes initiales ne formait une base.

3.3.10 Application (Tester l'appartenance à un groupe)**.** Soit

$$H = \begin{pmatrix} 6 & -8 & -9 \\ \boxed{2} & 1 & 0 \\ 0 & \boxed{4} & 3 \\ 0 & 0 & \boxed{3} \end{pmatrix} \qquad\qquad V = \begin{pmatrix} -7 \\ 1 \\ -1 \\ 3 \end{pmatrix}.$$

Le vecteur V est-il combinaison entière des colonnes C_i de la matrice H ?

Réponse : on cherche V sous la forme $V = a_1C_1 + a_2C_2 + a_3C_3$. On trouve successivement $a_3 = 1$, $a_2 = -1$, $a_1 = 1$, et on vérifie que le premier coefficient -7 est compatible avec ces contraintes.

3.3.11 Application (Résoudre un système linéaire à coefficient entier)**.** On veut résoudre l'équation $2x + 3y + 5z = 0$. Sous forme matricielle :

$$(2\ 3\ 5) \begin{pmatrix} x \\ y \\ z \end{pmatrix} = (0).$$

On met la matrice $A = (2\ 3\ 5)$ sous forme normale de Hermite $AU = H$:

$$(2\ 3\ 5) \begin{pmatrix} -4 & -3 & -1 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix} = (0\ 0\ 1)$$

Les colonnes de U forment une base de $\mathbf{Z}^3$. Les 2 premières colonnes de la matrice U forment donc une base de l'espace Sol des solutions, c'est à dire

$$\mathrm{Sol} = \left\{ \mathbf{Z} \begin{pmatrix} -4 \\ 1 \\ 1 \end{pmatrix} + \mathbf{Z} \begin{pmatrix} -3 \\ 2 \\ 0 \end{pmatrix} \right\} \simeq \mathbf{Z}^2.$$

3.3.12 Application (Compléter une base)**.** On veut compléter le vecteur (2 3 5) en une base de $\mathbf{Z}^3$ (c'est possible car les coefficients sont premiers entre eux dans leur ensemble). On considère cette fois l'inverse de la matrice qui amène à la forme normale de Hermite :

$$\begin{pmatrix} 0 & 0 & 1 \\ -1 & -1 & -3 \\ 2 & 3 & 5 \end{pmatrix} \begin{pmatrix} -4 & -3 & -1 \\ 1 & 2 & 1 \\ 1 & 0 & 0 \end{pmatrix} = I_3$$

Les 3 lignes de la première matrice donnent la base souhaitée.

- Selon les sources la forme normale de Hermite est échelonnée en lignes, en colonnes, ou co-échelonnée en colonnes. J'ai suivi Cohen sur ce dernier choix, et adapté des exemples de Coste qui échelonne en lignes.
- La forme normale de Smith pourrait aussi être abordée. Et en fait on peut aussi résoudre les systèmes linéaires avec cette autre forme normale.

Chapitre 4

Dualité

INTRODUCTION À ÉCRIRE.

1) dualité tout court, sens géométrique, encore mieux en projectif.
   ▷ importance du bidual pour avoir une forme symétrique.
2) un aspect particulièrement remarquable est qu'on peut faire de la géométrie sur les espaces d'équations ; en particulier on peut considérer des lieux définis par des équations linéaires, et un truc frappant est que le dual de l'espace des équations linéaires est l'espace de départ (ça ne fonctionne qu'en dimension finie)
3) version matriciellle de la dualité : une matrice peut être vue comme une suite de colonnes, ou une suite de lignes !
4) la preuve du thorème de bidualité est très simple, on appelle ça théorème pour l'importance de ses implications.
5) dualité par rapport à une FQ.

4.1 – Formes linéaires, espace dual

4.1.1 Définition. *Soit E un $\mathbf{k}$-espace vectoriel de dimension finie n. Une* forme linéaire *est une application linéaire $E \to \mathbf{k}$.*

*L'ensemble de toutes les formes linéaires sur E est un $\mathbf{k}$-espace vectoriel de dimension finie n que l'on connaît bien : $\mathcal{L}(E, \mathbf{k})$. On l'appelle l'*espace dual *de E, et on le note $E^\star$. (On peut aussi trouver la notation $E^\vee$, mais nous ne l'utiliserons pas dans ce texte).*

Avant d'introduire la notion de base duale, rappelons un fait bien connu d'algèbre linéaire. Si E, F sont deux $\mathbf{k}$-espaces vectoriels munis de bases respectives $(e_1, \dots, e_n)$ et $(f_1, \dots, f_m)$, alors il existe une unique famille d'applications linéaires $(\varphi_i^s)_{1 \leqslant i \leqslant n, 1 \leqslant s \leqslant m}$ de E dans F telle que

$$\forall i, j \in [\![1, n]\!],\ \forall s \in [\![1, m]\!] : \quad \varphi_i^s(e_j) = \delta_{ij}.f_s,$$

et cette famille constitue une base de $\mathcal{L}(E, F)$. Matriciellement, c'est juste le fait que les matrices avec un seul coefficient égal à 1 et des 0 partout ailleurs forment une base de l'espace des matrices rectangulaires $m \times n$. En appliquant ce fait à $E^\star = \mathcal{L}(E, \mathbf{k})$ on obtient :

4.1.2 Définition. *Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E. Il existe une unique famille $(e_1^\star, \dots, e_n^\star)$ de formes linéaires sur E telle que*

$$\forall i, j \in [\![1, n]\!] : \quad e_i^\star(e_j) = \delta_{ij}.$$

Cette famille est une base de $E^\star$, on l'appelle la base duale *de $\mathcal{B}$ et on la note $\mathcal{B}^\star$.*

Attention, la notation pour la base duale peut être trompeuse : chaque forme $e_i^\star$ dépend de l'ensemble des vecteurs de la base $\mathcal{B}$. Par contre si on se donne un seul vecteur $e \in E$, il n'y a pas de forme $e^\star$ bien définie. En effet il existe en général plein de formes linéaires ℓ telles que $\ell(e) = 1$ (en fait une infinité, dès que $\mathbf{k}$ est infini et $n > 1$).

Dans le système de coordonnées $(x_1, \dots, x_n)$ défini par la base $\mathcal{B}$, $e_i^\star$ est simplement la fonction linéaire qui à tout vecteur $x \in E$ associe sa i-ème coordonnée $x_i \in \mathbf{k}$. Ainsi une fois une base de E choisie, l'espace dual $E^\star$ s'identifie à l'espace vectoriel des polynômes *homogènes* de degré 1 en les n variables $x_1, \dots, x_n$, c'est-à-dire aux polynômes de la forme

$$a_1 x_1 + \cdots + a_n x_n, \quad a_1, \dots, a_n \in \mathbf{k}.$$

La famille $(x_1, \dots, x_n)$ forme une base de cet espace vectoriel, et correspond à la base duale via cette identification.

De la même façon, on verra plus loin que l'espace vectoriel des formes quadratiques sur E s'identifie aux polynômes homogènes de degré 2 en $x_1, \dots, x_n$, qui sont les polynômes de la forme

$$\sum_{1 \leqslant i \leqslant j \leqslant n} a_{ij} x_i x_j, \quad a_{ij} \in \mathbf{k}.$$

4.1.3 Lemme. *Soit $\mathcal{B} = (e_1, \dots, e_n)$ une base de E, et $\ell \in E^\star$. Alors*

$$\ell = \ell(e_1).e_1^\star + \cdots + \ell(e_n).e_n^\star,$$

autrement dit les coordonnées de ℓ dans la base duale sont $\big(\ell(e_1), \dots, \ell(e_n)\big) \in \mathbf{k}^n$.

Preuve. Soit $x \in E$, de coordonnées $(x_1, \dots, x_n) \in \mathbf{k}^n$ dans la base $\mathcal{B}$. On a

$$\begin{aligned} \ell(x) = \ell(x_1.e_1 + \cdots + x_n.e_n) &= x_1 \ell(e_1) + \cdots + x_n \ell(e_n) \\ &= e_1^\star(x)\ell(e_1) + \cdots + e_n^\star(x)\ell(e_n) \\ &= \big(\ell(e_1).e_1^\star + \cdots + \ell(e_n).e_n^\star\big)(x). \end{aligned}$$

Ainsi les deux formes linéaires ℓ et $\ell(e_1).e_1^\star + \cdots + \ell(e_n).e_n^\star$ prennent les mêmes valeurs en tout $x \in E$, et sont donc égales. □

4.1.4 Exemple. Soit f une fonction à valeurs réelles définie sur un ouvert $U \subseteq \mathbf{R}^n$, différentiable en un point $p \in U$. La différentielle de f en p, df_p est une forme linéaire sur $\mathbf{R}^n$.[1] Dans ce contexte, on a l'habitude de noter $(dx_1, \dots, dx_n)$ la base duale de la base canonique de $\mathbf{R}^n$. On a alors

$$df_p = \tfrac{\partial f}{\partial x_1}(p).dx_1 + \cdots + \tfrac{\partial f}{\partial x_n}(p).dx_n,$$

ainsi les coordonnées de df_p dans la base canonique de $(\mathbf{R}^n)^\star$ sont $\big(\tfrac{\partial f}{\partial x_1}(p), \dots, \tfrac{\partial f}{\partial x_n}(p)\big)$.

1. Plus généralement, si f est une fonction à valeurs réelles définie sur un ouvert U d'une variété différentiable M, différentiable en un point $p \in U$, la différentielle de f en p est une forme linéaire sur l'espace tangent T_pM, qui est un $\mathbf{R}$-espace vectoriel de dimension $\dim(M)$.

4.1.5 Lemme (changement de base)**.** *Soit $\mathcal{B}_1$ et $\mathcal{B}_2$ deux bases de E. On a*

$$\mathrm{Mat}_{\mathcal{B}_1^\star}(\mathcal{B}_2^\star) = (\mathrm{Mat}_{\mathcal{B}_1}(\mathcal{B}_2)^\mathsf{T})^{-1}.$$

Preuve. Nous allons démontrer la formule $\mathrm{Mat}_{\mathcal{B}_2^\star}(\mathcal{B}_1^\star) = (\mathrm{Mat}_{\mathcal{B}_1}(\mathcal{B}_2))^\mathsf{T}$, dont on déduit aussitôt la formule voulue. Notons $\mathcal{B}_1 = (e_1, \dots, e_n)$, $\mathcal{B}_2 = (\varepsilon_1, \dots, \varepsilon_n)$, et $\mathrm{Mat}_{\mathcal{B}_1}(\mathcal{B}_2) = (a_{ij})_{1\leqslant i,j\leqslant n}$, de sorte que pour tout $j = 1, \dots, n$ on a $\varepsilon_j = \sum_i a_{ij}.e_i$. On calcule $\mathrm{Mat}_{\mathcal{B}_2^\star}(\mathcal{B}_1^\star)$ en utilisant le lemme 4.1.3. Soit $j \in [\![1, n]\!]$. On a

$$\begin{aligned} e_j^\star &= e_j^\star(\varepsilon_1).\varepsilon_1^\star + \cdots + e_j^\star(\varepsilon_n).\varepsilon_n^\star \\ &= a_{j1}.\varepsilon_1^\star + \cdots + a_{jn}.\varepsilon_n^\star, \end{aligned}$$

puisque $e_j^\star(\varepsilon_i)$ est la j-ème coordonnée de ε_i dans la base $(e_1, \dots, e_n)$, c'est-à-dire a_{ji}. On a donc bien

$$\mathrm{Mat}_{\mathcal{B}_2^\star}(\mathcal{B}_1^\star) = (a_{ji})_{1\leqslant i,j\leqslant n} = \mathrm{Mat}_{\mathcal{B}_1}(\mathcal{B}_2)^\mathsf{T}$$

comme annoncé. □

4.1.6 Exemple. Soit E une droite (c'est le cas $n = 1$), et $e \neq 0$ un vecteur de E. Alors e constitue à lui seul une base de E, et pour tout $\lambda \in \mathbf{k}^*$, $(\lambda.e)$ est une autre base de E. On a :

$$(\lambda e)^\star = \frac{1}{\lambda} e^\star.$$

En effet :

$$(\lambda e)^\star(e) = \frac{1}{\lambda}(\lambda e)^\star(\lambda e) = \frac{1}{\lambda} = \frac{1}{\lambda} e^\star(e).$$

4.1.7 Définition. *Soit E, F deux $\mathbf{k}$-espaces vectoriels, et $u\colon E \to F$ une application linéaire. La* transposée *de u est l'application linéaire*

$$\begin{array}{rccc} u^\mathsf{T}\colon & F^\star & \longrightarrow & E^\star \\ & \ell & \longmapsto & \ell \circ u. \end{array}$$

Remarquons que la transposition renverse le sens des flèches. Il faut faire avec, il est dans la nature de la dualité de tout renverser.

4.1.8 Lemme. *Dans les notations de la définition 4.1.7, supposons E et F tous deux de dimension finie, et soit $\mathcal{B}_E$ et $\mathcal{B}_F$ des bases de E et F respectivement. On a*

$$\mathrm{Mat}_{\mathcal{B}_F^\star, \mathcal{B}_E^\star}(u^\mathsf{T}) = \mathrm{Mat}_{\mathcal{B}_E, \mathcal{B}_F}(u)^\mathsf{T}.$$

Autrement dit, la matrice de l'application transposée est la transposée de la matrice de l'application initiale. Cette formule était attendue, vu le choix de la terminologie. Elle donne une interprétation géométrique à l'opération de transposition matricielle, qui jusqu'à présent n'était qu'une manipulation combinatoire des entrées d'un tableau de nombres.

Preuve. À nouveau, les calculs pour établir cette formule sont basés sur la formule du lemme 4.1.3. Notons $\mathcal{B}_E = (e_1, \dots, e_n)$, $\mathcal{B}_F = (f_1, \dots, f_m)$, et $\mathrm{Mat}_{\mathcal{B}_E, \mathcal{B}_F}(u) = (a_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n}$. Soit $j \in [\![1, m]\!]$. On a $u^\mathsf{T}(f_j^\star) = u^\mathsf{T}(f_j^\star)(e_1).e_1^\star + \cdots + u^\mathsf{T}(f_j^\star)(e_n).e_n^\star$, et pour chaque $i = 1, \dots, n$,

$$u^\mathsf{T}(f_j^\star)(e_i) = (f_j^\star \circ u)(e_i) = f_j^\star(u(e_i))$$

est la j-ème coordonnée de $u(e_i)$ dans la base $(f_1, \dots, f_m)$, c'est-à-dire a_{ji}. On a donc $\mathrm{Mat}_{\mathcal{B}_F^\star, \mathcal{B}_E^\star}(u^\mathsf{T}) = (a_{ji})_{1\leqslant j\leqslant n, 1\leqslant i\leqslant m}$ comme attendu. □

4.1.9 Remarque. On peut utiliser cette formule pour prouver la formule de changement de base du lemme 4.1.5. On reprend les notations de ce lemme. On a

$$\mathrm{Mat}_{\mathcal{B}_1^\star}(\mathcal{B}_2^\star) = \mathrm{Mat}_{\mathcal{B}_2^\star,\mathcal{B}_1^\star}(\mathrm{id}_{E^\star}) = \mathrm{Mat}_{\mathcal{B}_1,\mathcal{B}_2}(\mathrm{id}_E)^\mathsf{T} = \mathrm{Mat}_{\mathcal{B}_2}(\mathcal{B}_1)^\mathsf{T},$$

en utilisant le lemme 4.1.8 pour la deuxième égalité, après avoir noté que $\mathrm{id}_{E^\star} = \mathrm{id}_E^\mathsf{T}$.

4.1.10 Présentation matricielle de la dualité. Une fois choisie une base $\mathcal{B} = (e_1, \ldots, e_n)$ pour E, on représente les éléments de E comme des vecteurs colonnes, ou autrement dit, des éléments de $\mathbf{k}^n = \mathcal{M}_{n,1}(\mathbf{k})$). De même, les élements de $E^\star = \mathcal{L}(E, \mathbf{k})$ sont représentés par leur matrice dans la base $\mathcal{B}$ et la base canonique de $\mathbf{k}$, qui est un vecteur ligne (autrement dit, un vecteur de $(\mathbf{k}^n)^\mathsf{T} = \mathcal{M}_{1,n}(\mathbf{k})$). L'évaluation $\ell(x)$ pour $x \in E$ et $\ell \in E^\star$ correspond au produit d'un vecteur ligne par un vecteur colonne, ce qui donne une matrice 1×1 que l'on identifie à un scalaire. La base $(e_1, \ldots, e_n)$ correspond aux vecteurs colonnes

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

et la base duale $(e_1^\star, \ldots, e_n^\star)$ correspond aux vecteurs lignes

$$(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1).$$

La formule du lemme 4.1.3 dit que la colonne des coordonnées d'une forme linéaire $\ell \in E^\star$ dans la base $\mathcal{B}^\star$ est la transposée de sa matrice (ligne) dans les bases $\mathcal{B}$ de E et base canonique de $\mathbf{k}$.

Reprenons la formule du lemme 4.1.8 en ces termes. On reprend la notation de ce lemme, et on appelle $A = \mathrm{Mat}_{\mathcal{B}_E,\mathcal{B}_F}(u)$. Soit $\ell \in F^\star$. Voyant $\ell \in \mathcal{L}(F, \mathbf{k})$, on écrit sa matrice

$$L = \mathrm{Mat}_{\mathcal{B}_F,\mathcal{B}_{\mathrm{can}}}(\ell) \in (\mathbf{k}^n)^\mathsf{T} ;$$

les coordonnées $\mathrm{Mat}_{\mathcal{B}_F^\star}(\ell)$ de ℓ dans $\mathcal{B}_F^\star$ sont la colonne $L^\mathsf{T} \in \mathbf{k}^n$. Le vecteur $u^\mathsf{T}(\ell) \in E^\star$ est l'application linéaire $\ell \circ u \in \mathcal{L}(E, \mathbf{k})$, dont la matrice s'obtient par le produit matriciel

$$\mathrm{Mat}_{\mathcal{B}_E,\mathcal{B}_{\mathrm{can}}}(u^\mathsf{T}(\ell)) = LA \in (\mathbf{k}^n)^\mathsf{T}.$$

Les coordonnées $\mathrm{Mat}_{\mathcal{B}_E}(u^\mathsf{T}(\ell))$ de $u^\mathsf{T}(\ell)$ dans la base $\mathcal{B}_E^\star$ sont donc (c'est le lemme 4.1.8)

$$(LA)^\mathsf{T} = A^\mathsf{T} L^\mathsf{T} = A^\mathsf{T} \mathrm{Mat}_{\mathcal{B}_F^\star}(\ell),$$

ce qui prouve que $\mathrm{Mat}_{\mathcal{B}_F^\star,\mathcal{B}_E^\star}(u^\mathsf{T}) = A^\mathsf{T}$ comme on voulait.

4.1.11 Crochet de dualité. Si $\ell \in E^\star$ et $x \in E$, on note indifféremment $\ell(x)$ ou $\langle \ell, x \rangle$ la valeur de la forme ℓ évaluée en le vecteur x. Pour la deuxième notation on parle de "crochet de dualité". La notation est justifiée par la ressemblance avec le produit scalaire usuel, au niveau du calcul matriciel. On peut remarquer pour plus tard que l'expression $\langle \ell, x \rangle$ est "parfaitement" symétrique en ℓ et x,[2] ce qui est l'idée centrale du théorème de bidualité 4.3.5.

2. certains esprits dérangés en profitent pour s'autoriser la notation $\langle x, \ell \rangle$, et pire $x(\ell)$.

4.2 – Orthogonalité

4.2.1 Définition. *Soit $A \subseteq E$ un sous-ensemble. L'*orthogonal *de A est l'ensemble*

$$A^\perp = \{\ell \in E^\star :\ \forall x \in A,\ \langle \ell, x\rangle = 0\}$$

des formes linéaires qui s'annulent sur A.

Dans cette définition il n'y a aucune contrainte sur le sous-ensemble A. En revanche, quelle que soit la forme de A son orthogonal $A^\perp$ est toujours un sous-espace vectoriel de $E^\star$:

4.2.2 Proposition. *Soit $A \subseteq E$. L'orthogonal de A coïncide avec l'orthogonal de* $\mathrm{Vect}(A)$*, et est un sous-espace vectoriel de $E^\star$ de codimension* $\dim(\mathrm{Vect}(A))$*.*

Preuve. L'inclusion $\mathrm{Vect}(A)^\perp \subseteq A^\perp$ s'obtient directement par définition de l'orthogonal, puisque $A \subseteq \mathrm{Vect}(A)$. Réciproquement, soit $\ell \in A^\perp$. Par définition le noyau de ℓ contient A, donc il contient aussi $\mathrm{Vect}(A)$, puisque $\ker(\ell)$ est un sous-espace vectoriel et $\mathrm{Vect}(A)$ est le plus petit sous-espace vectoriel de E contenant A. Ainsi $\ell \in \mathrm{Vect}(A)^\perp$, et on a l'inclusion $A^\perp \subseteq \mathrm{Vect}(A)^\perp$, ce qui achève la preuve de l'égalité $\mathrm{Vect}(A)^\perp = A^\perp$.

Considérons une base $(e_1, \dots, e_r)$ de $\mathrm{Vect}(A)$, et complétons la en une base $(e_1, \dots, e_r, \dots, e_n)$ de E. On a $\mathrm{Vect}(A) = \mathrm{Vect}(e_1, \dots, e_r)$, donc

$$A^\perp = \mathrm{Vect}(e_1, \dots, e_r)^\perp = \{e_1, \dots, e_r\}^\perp = \big\{\ell \in E^\star :\ \ell(e_1) = \dots = \ell(e_r) = 0\big\}.$$

Soit $\ell \in E^\star$, de coordonnées $(\alpha_1, \dots, \alpha_n)$ dans la base duale $(e_1^\star, \dots, e_n^\star)$. On a $\alpha_i = \ell(e_i)$ d'après le lemme 4.1.3, donc

$$\ell(e_1) = \dots = \ell(e_r) = 0 \quad \Longleftrightarrow \quad \alpha_1 = \dots = \alpha_r = 0$$

et finalement

$$A^\perp = \mathrm{Vect}(e_{r+1}^\star, \dots, e_n^\star)$$

est un sous-espace vectoriel de $E^\star$ de dimension $n - r$ comme il fallait démontrer. □

4.2.3 Lemme. *Soit E un espace vectoriel de dimension finie. Pour tout $A, B \subseteq E$, on a :*
1 $(A \cup B)^\perp = A^\perp \cap B^\perp$ *;*
2 $A^\perp + B^\perp = (A \cap B)^\perp$*.*

Preuve. La première identité est automatique :

$$(A \cup B)^\perp = \{x \in E :\ \forall \ell \in A, \ell(x) = 0 \text{ et } \forall \ell \in B, \ell(x) = 0\} = A^\perp \cap B^\perp.$$

La seconde ne l'est pas, et d'ailleurs elle est en général fausse en dimension infinie . On a toujours $A^\perp + B^\perp \subseteq (A \cap B)^\perp$: en effet pour tout $\ell_1 \in A^\perp$ et $\ell_2 \in B^\perp$, ℓ_1 et ℓ_2 s'annulent toutes les deux sur $A \cap B$ donc $\ell_1 + \ell_2 \in (A \cap B)^\perp$.

On va conclure à l'égalité par un argument de dimension. On a

$$\mathrm{codim}(A^\perp + B^\perp) = \mathrm{codim}(A^\perp) + \mathrm{codim}(B^\perp) - \mathrm{codim}(A^\perp \cap B^\perp)$$

par la formule de Grassmann. Par la première identité, on a $A^\perp \cap B^\perp = (A \cup B)^\perp$. Donc d'après la proposition 4.2.2,

$$\begin{aligned}\mathrm{codim}(A^\perp + B^\perp) &= \dim\big(\mathrm{Vect}(A)\big) + \dim\big(\mathrm{Vect}(B)\big) - \dim\big(\mathrm{Vect}(A \cup B)\big)\\ &= \dim\big(\mathrm{Vect}(A \cap B)\big)\\ &= \mathrm{codim}(A \cap B)^\perp,\end{aligned}$$

en utilisant à nouveau la formule de Grassmann et la proposition 4.2.2 pour les deuxième et troisième égalité. On a donc finalement $\dim(A^\perp + B^\perp) = \dim(A \cap B)^\perp$, ce qui conclut la preuve. □

4.3 – Bidualité

4.3.1 Définition. *Soit Λ un sous-ensemble de $E^\star$. L'*ensemble des zéros *de Λ est*

$$\Lambda^o = \{x \in E : \ \forall \ell \in \Lambda, \ \ell(x) = 0\}.$$

Il s'agit de la notion duale de la notion d'orthogonal, en conséquence les ensembles de zéros jouissent de propriétés analogues à celles des orthogonaux. En particulier, on peut démontrer en utilisant le pivot de Gauss que pour tout $\Lambda \subseteq E^\star$, l'ensemble des zéros Λ^o est un sous-espace vectoriel de E de codimension $\dim(\mathrm{Vect}(\Lambda))$.

4.3.2 Exemple. Soit $\ell_1, \ldots, \ell_r \in E^\star$, et prenons $\Lambda = \{\ell_1, \ldots, \ell_r\}$. Dans ce cas Λ^o est le sous-espace de E défini par les équations $\ell_1(x) = \ldots = \ell_r(x)$, de codimension $\mathrm{rg}(\ell_1, \ldots, \ell_r)$ dans E.

En fait nous allons utiliser le théorème de bidualité 4.3.5 pour identifier définitivement ces deux notions, et nous n'aurons alors plus rien à faire pour appliquer les résultats valables pour les orthogonaux aux ensembles de zéros. Une fois que ce sera fait, on n'utilisera plus jamais la notation Λ^o.

4.3.3 Définition. *Soit $\mathcal{B}' = (\ell_1, \ldots, \ell_n)$ une base de $E^\star$. Il existe une unique base $\mathcal{B}$ de E telle que $\mathcal{B}'$ soit la base $\mathcal{B}^\star$ duale de $\mathcal{B}$. La base $\mathcal{B}$ est dite* base antéduale *de $\mathcal{B}'$.*

Nous allons remettre à plus tard la démonstration de l'existence et l'unicité de la base antéduale. À nouveau, notre approche consistera à identifier les deux notions duales de base duale et base antéduale *via* le théorème de bidualité 4.3.5, puis à appliquer la construction de la base duale à $\mathcal{B}'$.

4.3.4 Exemple. Prenons $E = \mathbf{k}[X]_{n-1}$ l'espace vectoriel des polynômes de degré $\leqslant n-1$. Soit $x_1, \ldots, x_n \in \mathbf{k}$ des scalaires deux à deux distincts (en supposant que $\mathbf{k}$ soit suffisamment gros pour que ce soit possible). Pour tout $i = 1, \ldots, n$ on considère la forme linéaire ev_i d'évaluation en x_i :

$$\forall P \in E : \quad \mathrm{ev}_i(P) = P(x_i) \in \mathbf{k}.$$

La famille $\mathcal{B}' = (\mathrm{ev}_1, \ldots, \mathrm{ev}_n)$ est une base de $E^\star$. Pour s'en convaincre, une possibilité est d'écrire sa matrice dans la base duale de la base canonique $\mathcal{B} = (1, X, \ldots, X^{n-1})$ de E,

$$\mathrm{Mat}_{\mathcal{B}^\star}(\mathcal{B}') = \begin{pmatrix} 1 & \ldots & \ldots & 1 \\ x_1 & \ldots & \ldots & x_n \\ \vdots & & & \vdots \\ x_1^{n-1} & \ldots & \ldots & x_n^{n-1} \end{pmatrix},$$

dont on voit qu'elle est inversible en réalisant que c'est une matrice de Vandermonde.

Les polynômes interpolateurs de Lagrange permettent d'expliciter la base antéduale de $\mathcal{B}'$. Pour tout $i = 1, \ldots, n$, on pose

$$L_i = \frac{\prod_{j \neq i}(X - x_j)}{\prod_{j \neq i}(x_i - x_j)} \in E.$$

La définition est faite pour que $L_i(x_j) = \delta_{ij}$, on a donc $\mathcal{B}' = (L_1, \ldots, L_n)^\star$.

4.3.5 Théorème (de bidualité)**.** *Soit E un $\mathbf{k}$-espace vectoriel de dimension finie. Pour tout $x \in E$, on note ev_x la forme linéaire $\ell \in E^\star \mapsto \ell(x) \in \mathbf{k}$ définie sur le dual de E. L'application $x \in E \mapsto \mathrm{ev}_x \in (E^\star)^\star$ réalise un isomorphisme $E \cong (E^\star)^\star$.*

L'isomorphisme $E \cong (E^\star)^\star$ ci-dessus est canonique en ce qu'il ne dépend d'aucun choix (ni choix de base, ni choix de rien du tout). En fait il est tellement naturel qu'après les explications données aux paragraphes ci-dessous on ne distinguera plus jamais $(E^\star)^\star$ de E. Le crochet de dualité introduit en 4.1.11 permet de bien comprendre à quel point cet isomorphisme est naturel : étant donné $\ell \in E^\star$ et $x \in E$, les formes linéaires ℓ sur E et ev_x sur $E^\star$ s'écrivent respectivement $y \in E \mapsto \langle \ell, y\rangle$ et $\varphi \in E^\star \mapsto \langle \varphi, x\rangle$.

Le théorème 4.3.5 est fondamentalement de nature géométrique. Il est faux en général en dimension infinie, ou si on remplace les $\mathbf{k}$-espaces vectoriels par des modules sur un anneau commutatif.

Preuve. Puisque $\dim(E^\star) = \dim(E)$, il suffit de montrer que l'application linéaire $x \mapsto \mathrm{ev}_x$ est injective. Soit $x \in E$ non nul. On complète la famille (x) en une base $(x, e_2, \ldots, e_n)$ de E, et on considère la base duale $(x^\star, e_2^\star, \ldots, e_n^\star)$ de $E^\star$. On a $\mathrm{ev}_x(x^\star) = x^\star(x) = 1$, donc $\mathrm{ev}_x \neq 0$. Ceci prouve que le noyau de $x \mapsto \mathrm{ev}_x$ est réduit à $\{0\}$, ce qui achève la preuve. □

4.3.1 – Bidualité et orthogonalité

4.3.6 Lemme. *Soit $\Lambda \subseteq E^\star$. Les sous-espaces vectoriels Λ^o de E et $\Lambda^\perp$ de $(E^\star)^\star$ sont identifiés par l'isomorphisme $E \cong (E^\star)^\star$.*

Preuve. D'une part

$$\Lambda^o = \{x \in E :\ \forall \ell \in \Lambda,\ \ell(x) = 0\},$$

et d'autre part

$$\begin{aligned}\Lambda^\perp &= \{\varphi \in (E^\star)^\star :\ \forall \ell \in \Lambda,\ \varphi(\ell) = 0\}\\ &\cong \{x \in E :\ \forall \ell \in \Lambda,\ \mathrm{ev}_x(\ell) = 0\},\end{aligned}$$

où le dernier isomorphisme est donné par l'isomorphisme canonique $(E^\star)^\star \cong E$, qui dit que pour tout $\varphi \in (E^\star)^\star$, il existe un unique $x \in E$ tel que $\varphi = \mathrm{ev}_x$. Finalement, on a donc bien

$$\Lambda^\perp \cong \Lambda^o = \{x \in E :\ \forall \ell \in \Lambda,\ \langle \ell, x\rangle = 0\}.$$

□

Dorénavant nous n'utiliserons plus jamais la notation Λ^o, et verrons toujours $\Lambda^\perp$ (pour $\Lambda \subseteq E^\star$) comme un sous-espace de E.

4.3.7 Lemme. *Soit $A \subseteq E$. On a $(A^\perp)^\perp = \mathrm{Vect}(A)$.*

Preuve. On a automatiquement $A \subseteq (A^\perp)^\perp$, puisque par définition de $A^\perp$ pour tout $a \in A$ et $\ell \in A^\perp$ on a $\ell(a) = 0$. On a donc $\mathrm{Vect}(A) \subseteq (A^\perp)^\perp$ puisque $(A^\perp)^\perp$ est un sous-espace vectoriel de E, et ces deux sous-espaces sont nécessairement égaux car de même dimension d'après la Proposition 4.2.2. □

4.3.8 Remarque. En dimension infinie, l'inclusion $\mathrm{Vect}(A) \subseteq (A^\perp)^\perp$ peut être stricte, comme le montre l'exemple suivant. Prenant un peu d'avance sur la dualité par rapport à une forme quadratique, considérons la forme bilinéaire non-dégénérée sur $\mathbf{k}[X]$ standard

$$\langle a_d X^d + \cdots + a_0, b_d X^d + \cdots + b_0 \rangle = \sum a_i b_i,$$

et $F = \{P : P(1) = 0\}$. On vérifie que $F^\perp = \{0\}$ et donc $(F^\perp)^\perp = \mathbf{k}[X]$.

4.3.2 – Application de la bidualité à la construction de bases antéduales

4.3.9 Lemme. *Soit $\mathcal{B}$ une base de E. On a $(\mathcal{B}^\star)^\star = \mathcal{B}$ via l'isomorphisme $E \cong (E^\star)^\star$.*

Preuve. Notons $\mathcal{B} = (e_1, \ldots, e_n)$ et $\mathcal{B}^\star = (e_1^\star, \ldots, e_n^\star)$. Pour tout $i, j \in [\![1, n]\!]$ on a

$$\mathrm{ev}_{e_i}(e_j^\star) = \langle e_j^\star, e_i \rangle = \delta_{ij},$$

donc $(\mathrm{ev}_{e_1}, \ldots, \mathrm{ev}_{e_n})$ est la base duale de $(e_1^\star, \ldots, e_n^\star)$, comme il fallait démontrer. □

4.3.10 Corollaire. *Soit $\mathcal{B}' = (\ell_1, \ldots, \ell_n)$ une base de $E^\star$. Il existe une unique base $\mathcal{B}$ de E telle que $\mathcal{B}'$ soit la base $\mathcal{B}^\star$ duale de $\mathcal{B}$.*

Preuve. Soit $\mathcal{B} = (\mathcal{B}')^\star$. D'après le lemme précédent $\mathcal{B}^\star = \mathcal{B}'$, donc $\mathcal{B}$ est une base antéduale de $\mathcal{B}'$. D'autre part, pour toute base $\mathcal{B}_1$ antéduale de $\mathcal{B}'$, on a $\mathcal{B}_1 = (\mathcal{B}_1^\star)^\star = (\mathcal{B}')^\star = \mathcal{B}$, donc $\mathcal{B}$ est l'unique base antéduale de $\mathcal{B}'$, comme il fallait démontrer. □

4.3.3 – Involutivité de la transposition

4.3.11. Terminons par la preuve qui rend fou : soit E et F deux $\mathbf{k}$-espaces vectoriels de dimensions finies, et $u \in \mathcal{L}(E, F)$. Nous allons montrer que *via* $E \cong (E^\star)^\star$ on a $(u^\mathsf{T})^\mathsf{T} = u$. Il serait raisonnable d'écrire la matrice de u dans des bases de E et F, et d'utiliser le fait que pour une matrice M, $(M^\mathsf{T})^\mathsf{T} = M$ sans difficulté. Mais alors on ne devient pas fou, donc c'est tricher.

Par définition, pour $\varphi \in (E^\star)^\star$ on a $(u^\mathsf{T})^\mathsf{T}(\varphi) = \varphi \circ u^\mathsf{T}$. Il s'agit de voir ce que ça donne pour $\varphi = \mathrm{ev}_x$, $x \in E$: pour tout $\ell \in E^\star$,

$$\begin{aligned}(u^\mathsf{T})^\mathsf{T}(\mathrm{ev}_x)(\ell) &= (\mathrm{ev}_x \circ u^\mathsf{T})(\ell) \\ &= \mathrm{ev}_x(u^\mathsf{T}(\ell)) \\ &= \mathrm{ev}_x(\ell \circ u) \\ &= \ell(u(x)) \\ &= \mathrm{ev}_{u(x)}(\ell),\end{aligned}$$

donc on a $(u^\mathsf{T})^\mathsf{T}(\mathrm{ev}_x) = \mathrm{ev}_{u(x)}$ comme il fallait démontrer.

4.4 – Correspondance entre sous-espaces de E et de son dual

L'énoncé ci-dessous est simplement une synthèse de faits établis plus haut. Il explicite le fait qu'un sous-espace de E peut être considéré indifféremment que l'ensemble des vecteurs qu'il contient (ses *points*) ou comme l'ensemble des formes linéaires s'annulant sur lui (ses *équations*).

4.4.1 Proposition. *Soit $k \in [\![0, n]\!]$. On note $\mathbf{Gr}(k, E)$ l'ensemble des sous-espaces de dimension k de E, et $\mathbf{Gr}(n-k, E^\star)$ l'ensemble des sous-espaces de dimension $n-k$ de $E^\star$. Les applications*

$$F \in \mathbf{Gr}(k, E) \mapsto F^\perp \in \mathbf{Gr}(n-k, E^\star) \quad et \quad \Lambda \in \mathbf{Gr}(n-k, E^\star) \mapsto \Lambda^\perp \in \mathbf{Gr}(k, E)$$

sont bijectives et réciproques l'une de l'autre.

Preuve. Les applications sont bien définies en vertu de la proposition 4.2.2, et réciproques l'une de l'autre d'après le lemme 4.3.7. □

Cette proposition généralise l'énoncé suivant pour les hyperplans, qu'on a appris quand on était petit.
Soit H un hyperplan de E. Il existe une forme linéaire ℓ sur E telle que $H = \ker(\ell)$. Réciproquement, le noyau d'une forme linéaire non nulle est un hyperplan de E, et deux formes linéaires ont le même noyau si et seulement si elles sont proportionnelles.

À présent, on voit que le bon objet à considérer n'est pas une forme linéaire de noyau H, mais l'ensemble de toutes ces formes linéaires (plus 0, ou alors l'ensemble de toutes les formes linéaires dont le noyau contient H), qui est une droite vectorielle dans l'espace dual. Il n'y a plus qu'un pas à franchir pour considérer des espaces projectifs !

4.4.2 Exemple. Soit f une fonction à valeurs réelles définie sur un ouvert $U \subseteq \mathbf{R}^n$, différentiable en un point $p \in U$. Le noyau de la différentielle $df_p \in (\mathbf{R}^n)^\star$ est un hyperplan de $\mathbf{R}^n$, qui est la direction de l'espace tangent en p de l'hypersurface V d'équation $f(x) = f(p)$:

$$T_pV = p + \ker(df_p).$$

FAIRE UN DESSIN ?

4.4.3 Remarque. Caché dans la proposition 4.4.1 se trouve le fait que le rang d'une matrice est égal au rang de sa transposée. Contrairement à ce qu'on pourrait croire trop rapidement, cet énoncé n'est pas trivial et a un contenu mathématique non vide.

4.4.4 Proposition. *Soit une matrice $A \in \mathcal{M}_{mn}(\mathbf{k})$. On a $\mathrm{rg}(A) = \mathrm{rg}(A^\mathsf{T})$.*

Preuve. Les lignes $L_1, \ldots, L_m$ de A sont des formes linéaires sur $\mathbf{k}^n$, et on a

$$\{L_1, \ldots, L_m\}^\perp = \ker(A).$$

La codimension de $\{L_1, \ldots, L_m\}^\perp$ égale le rang de la famille $(L_1, \ldots, L_m)$ par la proposition 4.2.2. D'autre part, le théorème du rang nous dit que la codimension du noyau de A égale la dimension de l'image de A, c'est-à-dire le rang des colonnes de A. On en conclut que le rang des lignes de A égale le rang des colonnes de A, autrement dit le rang de A égale le rang de sa transposée. □

Nous proposons deux autres preuves de cet énoncé.

Preuve par pivot de Gauss. On a vu que A est équivalente à la matrice $J_r = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0) \in \mathcal{M}_{mn}(\mathbf{k})$, où r est le rang de A (théorème 1.2.5). Autrement dit il existe $P \in \mathrm{GL}_m(\mathbf{k})$ et $Q \in \mathrm{GL}_n(\mathbf{k})$ telles que $PAQ = J_r$. On en déduit que $Q^\mathsf{T} A^\mathsf{T} P^\mathsf{T} = J_r^\mathsf{T}$, et donc A^T est équivalente à $J_r^\mathsf{T} = \mathrm{diag}(1, \ldots, 1, 0, \ldots, 0) \in \mathcal{M}_{nm}(\mathbf{k})$, ce qui implique que $\mathrm{rg}(A^\mathsf{T}) = r = \mathrm{rg}(A)$. □

Preuve par calcul de mineurs. Le rang d'une matrice est la taille de son plus grand mineur non-nul (proposition 2.5.6). Puisque le déterminant d'une matrice est égal au déterminant de sa transposée, A et A^T ont les mêmes mineurs, et donc ont même rang. □

4.4.4.1 Application. À FINIR DE RÉDIGER ET COMPLÉTER. Calculer un système d'équation de $\mathrm{Vect}(u_1,\dots,u_r)$ connaissant les coordonnées de chaque u_i dans $\mathcal{B}$, sous l'hypothèse que les u_i sont linéairement indépendants. i) en trouvant un mineur non-nul dans la matrice des u_i ; ii) il est bon de parler du pivot dans ce contexte.

Une fois qu'on a $A_r \in \mathrm{GL}_r(\mathbf{k})$,

$$(4.4.4.1)\qquad \left(\begin{array}{c|c} & x_1 \\ A_r & \vdots \\ & x_r \\ \hline L_{r+1} & x_{r+1} \\ \vdots & \vdots \\ L_n & x_n \end{array}\right) \iff \forall i=1,\dots,n-r,\ \det\left(\begin{array}{c|c} & x_1 \\ A_r & \vdots \\ & x_r \\ \hline L_{r+i} & x_{r+i} \end{array}\right) = 0.$$

En effet, on cherche les équations d'un espace de dimension r. Les équations du côté droit de (4.4.4.1) sont $n-r$ équations linéaires indépendantes vérifiées par les vecteurs de notre espace. Elles suffisent pour définir notre sous-espace.

On retrouve essentiellement la formule pour l'équation d'un plan dans $\mathbf{k}^3$ avec le produit vectoriel de deux vecteurs générateurs.

Un autre bon point pour la géométrie projective : si on la connaît, la méthode ci-dessus permet aussi d'écrire les équations des sous-espaces affines.

4.4.4.2 Une autre application. À METTRE EN FORME POUR UN AGRÉGATIF. Pour l'instant j'écris en projectif. Pour avoir l'équation en $(x:y:z)$ de la droite passant par deux points $(a:b:c)$ et $(a':b':c')$, on écrit

$$\begin{vmatrix} a & a' & x \\ b & b' & y \\ c & c' & z \end{vmatrix} = 0,$$

c'est l'application précédente. L'opération duale consiste à chercher les coordonnées du point d'intersection de deux droites. Si on considère à présent deux points du plan dual $(\alpha:\beta:\gamma),(\alpha':\beta':\gamma') \in \check{\mathbf{P}}^2$, représentant les équations de deux droites dans le plan initial $\mathbf{P}^2$, l'équation en $(\xi:\upsilon:\zeta) \in \check{\mathbf{P}}^2$ de la droite de $\check{\mathbf{P}}^2$ reliant ces deux points est

$$\begin{vmatrix} \alpha & \alpha' & \xi \\ \beta & \beta' & \upsilon \\ \gamma & \gamma' & \zeta \end{vmatrix} = 0 \iff \begin{vmatrix} \beta & \beta' \\ \gamma & \gamma' \end{vmatrix}\xi - \begin{vmatrix} \alpha & \alpha' \\ \gamma & \gamma' \end{vmatrix}\upsilon + \begin{vmatrix} \alpha & \alpha' \\ \beta & \beta' \end{vmatrix}\zeta = 0;$$

puisque l'équation de cette droite est donnée par les coordonnées du point d'intersection des deux droites $(\alpha:\beta:\gamma)^\perp,(\alpha':\beta':\gamma')^\perp \subseteq \mathbf{P}^2$, on en déduit :

$$\begin{cases} \alpha x + \beta y + \gamma z = 0 \\ \alpha' x + \beta' y + \gamma' z = 0 \end{cases} \iff \begin{pmatrix} x \\ y \\ z \end{pmatrix} \in \mathbf{k}.\begin{pmatrix} \begin{vmatrix} \beta & \beta' \\ \gamma & \gamma' \end{vmatrix} \\ -\begin{vmatrix} \alpha & \alpha' \\ \gamma & \gamma' \end{vmatrix} \\ \begin{vmatrix} \alpha & \alpha' \\ \beta & \beta' \end{vmatrix} \end{pmatrix}.$$

4.4.1 – Dualité entre image et noyau

4.4.5 Proposition. *Soit u une application linéaire entre deux espaces vectoriels E et F de dimensions finies. On a $\mathrm{im}(u^\mathsf{T}) = (\ker u)^\perp$ et $\ker(u^\mathsf{T}) = (\mathrm{im}(u))^\perp$.*

Preuve. Une nouvelle fois, on va montrer une inclusion (l'autre étant fausse en général en dimension infinie), puis conclure par un argument de dimension. Soit $\ell \in \mathrm{im}(u^\mathsf{T})$ et $x \in \ker u$. Il existe $\varphi \in F^\star$ telle que $\ell = u^\mathsf{T}(\varphi) = \varphi \circ u$. On a donc

$$\ell(x) = \varphi(u(x)) = \varphi(0) = 0,$$

et ainsi $\ell \in (\ker u)^\perp$. On en conclut que $\mathrm{im}(u^\mathsf{T}) \subseteq (\ker u)^\perp$.

La dimension de $\mathrm{im}(u^\mathsf{T})$ est le rang de u^T, qui est égal au rang de u d'après le lemme 4.1.8 (dans des bases duales, les matrices de u et u^T sont transposées l'une de l'autre) et la proposition 4.4.4 ci-dessus. D'autre part, $(\ker u)^\perp$ est de dimension $n - \dim(\ker u)$ d'après la proposition 4.2.2, ce qui égale $\mathrm{rg}(u)$ d'après le théorème du rang. Ainsi $\mathrm{im}(u^\mathsf{T})$ et $(\ker u)^\perp$ ont la même dimension, donc sont égaux puisque l'un est inclus dans l'autre.

On obtient l'autre identité en appliquant celle qu'on vient de montrer à u^T, en tirant parti de 4.3.11 :

$$\mathrm{im}(u) = \mathrm{im}((u^\mathsf{T})^\mathsf{T}) = (\ker(u^\mathsf{T}))^\perp,$$

puis en prenant l'orthogonal

$$(\mathrm{im}(u))^\perp = (\ker(u^\mathsf{T})^\perp)^\perp = \ker(u^\mathsf{T})$$

comme on voulait. □

4.4.6 Corollaire. *Si u est injective (resp. surjective), alors u^T est surjective (resp. injective).*

Preuve. Si u est injective, alors $\ker u = \{0\}$ donc $\mathrm{im}(u^\mathsf{T}) = 0^\perp = E^\star$, et u^T est surjective. Si u est surjective, alors $\ker u^\mathsf{T} = (\mathrm{im}\, u)^\perp = E^\perp = \{0\}$, donc u^T est injective. □

4.4.7 Dualité et quotients. Dans ce paragraphe, on considère F sous-espace vectoriel de E. On a une identification canonique

$$(E/F)^\star \cong F^\perp.$$

En effet, c'est un cas particulier de l'identification canonique entre $\mathcal{L}(E/F, G)$ et le sous-espace de $\mathcal{L}(E, G)$ des applications linéaires nulles sur F, déjà vu dans le chapitre d'algèbre linéaire. Dans cet isomorphisme, la transposée de la projection canonique $\pi : E \twoheadrightarrow E/F$ s'identifie à l'inclusion canonique $\iota : F^\perp \hookrightarrow E^\star$.

En géométrie projective, ceci se traduit par le fait que l'opération duale de 'projeter depuis un point' est 'prendre une section hyperplane'.

Chapitre 5

Sommes directes

5.1 – Composantes selon une somme directe

5.1.1 Définition. *Soit $E_1, \dots, E_r$ des sous-espaces vectoriels d'un espace vectoriel E. On dit que $E_1, \dots, E_r$ sont en* somme directe*, ou que la somme $\sum_{i=1}^r E_i$ est directe, si la condition suivante est vérifiée :*

$$(5.1.1.1) \quad \forall (x_1, \dots, x_r) \in E_1 \times \cdots \times E_r : \quad x_1 + \cdots + x_r = 0 \iff x_1 = \cdots = x_r = 0.$$

On laisse au lecteur le soin de vérifier que la condition (5.1.1.1) est équivalente à la condition suivante :

$$(5.1.1.2) \quad \forall (x_1, \dots, x_r), (x'_1, \dots, x'_r) \in E_1 \times \cdots \times E_r : \\ x_1 + \cdots + x_r = x'_1 + \cdots + x'_r \iff x_1 = x'_1, \ \dots, \ x_r = x'_r.$$

Ainsi, la somme $\sum_{i=1}^r E_i$ est directe si et seulement si pour tout vecteur $x \in \sum_{i=1}^r E_i$, il existe un unique r-uplet $(x_1, \dots, x_r) \in E_1 \times \cdots \times E_r$ tel que $x = x_1 + \cdots + x_r$. Dans ces conditions, pour tout $i_0 = 1, \dots, r$, on dit que le vecteur $x_{i_0} \in E_{i_0}$ est la *composante de x selon E_{i_0} relativement à la somme directe $\bigoplus_{i=1}^r E_i$*. Souvent, on omettra de dire "relativement à la somme directe $\bigoplus_{i=1}^r E_i$" ; il est important de garder à l'esprit qu'il s'agit d'un abus de langage.

5.1.2 Proposition. *Soit $E_1, \dots, E_r$ des sous-espaces vectoriels de E tels que $E = \bigoplus_{i=1}^r E_i$. Soit $i_0 \in [\![1, r]\!]$. L'application $\tilde{p}_{i_0}$ qui à tout vecteur $x \in E$ associe sa composante selon E_{i_0} relativement à la somme directe $\bigoplus_{i=1}^r E_i$ est une application linéaire de E dans E_{i_0}.*

5.1.3. Dans l'énoncé ci-dessus, on sera attentif au fait que l'hypothèse "$E = \bigoplus_{i=1}^r E_i$" affirme deux choses : (i) $E = \sum_{i=1}^r E_i$, autrement dit tout vecteur $x \in E$ peut s'écrire comme somme de vecteurs $x_1 \in E_1, \dots, x_r \in E_r$, et (ii) la somme $\sum_{i=1}^r E_i$ est directe. La condition "$E = \bigoplus_{i=1}^r E_i$" est donc équivalente à la condition suivante :

$$\forall x \in E : \quad \exists ! (x_1, \dots, x_r) \in E_1 \times \cdots \times E_r \quad \text{tel que} \quad x = x_1 + \cdots + x_r.$$

Ceci peut aussi s'exprimer de la façon suivante, qui éclaire d'un jour intéressant la notion de somme directe : $E = \bigoplus_{i=1}^r E_i$ vaut si et seulement si l'application linéaire

$$(x_1, \dots, x_r) \in E_1 \times \cdots \times E_r \longmapsto x_1 + \cdots + x_r \in E$$

est un isomorphisme.

Une nouvelle fois, le lecteur auquel est destiné ce livre devrait être en mesure d'établir lui-même la proposition 5.1.2, et nous lui conseillons vivement de le faire.

Avant d'aller plus loin, nous rappelons un critère pratique commode pour démontrer qu'une somme est directe.

5.1.4 Exercice. Soit $E_1, \dots, E_r$ des sous-espaces vectoriels d'un espace vectoriel E. Montrer que les conditions suivantes sont équivalentes :

(i) la somme $\sum_i E_i$ est directe ;
(ii) pour tout $i = 1, \dots, r$, les deux sous-espaces E_i et $\sum_{j\neq i} E_j$ sont en somme directe ;
(iii) pour tout $i = 1, \dots, r$, l'intersection $E_i \cap \left(\sum_{j\neq i} E_j\right)$ est réduite à $\{0\}$.

Nous poursuivons en expliquant comment une application linéaire se décompose relativement à des décompositions en sommes directes des espaces d'arrivée et de départ. Le résultat suivant est analogue à l'énoncé bien connu disant qu'étant donné deux espaces vectoriels E et F munis de bases $\mathcal{B}$ et $\mathcal{D}$ respectivement, une application linéaire f est équivalente à la donnée pour tout $e \in \mathcal{B}$ des coordonnées de $f(e)$ dans la base $\mathcal{D}$, autrement dit à la donnée de sa matrice dans les bases $\mathcal{B}$ et $\mathcal{D}$.

5.1.5 Proposition. *Soit E, F deux espaces vectoriels, munis respectivement des décompositions en sommes directes $E = \bigoplus_{j=1}^{r} E_j$ et $F = \bigoplus_{i=1}^{s} F_i$.*

5.1.5.1. *Soit $f \in \mathcal{L}(E, F)$. Pour tout $(i, j) \in [\![1, s]\!] \times [\![1, r]\!]$, l'application f_{ij} qui à $x \in E_j$ associe la composante de $f(x)$ selon F_i relativement à la somme directe $F_1 \oplus \cdots \oplus F_s$ est une application linéaire $f_{ij} \in \mathcal{L}(E_j, F_i)$.*

5.1.5.2. *Pour toute famille d'applications linéaires $(\tilde{f}_{ij}) \in \prod_{(i,j)\in[\![1,s]\!]\times[\![1,r]\!]} \mathcal{L}(E_j, F_i)$, il existe une unique application linéaire $f \in \mathcal{L}(E, F)$ telle que $f_{ij} = \tilde{f}_{ij}$, où les f_{ij} sont les applications associées à f comme en 5.1.5.1.*

On appellera *composantes de f selon les décompositions $E = \bigoplus_{j=1}^{r} E_j$ et $F = \bigoplus_{i=1}^{s} F_i$* les applications linéaires $f_{ij} \in \mathcal{L}(E_j, F_i)$ comme en 5.1.5.1 ci-dessus.

Preuve. Pour 5.1.5.1, il suffit d'observer que

$$f_{ij} = \tilde{p}_i \circ f|_{E_j} ,$$

où $\tilde{p}_i$ est l'application i-ème composante relativement à la décomposition $F = F_1 \oplus \cdots \oplus F_s$ comme en 5.1.2 : la restriction $f|_{E_j}$ est une application linéaire $E_i \to F$, et l'application i-ème composante $\tilde{p}_i$ est linéaire $F \to F_i$, donc la composition $\tilde{p}_i \circ f|_{E_j}$ est bien une application linéaire $E_j \to F_i$.

Pour 5.1.5.2, commençons par la partie "unicité". Soit $f, g \in \mathcal{L}(E, F)$ telles que $f_{ij} = g_{ij}$ pour tout $(i, j) \in [\![1, s]\!] \times [\![1, r]\!]$, où f_{ij} et g_{ij} sont définies à partir de f et g respectivement comme en 5.1.5.1. Il s'agit de démontrer que $f = g$. Soit $x \in E$, et écrivons le $x = x_1 + \cdots + x_r$, avec $(x_1, \dots, x_r) \in E_1 \times \cdots \times E_r$. On a

$$\begin{aligned} f(x) = f\Big(\sum_{j=1}^{r} x_j\Big) &= \sum_{j=1}^{r} f(x_j) \\ &= \sum_{j=1}^{r} \sum_{i=1}^{s} f_{ij}(x_j); \end{aligned} \tag{5.1.5.1}$$

l'égalité de droite sur la première ligne provient de la linéarité de f, et celle de la seconde ligne du fait que pour tout j, $f_{1j}(x_j), \dots, f_{sj}(x_j)$ sont les composantes de $f(x_j)$ selon la somme directe $F_1 \oplus \cdots \oplus F_s$. Ainsi, puisque $f_{ij} = g_{ij}$ pour tout (i,j), on a

$$\sum_{j=1}^{r}\sum_{i=1}^{s} f_{ij}(x_j) = \sum_{j=1}^{r}\sum_{i=1}^{s} g_{ij}(x_j),$$

et donc $f(x) = g(x)$, ce qui conclut la preuve du fait que $f = g$.

Enfin, démontrons la partie existence de 5.1.5.2. La formule (5.1.5.1) ci-dessus nous dit comment définir f, cette définition étant sans ambiguïté puisque $x_1, \dots, x_r$ sont uniquement déterminés par x par (5.1.1.2). Une autre façon de formuler les choses est de dire que l'application linéaire

$$f = \sum_{j=1}^{r}\sum_{i=1}^{s} \iota_i \circ \tilde{f}_{ij} \circ \tilde{q}_j$$

convient, où $\tilde{q}_j \in \mathcal{L}(E, E_j)$ est l'application j-ème composante relativement à la décomposition $E = E_1 \oplus \cdots \oplus E_r$, et $\iota_i \in \mathcal{L}(F_i, F)$ est l'injection canonique $F_i \hookrightarrow F$. Noter que pour tout (i,j), $\iota_i \circ \tilde{f}_{ij} \circ \tilde{q}_j \in \mathcal{L}(E,F)$, donc la somme de toutes ces applications linéaires est bien définie. □

5.1.6 Application : projecteurs. Considérons le cas particulier de la Proposition 5.1.5 où $E = F$ est muni d'une seule décomposition $E = \bigoplus_{i=1}^{r} E_i$. Pour tout $i_0 \in [\![1, r]\!]$, il existe un unique endomorphisme $p_{i_0} \in \mathcal{L}(E)$ tel que

$$\forall x \in E_{i_0} :\ p_{i_0}(x) = x, \quad \text{et} \quad \forall x \in \bigoplus_{i \neq i_0} E_i :\ p_{i_0}(x) = 0. \tag{5.1.6.1}$$

On peut le voir en appliquant 5.1.5.2 à la collection $(\tilde{f}_{ij})_{(i,j)\in[1,r]^2}$ définie par

$$\tilde{f}_{ij} = \begin{cases} \mathrm{id}_{E_{i_0}} & \text{si } (i,j) = (i_0, i_0) \\ 0 & \text{sinon.} \end{cases}$$

L'endomorphisme $p_{i_0} \in \mathcal{L}(E)$ s'appelle le *projecteur sur E_{i_0} relativement à la décomposition* $E = \bigoplus_{i=1}^{r} E_i$. On notera la subtile différence entre $p_{i_0} \in \mathcal{L}(E)$ et l'application i_0-ème composante $\tilde{p}_{i_0} \in \mathcal{L}(E, E_{i_0})$: pour $x \in E$ décomposé en $x = x_1 + \cdots + x_r$ selon $E = \bigoplus_{i=1}^{r} E_i$, on a $p_{i_0}(x) = \tilde{p}_{i_0}(x) = x_{i_0}$, mais dans le premier cas x_{i_0} est considéré comme un vecteur de E, tandis que dans le second cas il est considéré comme un vecteur de E_{i_0} ; autrement dit on a la relation $p_i = \iota_i \circ \tilde{p}_i$, où ι_i désigne l'inclusion $E_i \hookrightarrow E$. Il suit des conditions (5.1.6.1) que le projecteur p_i dépend seulement de E_{i_0} et de son supplémentaire $\bigoplus_{i \neq i_0} E_i$, et non pas de toute la décomposition $E = \bigoplus_{i=1}^{r} E_i$.

5.1.7 Projecteurs, II. Soit F et G deux sous-espaces vectoriels supplémentaires dans E. Le *projecteur sur F dans la direction de G* est l'endomorphisme $p \in \mathcal{L}(E)$ défini par les conditions :

$$\forall x \in F :\ p(x) = x, \quad \text{et} \quad \forall x \in G :\ p(x) = 0.$$

On prendra bien garde au fait que le projecteur p dépend à la fois de F et de G, comme nous l'illustrons dans la figure ci-dessous [1] (on note p et q les projections sur F et G relativement à $F \oplus G$, et p' et q' les projections sur F et G' relativement à $F \oplus G'$).

1. admirez au passage l'illusion d'optique : nous vous jurons que si vous superposez les deux figures, vous verrez le vecteur x est bien le même à gauche et à droite.

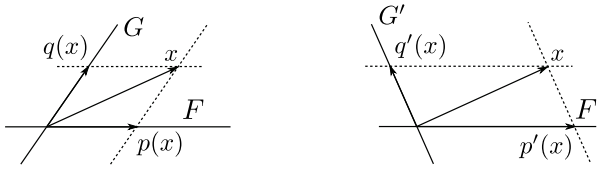


FIGURE 5.1 – Projections sur F dans les directions de G et G' respectivement

Un endomorphisme $f \in \mathcal{L}(E)$ est un *projecteur* s'il existe deux sous-espaces F et G supplémentaires dans E tels que f est le projecteur sur F dans la direction de G.

5.1.8 Proposition. *Un endomorphisme $p \in \mathcal{L}(E)$ est un projecteur si et seulement si $p \circ p = p$. Dans ce cas, c'est le projecteur sur* $\mathrm{im}(p)$ *dans la direction de* $\ker(p)$.

Preuve. Supposons $p \circ p = p$. Montrons que $E = \ker(p) \oplus \mathrm{im}(p)$. Pour tout $x \in E$, on a $x = x - p(x) + p(x)$ avec $x - p(x) \in \ker(p)$ et $p(x) \in \mathrm{im}(p)$. De plus si $y \in \ker(p) \cap \mathrm{im}(p)$, on a $y = p(x)$ pour un certain $x \in E$ et $0 = p(y) = (p \circ p)(x) = p(x) = y$. Ceci montrer que $E = \ker(p) \oplus \mathrm{im}(p)$, et par définition p est identiquement nul sur $\ker(p)$ et la relation $p \circ p = p$ implique que pour tout $x \in \mathrm{im}(p)$, $p(x) = x$. On conclut que p est le projecteur sur $\mathrm{im}(p)$ dans la direction de $\ker(p)$.

Réciproquement, supposons que p soit le projecteur sur F dans la direction d'un supplémentaire G. Soit $x \in E$, que l'on écrit $x = x_F + x_G$ avec $x_F \in F$, $x_G \in G$. Par définition, $p(x_F) = 0$ et $p(x_G) = x_G$, et on obtient

$$(p \circ p)(x)) = p(p(x_F) + p(x_G)) = p(0 + x_G) = x_G = p(x).$$

Ceci étant vrai pour tout $x \in E$, on conclut $p \circ p = p$. □

Enfin voici une autre propriété basique des projecteurs.

5.1.9 Lemme. *Considérons une décomposition $E = \bigoplus_{i=1}^{r} E_i$, et $p_1, \dots, p_r \in \mathcal{L}(E)$ les projecteurs associés. Alors on a*

$$p_1 + \cdots + p_r = \mathrm{id}_E.$$

Preuve. Soit $x \in E$. Pour $i \in [\![1, r]\!]$, notons $x_i = p_i(x) \in E_i \subseteq E$. Par définition $x = x_1 + \cdots + x_r$, et donc $x = p_1(x) + \cdots + p_r(x) = (p_1 + \cdots + p_r)(x)$. Ceci étant vrai pour tout $x \in E$, on conclut $\mathrm{id}_E = p_1 + \cdots + p_r$. □

5.2 – Matrices par blocs

Dans cette section, on interprète les composantes d'une application linéaire selon des décompositions des espaces de départ et d'arrivée en termes de matrices par blocs. Ceci permettra entre autre de donner une explication conceptuelle aux formules pour le produit de matrices par blocs.

5.2.1 Bases compatibles à une somme directe. Soit E un espace vectoriel dont on connaît une décomposition $E = E_1 \oplus \cdots \oplus E_r$. Soit $\mathcal{B}_1, \dots, \mathcal{B}_r$ des bases des sous-espaces vectoriels $E_1, \dots, E_r$ respectivement. Alors la famille de vecteurs $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_r)$ obtenue en concaténant $\mathcal{B}_1, \dots, \mathcal{B}_r$ est une base de E. Une base obtenue de la sorte sera dite *compatible à la décomposition $E = \bigoplus_i E_i$*.

5.2.2 Proposition. *Soit $E = \bigoplus_{j=1}^{r} E_j$, $F = \bigoplus_{i=1}^{s} F_i$, et $f \in \mathcal{L}(E,F)$. Dans des bases $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_r)$ et $\mathcal{D} = (\mathcal{D}_1, \dots, \mathcal{D}_s)$ compatibles aux décompositions de E et F respectivement, la matrice de f se décompose par blocs*

$$\mathrm{Mat}_{\mathcal{B},\mathcal{D}}(f) = \left(\begin{array}{c|c|c} M_{11} & \cdots & M_{1r} \\ \hline \vdots & & \vdots \\ \hline M_{s1} & \cdots & M_{sr} \end{array}\right)$$

où pour tout i, j, M_{ij} est la matrice dans les bases $\mathcal{B}_j$ et $\mathcal{D}_i$ de la composante $f_{ij} \in \mathcal{L}(E_j, F_i)$ de f.

La preuve de cet énoncé est essentiellement automatique, une fois assimilées les notions qu'il met en œuvre. Elle peut cependant s'avérer particulièrement désagréable si on fait des choix de notations malheureux ; nous utiliserons ainsi un double système d'indices, dont l'une des moitiés sera en fait en exposant.

Preuve. On note $\mathcal{B}_j = (e_j^1, \dots, e_j^{n_j})$ pour tout $j = 1, \dots, r$ (ainsi $n_1, \dots, n_r$ sont les dimensions respectives de $E_1, \dots, E_r$), et $\mathcal{D}_i = (\varepsilon_i^1, \dots, \varepsilon_i^{m_i})$ pour tout $i = 1, \dots, s$ (ainsi $m_1, \dots, m_s$ sont les dimensions respectives de $F_1, \dots, F_s$). On découpe la matrice $\mathrm{Mat}_{\mathcal{B},\mathcal{D}}(f)$ en blocs de tailles $m_i \times n_j$ comme indiqué ci-dessous.

$$\begin{array}{c} \\ m_1 \updownarrow \\ \\ m_s \updownarrow \end{array} \begin{array}{c} \overset{n_1}{\longleftrightarrow} \quad \overset{n_r}{\longleftrightarrow} \\ \left(\begin{array}{c|c|c} M_{11} & \cdots & M_{1r} \\ \hline \vdots & & \vdots \\ \hline M_{s1} & \cdots & M_{sr} \end{array}\right) \end{array}$$

Pour tout $(i,j) \in [\![1,s]\!] \times [\![1,r]\!]$, on note

$$M_{ij} = (a_{ij}^{kl})_{1 \leqslant k \leqslant m_i,\, 1 \leqslant l \leqslant n_j}.$$

On veut montrer que pour tout i, j, on a l'égalité $M_{ij} = \mathrm{Mat}_{\mathcal{B}_j,\mathcal{D}_i}(f_{ij})$. Il s'agit donc de montrer que pour tout $l \in [\![1, n_j]\!]$,

$$f_{ij}(e_j^l) = \sum_{k=1}^{m_i} a_{ij}^{kl}.\varepsilon_i^k.$$

Or on a, considérant la matrice de f dans les bases $\mathcal{B}$ et $\mathcal{D}$,

$$f(e_j^l) = \sum_{i=1}^{s} \sum_{k=1}^{m_i} a_{ij}^{kl}.\varepsilon_i^k.$$

Puisque pour tout $i = 1, \dots, s$, le vecteur $\sum_{k=1}^{m_i} a_{ij}^{kl}.\varepsilon_i^k$ appartient au sous-espace F_i, les composantes de $f(e_j^l)$ selon la somme directe $F = \bigoplus_i F_i$ sont les $\sum_{k=1}^{m_i} a_{ij}^{kl}.\varepsilon_i^k$, pour $i = 1, \dots, s$. En particulier la composante selon F_i est

$$f_{ij}(e_j^l) = \sum_{k=1}^{m_i} a_{ij}^{kl}.\varepsilon_i^k,$$

comme il fallait démontrer. □

La proposition suivante donne les composantes d'une application linéaire composée $g \circ f$ en fonction des composantes de f et g. On en déduira la formule pour le produit de deux matrices par blocs.

5.2.3 Proposition. *On considère trois espaces vectoriels décomposés en sommes directes $E = E_1 \oplus \cdots \oplus E_r$, $F = F_1 \oplus \cdots \oplus F_s$, $G = G_1 \oplus \cdots \oplus G_t$, et deux applications linéaires $f \in \mathcal{L}(E,F)$ et $g \in \mathcal{L}(F,G)$ décomposées respectivement en*

$$(f_{ij})_{1\leqslant i\leqslant s,\, 1\leqslant j\leqslant r} \quad \textit{et} \quad (g_{ij})_{1\leqslant i\leqslant t,\, 1\leqslant j\leqslant s}$$

selon ces sommes directes. Alors, pour tout $j = 1, \ldots, r$ et $i = 1, \ldots, t$, la composante $(g \circ f)_{ij} \in \mathcal{L}(E_j, G_i)$ de $g \circ f$ relativement aux décompositions de E et G est

$$(g \circ f)_{ij} = \sum\nolimits_{1\leqslant k\leqslant s} g_{ik} \circ f_{kj}.$$

Le calcul à faire pour démontrer cette formule est exactement le même que celui qui établit la formule

$$\mathrm{Mat}(g \circ f) = \mathrm{Mat}(g) \times \mathrm{Mat}(f),$$

et nous le laissons en exercice au lecteur. C'est ce qui explique les formules pour un produit de matrices par blocs, qui se résument en disant "qu'on peut prétendre que les blocs sont des scalaires". On prendra garde cependant au fait que, dans la formule ci-dessous, les produits de blocs $N_{ik}M_{kj}$ ne sont pas commutatifs en général.

5.2.4 Corollaire. *Soit*

$$M = (M_{ij})_{1\leqslant i\leqslant s,\, 1\leqslant j\leqslant r} \quad \textit{et} \quad N = (N_{ij})_{1\leqslant i\leqslant t,\, 1\leqslant j\leqslant s}$$

deux matrices par blocs, où M_{ij} est de taille $m_i \times n_j$ et N_{ij} est de taille $p_i \times m_j$. Alors le produit NM est bien défini, et s'écrit par blocs $NM = (R_{ij})_{1\leqslant i\leqslant t,\, 1\leqslant j\leqslant r}$, avec

$$R_{ij} = \sum_{k=1}^{s} N_{ik}M_{kj}$$

de taille $p_i \times n_j$.

$$\begin{array}{c} \overset{m_1}{\longleftrightarrow} \quad \overset{m_s}{\longleftrightarrow} \\ \left(\begin{array}{c|c|c} N_{11} & \cdots & N_{1s} \\ \hline \vdots & & \vdots \\ \hline N_{t1} & \cdots & N_{ts} \end{array}\right) \end{array}_{\begin{array}{l} p_1 \updownarrow \\ \\ p_t \updownarrow \end{array}} \times \begin{array}{c} \overset{n_1}{\longleftrightarrow} \quad \overset{n_r}{\longleftrightarrow} \\ \left(\begin{array}{c|c|c} M_{11} & \cdots & M_{1r} \\ \hline \vdots & & \vdots \\ \hline M_{s1} & \cdots & M_{sr} \end{array}\right) \end{array}_{\begin{array}{l} m_1 \updownarrow \\ \\ m_s \updownarrow \end{array}} = \begin{array}{c} \overset{n_1}{\longleftrightarrow} \quad \overset{n_r}{\longleftrightarrow} \\ \left(\begin{array}{c|c|c} R_{11} & \cdots & R_{1r} \\ \hline \vdots & & \vdots \\ \hline R_{t1} & \cdots & R_{tr} \end{array}\right) \end{array}_{\begin{array}{l} p_1 \updownarrow \\ \\ p_t \updownarrow \end{array}}$$

Ce corollaire résulte des deux propositions 5.2.2 et 5.2.3. Nous laissons la preuve au lecteur.

5.3 – Décomposition du déterminant relativement à une somme directe

Le résultat suivant est une version de la formule de Laplace pour le déterminant d'une famille de vecteurs dans une base. Nous allons en donner une preuve basée sur le théorème fondamental 2.6.1, qui est un peu moins exigeante techniquement que celle donnée à la section 2.4. En corollaire nous retrouverons la formule de Laplace pour les matrices (avec la petite réserve *a priori* qu'il faut des coefficients dans un corps).

5.3.1 Proposition. *On considère un espace vectoriel décomposé en somme directe $E = F \oplus G$, muni d'une base $\mathcal{B} = (\mathcal{B}_F, \mathcal{B}_G)$ compatible à la décomposition. On note p et q les projections sur F et G relativement à cette décomposition. On appelle n la dimension de E, r celle de F. Pour tout $(u_1, \dots, u_n) \in E^n$, on a*

$$(-1)^{\frac{r(r+1)}{2}} \det_{\mathcal{B}}(u_1, \dots, u_n) = \sum_{j_1 < \cdots < j_r} (-1)^{j_1 + \cdots + j_r} \det_{\mathcal{B}_F}(p(u_{j_1}), \dots, p(u_{j_r})) \det_{\mathcal{B}_G}(q(u_{j_{r+1}}), \dots, q(u_{j_n})),$$

où pour tout choix de $j_1 < \cdots < j_r$, les indices $j_{r+1} < \cdots < j_n$ sont définis par la condition

$$\{j_1 < \cdots < j_r\} \coprod \{j_{r+1} < \cdots < j_n\} = [\![1, n]\!].$$

Preuve. On considère la fonction $\Delta : E^n \to \mathbf{k}$ qui à tout n-uplet de vecteurs $(u_1, \dots, u_n)$ associe le scalaire

$$\sum_{j_1 < \cdots < j_r} (-1)^{j_1 + \cdots + j_r} \det_{\mathcal{B}_F}(p(u_{j_1}), \dots, p(u_{j_r})) \det_{\mathcal{B}_G}(q(u_{j_{r+1}}), \dots, q(u_{j_n})). \tag{5.3.1.1}$$

Nous allons montrer que Δ est une forme n-linéaire alternée qui prend la valeur $(-1)^{r(r+1)/2}$ sur la base $\mathcal{B}$, ce qui démontrera la formule. Le caractère n-linéaire est manifeste, et nous n'en dirons donc pas plus.

Calculons $\Delta(\mathcal{B})$. On note $\mathcal{B}_F = (e_1, \dots, e_r)$ et $\mathcal{B}_G = (e_{r+1}, \dots, e_n)$. On a

$$p(e_j) = \begin{cases} e_j & \text{si } j \leqslant r \\ 0 & \text{si } j \geqslant r+1 \end{cases} \quad \text{et} \quad p(e_j) = \begin{cases} 0 & \text{si } j \leqslant r \\ e_j & \text{si } j \geqslant r+1. \end{cases}$$

Ainsi le seul terme non nul dans la somme (5.3.1.1) pour $(u_1, \dots, u_n) = (e_1, \dots, e_n)$ correspond à

$$\{j_1 < \cdots < j_r\} = [\![1, r]\!].$$

On en déduit que $\Delta(\mathcal{B}) = (-1)^{r(r+1)/2}$, puisque $\det_{\mathcal{B}_F}(\mathcal{B}_F) = \det_{\mathcal{B}_G}(\mathcal{B}_G) = 1$.

Reste à démontrer que la forme n-linéaire Δ est alternée. Considérons $(u_1, \dots, u_n) \in E^n$ tel qu'il existe deux indices $a < b$ tels que $u_a = u_b$. Nous allons montrer que $\Delta(u_1, \dots, u_n) = 0$. Pour tout $J = \{j_1 < \cdots < j_r\}$, si a et b sont tous les deux dans J (respectivement, tous les deux dans le complémentaire de J), alors

$$\det_{\mathcal{B}_F}(p(u_{j_1}), \dots, p(u_{j_r})) = 0 \qquad (\text{respectivement, } \det_{\mathcal{B}_G}(q(u_{j_{r+1}}), \dots, q(u_{j_n})) = 0)$$

puisque $\det_{\mathcal{B}_F}$ et $\det_{\mathcal{B}_G}$ sont alternées. On en déduit que les seuls termes non-nuls dans la somme (5.3.1.1) sont ceux correspondant aux J tels que $J \cap \{a, b\}$ contient un unique

élément ; un tel J s'écrit ou bien $J_0 \cup \{a\}$, ou bien $J_0 \cup \{b\}$, pour J_0 une partie à $r-1$ éléments de $[\![1,n]\!] \setminus \{a,b\}$. La conclusion de tout ceci est que

$$\begin{aligned}(5.3.1.2)\quad &\Delta(u_1,\dots,u_n)\\ &= \sum_{J_0 \subseteq [\![1,n]\!]\setminus\{a,b\}} (-1)^{|J_0|}\Big[(-1)^a \det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{a\}\big)\cdot \det_{\mathcal{B}_G}\big(q(u_j), j\notin J_0\cup\{a\}\big)\\ &\qquad\qquad + (-1)^b \det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{b\}\big)\cdot \det_{\mathcal{B}_G}\big(q(u_j), j\notin J_0\cup\{b\}\big)\Big],\end{aligned}$$

où $|J_0| = \sum_{j\in J_0} j$, et par $\det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{a\}\big)$ on entend que les $p(u_j)$ sont rangés selon l'ordre croissant de leurs indices j.

Nous allons voir que pour chaque J_0 le terme entre crochet est nul. Puisque on suppose $u_a = u_b$, les vecteurs u_j pour $j \in J_0\cup\{a\}$ sont les mêmes dans leur ensemble que les vecteurs u_j pour $j \in J_0\cup\{b\}$. Ainsi les deux expressions

$$(5.3.1.3)\qquad \det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{a\}\big) \quad\text{et}\quad \det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{b\}\big)$$

diffèrent seulement par l'ordre dans lequel apparaissent les vecteurs dans le déterminant, et de ce fait ces deux déterminants sont égaux au signe près. Pour expliciter ce signe, appelons k la position dans laquelle a apparaît dans $J_0\cup\{a\}$ selon l'ordre croissant pour les indices, et k' la position dans laquelle apparaît b dans $J_0\cup\{b\}$; puisque $a<b$, on a $k \leqslant k'$. En permutant les entrées du second déterminant de (5.3.1.3) selon le $(k'-k+1)$-cycle $(k\ k+1\ \cdots\ k')$, on obtient le premier déterminant ;[2] on a donc

$$\det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{b\}\big) = (-1)^{k'-k}\det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{a\}\big)$$

puisque $\det_{\mathcal{B}_F}$ est alterné. On obtient de la même façon

$$\det_{\mathcal{B}_G}\big(q(u_j), j \notin J_0\cup\{b\}\big), = (-1)^{l-l'}\det_{\mathcal{B}_G}\big(q(u_j), j \notin J_0\cup\{a\}\big)$$

où l et l' sont les positions respectives de b et a dans $[\![1,n]\!]\setminus\big(J_0\cup\{a\}\big)$ et $[\![1,n]\!]\setminus\big(J_0\cup\{b\}\big)$; puisque $a<b$, on a $l \geqslant l'$. Le résultat est que pour tout J_0, le terme entre crochets dans la somme au second membre de (5.3.1.2) est

$$(5.3.1.4)\qquad \big[(-1)^a + (-1)^b(-1)^{k'-k}(-1)^{l-l'}\big]\det_{\mathcal{B}_F}\big(p(u_j), j \in J_0\cup\{a\}\big)\cdot \det_{\mathcal{B}_G}\big(q(u_j), j\notin J_0\cup\{a\}\big)),$$

2. écrivons ceci de manière plus formelle, dans l'idée que ceci pourrait contre toute attente aider certains lecteurs à suivre l'argument. Écrivant

$$J_0\cup\{a\} = \{j_1 < \cdots < j_k = a < \cdots < j_r\} \quad\text{et}\quad J_0\cup\{b\} = \{j'_1 < \cdots < j'_{k'} = b < \cdots < j'_r\},$$

on a

$$j'_t = \begin{cases} j_t & \text{si } t<k\\ j_{t+1} & \text{si } k\leqslant t<k'\\ j_t & \text{si } k'<t\end{cases}$$

soit $j'_t = j_{\sigma(t)}$ pour tout $t\neq k'$ en posant $\sigma = (k\ k+1\ \cdots\ k') \in \mathfrak{S}_r$. On a donc $u_{j'_t} = u_{j_{\sigma(t)}}$ pour $t\neq k'$; pour $t=k'$ d'autre part, puisque $u_a = u_b$,

$$u_{j'_{k'}} = u_b = u_a = u_{j_k} = u_{j_{\sigma(k')}},$$

si bien que finalement

$$\big(u_{j_{\sigma(1)}},\dots,u_{j_{\sigma(r)}}\big) = (u_{j'_1},\dots,u_{j'_r}).$$

et nous allons voir que les deux puissances de -1 dans le facteur entre crochets ont des signes opposés.

La différence $k' - k$ est le nombre d'entiers dans J_0 compris entre a et b. De même, la différence $l - l'$ est le nombre d'entiers dans $[\![1, n]\!] \setminus (J_0 \cup \{a, b\})$ compris entre a et b. On en déduit que $k' - k + l - l'$ est le nombre d'entiers dans $[\![1, n]\!] \setminus (\{a, b\})$ compris entre a et b ; autrement dit,

$$k' - k + l - l' = b - a - 1.$$

Le facteur entre crochets dans (5.3.1.4) est donc nul comme nous l'avions annoncé, et ceci conclut la preuve. □

5.3.2 Corollaire. *Soit $A = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbf{k})$. Pour tout $I = \{i_1 < \cdots < i_p\}$, on a :*

$$\det(A) = \sum_{J = \{j_1 < \cdots < j_p\}} (-1)^{|I| + |J|} \det(A_{IJ}) \det(A_{\bar{I}\bar{J}}),$$

où $|J| = j_1 + \cdots + j_p$.

5.4 – Quotients

Il sera commode d'utiliser la notion de quotient pour raisonner par récurrence dans le cadre de la réduction des endomorphismes. Ici nous donnons quelques rappels autour de cette notion.

5.4.1 Quotient par un sous-espace vectoriel. Soit E un $\mathbf{k}$-espace vectoriel et F un sous-espace vectoriel de E. $(E, +)$ est un groupe abélien, et $(F, +)$ est un sous-groupe (distingué, forcément) de $(E, +)$. On a donc un groupe quotient $(E/F, +)$. L'opération de multiplication par les scalaires définie par

$$\forall \bar{x} \in E,\ \forall \lambda \in \mathbf{k} :\ \lambda.\bar{x} = \overline{\lambda.x}$$

munit E/F d'une structure de $\mathbf{k}$-espace vectoriel canoniquement induite par celle de E. (On laisse toutes les vérifications élémentaires mais nécessaires au lecteur).

L'application $x \in E \mapsto \overline{x} \in E/F$ est une application linéaire surjective, dont le noyau est F. On l'appelle la *projection canonique*.

5.4.2 Lemme. *Soit F un sous-espace vectoriel de E, et considérons une base $\mathcal{B} = (e_1, \ldots, e_p, \ldots, e_n)$ telle que $(e_1, \ldots, e_p)$ est une base de F. Alors $(\overline{e}_{p+1}, \ldots, \overline{e}_n)$ est une base de E/F.*

Une base de E satisfaisant à l'hypothèse du lemme ci-dessus sera dite *compatible à F*. Étant donné une base compatible à F, les bases $(e_1, \ldots, e_p)$ de F et $(\overline{e}_{p+1}, \ldots, \overline{e}_n)$ de E/F comme ci-dessus seront dites *induites* par $\mathcal{B}$.

Preuve. Montrons que la famille $(\bar{e}_{p+1}, \ldots, \bar{e}_n)$ est libre. Soit $\lambda_{p+1}, \ldots, \lambda_n \in \mathbf{k}$ tels que $\lambda_{p+1}.\bar{e}_{p+1} + \cdots + \lambda_n.\bar{e}_n = 0$. Alors

$$\overline{\lambda_{p+1}.e_{p+1} + \cdots + \lambda_n.e_n} = \lambda_{p+1}.\overline{e_{p+1}} + \cdots + \lambda_n.\overline{e_n} = 0,$$

donc $\lambda_{p+1}.e_{p+1} + \cdots + \lambda_n.e_n \in F$. Puisque $(e_1, \ldots, e_p)$ est une base de F, il existe donc des scalaires $\mu_1, \ldots, \mu_p \in \mathbf{k}$ tels que

$$\begin{aligned} \lambda_{p+1}.e_{p+1} + \cdots + \lambda_n.e_n = \mu_1.e_1 + \cdots + \mu_p.e_p \\ \iff -\mu_1.e_1 - \cdots - \mu_p.e_p + \lambda_{p+1}.e_{p+1} + \cdots + \lambda_n.e_n = 0. \end{aligned}$$

Par liberté de la famille $(e_1, \ldots, e_p, \ldots, e_n)$, l'égalité de droite implique $\mu_1 = \cdots = \mu_p = \lambda_{p+1} = \cdots = \lambda_n = 0$. On a ainsi bien montré que $(\bar{e}_{p+1}, \ldots, \bar{e}_n)$ est libre.

Montrons que la famille $(\bar{e}_{p+1}, \ldots, \bar{e}_n)$ engendre E/F. Soit $\xi \in E/F$. On choisit un représentant $x \in E$ de ξ, que l'on écrit dans la base $\mathcal{B}$, $x = a_1.e_1 + \cdots + a_n.e_n$. Passant aux classes modulo F :

$$\xi = \overline{x} = a_1.\bar{e}_1 + \cdots + a_p.\bar{e}_p + a_{p+1}.\bar{e}_{p+1} + \cdots + a_n.\bar{e}_n = a_{p+1}.\bar{e}_{p+1} + \cdots + a_n.\bar{e}_n,$$

où la dernière égalité vient du fait que $\bar{e}_1 = \cdots = \bar{e}_p = 0$ puisque $e_1, \ldots, e_p \in F$. Ainsi tout $\xi \in E/F$ est combinaison linéaire de $\bar{e}_{p+1}, \ldots, \bar{e}_n$, et on a bien démontré que la famille $(\bar{e}_{p+1}, \ldots, \bar{e}_n)$ engendre E/F. □

5.4.3 Corollaire. *Le quotient E/F est de dimension finie, et*

$$\dim(E/F) = \dim E - \dim F.$$

Le cas $p = 2$ du résultat suivant est en quelque sort le lemme 5.4.2 "dans l'autre sens". Il sera commode d'avoir la version pour p arbitraire toute prête. Une suite de sous-espaces vectoriels $F_0, F_1, \ldots, F_p$ comme dans l'énoncé ci-dessous s'appelle une *filtration* de E.

5.4.4 Lemme. *Considérons une suite de sous-espaces vectoriels*

$$\{0\} = F_0 \subseteq F_1 \subseteq \cdots \subseteq F_p = E \ ;$$

on appelle une telle suite On considère des vecteurs

$$\begin{gathered} \varepsilon_{1,1}, \ldots, \varepsilon_{1,r_1} \in F_1 \\ \vdots \\ \varepsilon_{p,1}, \ldots, \varepsilon_{p,r_p} \in F_p \end{gathered}$$

tels que pour tout $i = 1, \ldots, p$ les classes $\bar{\varepsilon}_{1,1}, \ldots, \bar{\varepsilon}_{i,r_i}$ modulo F_{i-1} constituent une base de F_i/F_{i-1}. Alors la famille (ε_{ij}) est une base de E.

Preuve. Par récurrence sur p, il suffit de savoir le faire pour $p = 2$. Soit F un sous-espace vectoriel de E, $(a_1, \ldots, a_s)$ base de F, et $b_1, \ldots, b_r \in E$ tels que $(\bar{b}_1, \ldots, \bar{b}_r)$ soit une base du quotient E/F. Montrons que $(a_1, \ldots, a_s, b_1, \ldots, b_r)$ est une base de E.

Soit $\lambda_1, \ldots, \lambda_s, \mu_1, \ldots, \mu_r$ tels que

$$\lambda_1.a_1 + \cdots + \lambda_p.a_s + \mu_1.b_1 + \cdots + \mu_r.b_r = 0. \tag{5.4.4.1}$$

Après projection dans E/F, reste

$$\mu_1.\bar{b}_1 + \cdots + \mu_r.\bar{b}_r = 0,$$

qui implique $\mu_1 = \cdots = \mu_r = 0$. (5.4.4.1) devient alors une combinaison linéaire nulle de $e_1, \ldots, e_p$, qui à son tour donne $\lambda_1 = \cdots = \lambda_s = 0$. Ceci prouve la liberté.

D'autre part, soit $x \in E$. Sa projection $\bar{x}$ dans le quotient E/F s'écrit comme une combinaison linéaire

$$\bar{x} = \mu_1.\bar{b}_1 + \cdots + \mu_r.\bar{b}_r.$$

Ainsi $x - (\mu_1.b_1 + \cdots + \mu_r.b_r)$ est dans F, puisque sa classe modulo F est nulle, en conséquence de quoi il s'écrit comme combinaison linéaire de $a_1, \ldots, a_s$. □

5.4.1 – Quotients et supplémentaires

À présent, nous voulons expliquer les relations entre le quotient E/F et les supplémentaires de F dans E.

5.4.5 Lemme. *On considère F sous-espace vectoriel de E, et F' un supplémentaire de F dans E.*
(a) *La projection canonique $\pi : E \to E/F$ induit par restriction un isomorphisme $F' \xrightarrow{\cong} E/F$.*
(b) *L'application composante selon F' dans la décomposition $E = F \oplus F'$, $\tilde{p} : E \to F'$, induit par passage au quotient un isomorphisme $E/F \xrightarrow{\cong} F'$.*

Preuve. (a) La projection canonique $\pi : E \to E/F$ est surjective, de noyau F. Le noyau de la restriction $\pi|_{F'}$ est donc $F' \cap F = \{0\}$, et ainsi $\pi|_{F'}$ est injective. Montrons qu'elle est également surjective. Soit $\xi \in E/F$. On choisit $x \in E$ représentant ξ, et on l'écrit $x = x_F + x'_F$, avec $x_F \in F$ et $x'_F \in F'$. Alors

$$\xi = \overline{x} = \pi(x) = \pi(x_F + x'_F) = \pi(x'_F),$$

la dernière égalité provenant de la linéarité de π et du fait que $\pi(x_F) = 0$. Ainsi $\pi|_{F'}$ est surjective et inective, et nous avons démontré que c'est un isomorphisme $F' \xrightarrow{\cong} E/F$. □

5.4.6 Quotient et supplémentaire. Choisissons un supplémentaire F' de F dans E. On dispose de la projection $\tilde{p} : E \to F'$ dans la direction de F, qui est surjective et de noyau F. D'après la propriété universelle du quotient, il existe un isomorphisme canonique entre F' et E/F. Cette réalisation du quotient dépend d'un choix et n'est donc pas canonique.

Le choix d'un supplémentaire de F dans E équivaut au choix d'une injection linéaire $j : E/F \hookrightarrow E$ telle que $\mathrm{im}(j) \cap F = \{0\}$.

Il est important de noter qu'il n'existe pas de supplémentaire canonique à F (sauf à adjoindre des structures supplémentaires, par exemple une structure euclidienne). Cependant le bloc B de (6.4.1.1) a un sens canonique : il ne dépend que du F, et pas du choix du supplémentaire F'. Pour le voir, nous allons considérer le quotient E/F.

5.4.2 – Propriété universelle du quotient

Ce qui suit n'est pas strictement nécessaire pour la lecture des sections sur la réduction des endomorphismes. C'est cependant important du point de vue théorique, et nous l'utiliserons à l'occasion. Il n'est pas nécessaire ici de supposer E de dimension finie.

5.4.7 Proposition. *Soit E un espace vectoriel, et F un sous-espace vectoriel de E. Le quotient E/F jouit des deux propriétés suivantes :*
(a) il existe une application linéaire $\pi : E \to E/F$ surjective et de noyau F ;
(b) pour tout espace vectoriel G, l'application linéaire (c'en est une)

$$u \in \mathcal{L}(E/F, G) \mapsto u \circ \pi \in \mathcal{L}(E, G)$$

induit un isomorphisme (fonctoriel...)

$$\Phi : \mathcal{L}(E/F, G) \cong \ker\big(\mathrm{restr} : \mathcal{L}(E, G) \to \mathcal{L}(F, G)\big).$$

S'il existe E' jouissant lui aussi de l'une ou l'autre de ces deux propriétés, alors il existe un unique isomorphisme $\varphi : E/F \cong E'$ tel que $\pi' = \varphi \circ \pi$.

5.4.7.1 Preuves. On laisse au lecteur la preuve de la propriété (a) pour E/F et sa projection canonique. On va montrer que les propriétés (a) et (b) sont équivalentes, puis qu'un **k**-ev satisfaisant à ces propriétés s'identifie canoniquement à E/F.

Soit E' vérifiant (a) (i.e. on suppose E' **k**-ev muni de $\pi' : E \to E'$ linéaire, surjective et de noyau F), et montrons que (b) vaut pour E' (cela montrera au passage que (b) vaut bien pour le quotient). On commence par constater que $u \mapsto u \circ \pi'$ donne bien une application linéaire

$$\Phi_{\pi'} : \mathcal{L}(E', G) \to \ker\big(\text{restr} : \mathcal{L}(E, G) \to \mathcal{L}(F, G)\big)$$

pour tout **k**-ev G. Reste à voir que c'est effectivement un isomorphisme. Si $u \circ \pi' = 0$, alors $u = 0$ car π' est surjective, d'où l'injectivité de $\Phi_{\pi'}$.

Pour la surjectivité, on construit à la main un antécédent pour toute application linéaire $v : E \to G$ s'annulant sur F. Pour tout $y \in E'$ il existe $x \in E$ tel que $y = \pi'(x)$, et on pose $u(y) := v(x)$; ça ne dépend pas du choix de x car $\ker(\pi') = F$ et $v|_F = 0$. Ceci définit une application linéaire $u : E' \to G$ qui visiblement factorise comme il faut (i.e. $v = u \circ \pi'$). □

Réciproquement, soit E' vérifiant (b), et montrons directement que (a) vaut pour E' (à nouveau, cela montrera au passage que le quotient catégoriel est bien le quotient défini par la relation de congruence). Précisément, on suppose qu'il existe une application linéaire $\pi' : E \to E'$ tel que $\Phi_{\pi'}$ soit un isomorphisme pour tout G, et il s'agit de montrer que π' est surjective et de noyau F.

On regarde le $\Phi_{\pi'}$ pour $G = E'$: puisque $\pi' = \Phi_{\pi'}(\mathrm{id}_{E'})$, π' est dans le noyau de la restriction à F, donc $F \subseteq \ker(\pi')$. D'autre part, on regarde le $\Phi_{\pi'}$ pour $G = E/F$: puisque π est dans le noyau de la restriction à F, il existe un $u : E' \to E/F$ tel que $\pi = u \circ \pi'$. Ceci implique $\ker(\pi') \subseteq \ker(\pi) = F$, donc on conclut que $\ker(\pi') = F$.

D'autre part, si π' n'était pas surjective on pourrait contredire l'injectivité de $\Phi_{\pi'}$ pour n'importe quel $G \neq \{0\}$ de la manière suivante. Soit $x \in E'$ non atteint par π'. On considère G un **k**-ev non nul et $u' : E' \to G$. Alors pour g n'importe quel vecteur non nul de G et ℓ une forme linéaire de noyau un hyperplan transverse à x, on a

$$(u' + \ell.g) \circ \pi' = u' \circ \pi',$$

ce qui contredit l'injectivité de $\Phi_{\pi'}$. □

Soit E' vérifiant (a). Alors $\pi' \in \mathcal{L}(E, E')$ est dans le noyau de la restriction à F, donc il existe un unique $\varphi : E/F \to E'$ tel que $\pi' = \varphi \circ \pi$. Reste à voir que ce φ est un isomorphisme. Puisque π est surjective, $\ker \varphi \neq \{0\} \Rightarrow \ker \pi' \supsetneq F$ et donc nécessairement φ est injective. D'autre part φ est surjective car π' l'est. □

Chapitre 6

Réduction des endomorphismes : Introduction

6.1 – Qu'est-ce que la réduction des endomorphismes ?

6.1.1 Principe. On veut réduire l'étude d'un endomorphisme $f \in \mathcal{L}(E)$ à l'étude d'endomorphismes f_i définis sur des espaces vectoriels strictement plus petits que E. Concrètement, on va chercher à décomposer f en une somme directe (ou en un produit, voir 5.1.3) d'endomorphismes de sous-espaces stricts [1] de E.

Idéalement, on voudrait que chacun des endomorphismes de cette décomposition soit "atomique", c'est-à-dire suivant l'étymologie du terme qu'il soit impossible à casser en morceaux plus petits. On verra que ce n'est pas toujours possible, mais on s'attachera à donner des conditions caractérisant l'existence d'une telle décomposition.

6.1.2 Somme directe d'endomorphismes. Commençons par préciser ce qu'on entend par "somme directe d'endomorphismes" (et mettons en garde le lecteur quant au fait que cette terminologie imagée n'est sans doute pas universellement reconnue).

On dira que f est une *somme directe d'endomorphismes* $f_1, \dots, f_r$, et on notera $f = f_1 \oplus \cdots \oplus f_r$, s'il existe une décomposition $E = \bigoplus_{i=1}^r E_i$ et des endomorphismes $f_i \in \mathcal{L}(E_i)$ pour tout $i = 1, \dots, r$, tels que pour tout $(x_1, \dots, x_r) \in E_1 \times \cdots \times E_r$,

$$f(x_1 + \cdots + x_r) = f_1(x_1) + \cdots + f_r(x_r).$$

6.1.3 Matrice diagonale par blocs. Pour expliciter cette notion de somme directe d'endomorphismes, il est bon d'en donner une version matricielle. La condition de 6.1.2 ci-dessus est équivalente à ce que les composantes $f_{ij} \in \mathcal{L}(E_j, E_i)$ de f selon la décomposition $E = \bigoplus_{i=1}^r E_i$ soient

$$f_{ij} = \begin{cases} f_i & \text{si } i = j \\ 0 & \text{si } i \neq j. \end{cases}$$

Lorsque cette condition est vérifiée, la matrice de f dans une base $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_r)$ compatible à la décomposition $E = \bigoplus_{i=1}^r E_i$ s'écrit

$$\begin{pmatrix} M_1 & & 0 \\ & \ddots & \\ 0 & & M_r \end{pmatrix},$$

1. Un sous-espace vectoriel $F \subseteq E$ est dit *strict* si $\{0\} \subsetneq F \subsetneq E$.

avec M_i la matrice de f_i dans la base $\mathcal{B}_i$ pour tout $i = 1, \dots, r$. Ainsi, une décomposition de f en somme directe d'endomorphismes correspond dans une base adaptée à une matrice diagonale par blocs.

La définition suivante est fondamentale pour la réduction des endomorphismes, et nous allons l'utiliser pour reformuler la condition d'existence d'une décomposition en somme directe.

6.1.4 Définition. *Soit $f \in \mathcal{L}(E)$. Un sous-espace vectoriel F de E est* stable par f *si pour tout $x \in F$, on a $f(x) \in F$.*

Étant donné une décomposition $E = \bigoplus_{i=1}^r E_i$ arbitraire, pour $j_0 \in [\![1, r]\!]$ le sous-espace E_{j_0} est stable par f si et seulement si parmi les composantes f_{ij} de f selon $E = \bigoplus_{i=1}^r E_i$ on a $f_{ij_0} = 0$ si $i \neq j_0$.

Si $f = f_1 \oplus \cdots \oplus f_r$ comme en 6.1.2 ci-dessus, chacun des E_i est stable par f. En fait, l'existence d'une décomposition $f = f_1 \oplus \cdots \oplus f_r$ équivaut à l'existence d'une décomposition $E = \bigoplus_{i=1}^r E_i$ où tous les E_i sont stables ; pour tout i, l'endomorphisme $f_i \in \mathcal{L}(E_i)$ est l'endomorphisme induit par f sur E_i au sens de la définition suivante.

(On observera au passage que sous l'hypothèse que E_i est stable, l'endomorphisme $f_i \in \mathcal{L}(E_i)$ est complètement déterminé par la donnée de E_i seulement, alors qu'en général la composante $f_{ii} \in \mathcal{L}(E_i)$ de f dépend de toute la décomposition $E = \bigoplus_j E_j$).

6.1.5 Définition. *Soit $f \in \mathcal{L}(E)$, et F un sous-espace de E stable par f. L'*endomorphisme induit *par f sur F est l'endomorphisme $f_F \in \mathcal{L}(F)$ défini par la condition*

$$\forall x \in F : \quad f_F(x) = f(x).$$

Une bonne façon de considérer la notion de sous-espace stable est d'observer que l'existence d'un "sous-endomorphisme" correspond à l'existence d'un sous-espace stable (le "sous-endomorphisme" étant alors l'endomorphisme induit).

Ceci dicte la définition d'endomorphisme "atomique". La terminologie classique veut qu'on les appelle plutôt "simples".

6.1.6 Définition. *Soit $f \in \mathcal{L}(E)$. L'endomorphisme f est* simple *si aucun sous-espace strict de E n'est stable par f.*

Ainsi, un endomorphisme est simple s'il ne possède pas de sous-endomorphisme non-trivial, de la même façon qu'un groupe est simple s'il ne possède pas de sous-groupe distingué non-trivial.

6.1.7 Exemples. Si E est de dimension 1, tout endomorphisme de E est simple.

Si $E = \mathbf{R}^2$, l'endomorphisme r_θ défini dans la base canonique par la matrice

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

est simple si $\theta \not\equiv 0 \bmod \pi$. Géométriquement c'est clair, puisque cet endomorphisme est une rotation d'angle θ ; cependant dans le contexte où nous nous trouvons il n'y a pas de produit scalaire, et donc pas d'angles. Voici une justification plus algébrique (qui anticipe un peu sur les notions que nous verrons plus tard) : puisque E est de dimension 2, tout

sous-espace stable strict est de dimension 1, donc contenu dans un sous-espace propre. Or le polynôme caractéristique de R_θ est

$$X^2 - (2\cos\theta)X + 1 = (X - e^{i\theta})(X - e^{-i\theta}),$$

donc R_θ n'a aucune valeur propre réelle. Ainsi r_θ ne peut pas avoir de sous-espace stable strict.

Si $E = \mathbf{C}^2$, l'endomorphisme $\tilde{r}_\theta$ défini dans la base canonique par la matrice R_θ n'est pas simple : les deux sous-espaces propres relatifs aux valeurs propres $e^{i\theta}$ et $e^{-i\theta}$ sont tous les deux des droites de $\mathbf{C}^2$ stables par $\tilde{r}_\theta$.

Cet exemple illustre en particulier que la simplicité n'est pas une notion invariante par extension des scalaires.

Comme annoncé plus haut, il est illusoire d'espérer pouvoir réduire en toute généralité un endomorphisme en somme d'endomorphismes simples. L'exemple suivant illustre les problèmes typiques qui se posent.

6.1.8 Exemple. Soit $E = \mathbf{k}^2$ et f l'endomorphisme donné dans la base canonique (e_1, e_2) par la matrice

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

Le seul sous-espace strict de E stable par f est la droite $\mathrm{Vect}(e_1)$. En effet, par le même argument qu'en 6.1.7 ci-dessus, tout sous-espace stable par f est contenu dans un sous-espace propre de f. Or son polynôme caractéristique est X^2, donc 0 est son unique valeur propre, et son seul sous-espace propre est $\ker(f) = \mathrm{Vect}(e_1)$.

Il est donc impossible de décomposer E en somme directe de sous-espaces stricts stables par f.

On donne un nom aux endomorphismes qu'il est possible de décomposer en sommes d'endomorphismes simples.

6.1.9 Définition. *Un endomorphisme $f \in \mathcal{L}(E)$ est* semi-simple *s'il est somme directe d'endomorphismes simples, c'est-à-dire s'il existe une décomposition $E = \bigoplus_{i=1}^r E_i$ en somme de sous-espaces stables par f telle que pour tout $i = 1, \ldots, r$, l'endomorphisme induit $f_{E_i} \in \mathcal{L}(E_i)$ est simple.*

L'endomorphisme de l'exemple 6.1.8 n'est pas semi-simple, et *a fortiori* pas simple non plus. Les endomorphismes $r_\theta \in \mathcal{L}(\mathbf{R}^2)$ et $\tilde{r}_\theta \in \mathcal{L}(\mathbf{C}^2)$ définis par la matrice de rotation R_θ de 6.1.7 sont tous les deux semi-simples.

Avec la définition ci-dessus, il est assez facile de voir que tout endomorphisme diagonalisable est semi-simple. Dans une large mesure, les endomorphismes semi-simples sont ceux qu'il est possible de diagonaliser quitte à étendre les scalaires (c'est le cas lorsque le corps de base est parfait, mais il existe en général des exemples pathologiques, voir section 8.3). Dans la section 8.3 nous donnons des conditions nécessaires et suffisantes pour la simplicité et la semi-simplicité portant sur les polynômes minimaux et caractéristiques.

Dans la littérature on trouve souvent une autre définition de la semi-simplicité, qui est la suivante. Nous allons voir dans la sous-section suivante que les deux versions sont équivalentes, même si ce n'est pas si évident à première vue.

6.1.10 Définition. *Un endomorphisme $f \in \mathcal{L}(E)$ est* semi-simple-bis *si pour tout sous-espace F stable par f, il existe un supplémentaire F' de F dans E qui est lui aussi stable par f.*

Nous conserverons ces deux notions de semi-simplicité et semi-simplicité-bis, car il est souvent instructif de les considérer chacune à part en mettant de côté le fait qu'elles sont équivalentes.

6.2 – Semi-simplicité et semi-simplicité-bis

Cette sous-section est consacrée à la preuve du théorème ci-dessous, et à divers compléments.

6.2.1 Théorème. *Les conditions de semi-simplicité et de semi-simplicité-bis sont équivalentes.*

Nous donnons ici une preuve élémentaire, mais il est aussi possible d'utiliser des outils plus élaborés de réduction des endomorphismes, voir la section 8.3.

Preuve du théorème 6.2.1. L'implication « semi-simple ⇒ semi-simple-bis » est le contenu de la proposition 6.2.4, et sa réciproque « semi-simple-bis ⇒ semi-simple » est le contenu du corollaire 6.2.3. □

6.2.2 Proposition. *Soit $f \in \mathcal{L}(E)$ un endomorphisme semi-simple-bis. Pour tout sous-espace F stable par E, l'endomorphisme induit $f_F \in \mathcal{L}(F)$ est lui aussi semi-simple-bis.*

Preuve. Soit $F_0 \subseteq F$ un sous-espace stable par f_F, ou de manière équivalente stable par f. Par semi-simplicité de f, il existe donc $\tilde{G}_0$ supplémentaire de F_0 dans E stable par f. Le sous-espace $\tilde{G}_0 \cap F$ est stable par f comme intersection de deux sous-espaces stables, et comme $\tilde{G}_0 \cap F \subseteq F$, ce même sous-espace est aussi stable par l'endomorphisme induit f_F. Pour conclure la preuve nous allons montrer que $G_0 := \tilde{G}_0 \cap F$ est un supplémentaire de F_0 dans F.

On a $F_0 \cap G_0 \subseteq F_0 \cap \tilde{G}_0 = \{0\}$, donc F_0 et G_0 sont en somme directe. Puisque F_0 et G_0 sont tous les deux contenus dans F, on a l'inclusion $F_0 \oplus G_0 \subseteq F$. Pour montrer l'inclusion inverse, considérons $x \in F$ et décomposons le selon $F_0 \oplus \tilde{G}_0 = E$: on écrit $x = y + y'$, avec $y \in F_0$ et $y' \in \tilde{G}_0$. Alors $y' = x - y \in F$, donc $y' \in \tilde{G}_0 \cap F = G_0$. Ceci prouve que $x \in F_0 + G_0$, et finalement $F = F_0 \oplus G_0$ comme attendu. □

6.2.3 Corollaire. *Soit $f \in \mathcal{L}(E)$ semi-simple-bis. Alors f est semi-simple.*

Preuve du corollaire. On raisonne par récurrence sur la dimension n de E. Si $n \leqslant 1$, f est lui-même simple et il n'y a rien à démontrer. Supposons donc $n > 1$ et le résultat démontré pour $\dim(E) < n$. Si f est simple, le résultat est trivial. Sinon, il existe F sous-espace strict de E stable par f. Puisque f est semi-simple, il existe F' supplémentaire de F dans E lui aussi stable par f. Puisque F est strict, on a $\dim(F) < n$ et $\dim(F') < n$. D'après la proposition 6.2.2, les endomorphismes induits f_F et $f_{F'}$ sont tous les deux semi-simples. On peut donc appliquer l'hypothèse de récurrence, qui nous dit qu'il existe deux décompositions $F = \bigoplus_i F_i$ et $F' = \bigoplus_i F'_i$ en sommes de sous-espaces stables par f_F et $f_{F'}$ respectivement, et donc stables par f, tels que les $(f_F)_{F_i} = f_{F_i}$ et $(f_{F'})_{F'_i} = f_{F'_i}$ sont simples. On a donc une décomposition $E = \bigoplus_i F_i \oplus \bigoplus_i F'_i$ comme on voulait. □

6.2.4 Proposition. *Soit $f \in \mathcal{L}(E)$. Si f est semi-simple, alors il est aussi semi-simple-bis.*

Preuve. Supposons qu'il existe une décomposition $E = \bigoplus_{i=1}^{r} E_i$ telle que pour tout $i \in [\![1, r]\!]$, E_i est stable par f et l'endomorphisme induit f_{E_i} est simple, et montrons qu'alors f est semi-simple. Soit F un sous-espace stable par f. Il s'agit de trouver un supplémentaire à F lui aussi stable par f. Considérons

$$\mathcal{I} = \{I \subseteq [\![1, r]\!] \text{ t.q. } F \cap (\textstyle\bigoplus_{i\in I} E_i) = \{0\}\}.$$

L'ensemble $\mathcal{I}$ est fini, donc il possède des éléments maximaux pour l'inclusion. Soit I_0 un élément maximal de $\mathcal{I}$. Le sous-espace $E_{I_0} = \bigoplus_{i\in I_0} E_i$ est stable par f, et nous allons montrer que c'est un supplémentaire de F dans E, ce qui conclura la preuve.

Les sous-espaces F et E_{I_0} sont en somme directe puisque $I_0 \in \mathcal{I}$, donc il suffit de prouver que $F \oplus E_{I_0} = E$. Puisque $E = \bigoplus_{i=1}^{r} E_i$, il suffit de prouver que E_{i_1} est contenu dans $F \oplus E_{I_0}$ pour tout $i_1 \in [\![1, r]\!]$. Si $i_1 \in I_0$, c'est clair. Sinon, $I_0 \cup \{i_1\}$ contient strictement I_0, donc par maximalité de I_0, l'ensemble $I_0 \cup \{i_1\}$ n'appartient pas à $\mathcal{I}$, et donc $F \cap (E_{I_0} \oplus E_{i_1}) \neq \{0\}$. Ainsi il existe $y \in F$ non nul, $x_0 \in E_{I_0}$ et $x_1 \in E_{i_1}$ tels que $y = x_0 + x_1$. Puisque $F \cap E_{I_0} = \{0\}$, x_1 est non nul, et puisque $x_1 = y - x_0$, $x_1 \in E_{i_1} \cap (F \oplus E_{I_0})$. Ainsi, $E_{i_1} \cap (F \oplus E_{I_0}) \neq \{0\}$. Mais $E_{i_1} \cap (F \oplus E_{I_0})$ est un sous-espace de E_{i_1} stable par f, donc c'est un sous-espace stable par $f_{E_{i_1}}$. Puisque $f_{E_{i_1}}$ est simple, on en déduit que $E_{i_1} \cap (F \oplus E_{I_0}) = E_{i_1}$, et donc E_{i_1} est tout entier contenu dans $F \oplus E_{I_0}$ comme on voulait démontrer. □

Pour conclure cette section, nous allons donner une preuve différente de l'implication « semi-simple ⇒ semi-simple-bis » qui sera l'occasion de voir quelques lemmes utiles et instructifs.

6.2.5 Lemme. *On considère un endomorphisme $f \in \mathcal{L}(E)$.*
6.2.5.1. *Soit $E = F \oplus F'$ une décomposition telle que F et F' sont tous les deux stables par f. Alors les deux projecteurs $p, p' \in \mathcal{L}(E)$ sur F et F' respectivement relativement à cette décomposition commutent à f.*
6.2.5.2. *Soit $p \in \mathcal{L}(E)$ un projecteur commutant à f. Pour tout sous-espace F stable par f, le sous-espace $p(F)$ est stable par f.*

Preuve. Commençons par prouver 6.2.5.1. Par symétrie il suffit de démontrer le résultat pour p. Considérons un vecteur $x \in E$, et décomposons le en $x = x_F + x'_F$ avec $x_F \in F$, $x'_F \in F'$. On a

$$f(x) = f(x_F) + f(x'_F)$$

par linéarité, et $f(x_F) \in F$, $f(x'_F) \in F'$ par stabilité de F et F' par f. Donc

$$p(f(x)) = f(x_F) = f(p(x))$$

comme il fallait démontrer.

Venons en maintenant à 6.2.5.2. Soit F stable par f. Soit $x \in p(F)$, et montrons que $f(x) \in p(F)$. Il existe $y \in F$ tel que $x = p(y)$, et alors

$$f(x) = f(p(y)) = p(f(y))$$

appartient à $p(F)$ puisque $f(y) \in F$ par stabilité de F. □

6.2.6 Lemme. *Soit E un espace vectoriel muni d'une décomposition $E = F \oplus F'$, et considérons $p \in \mathcal{L}(E)$ le projecteur sur F dans la direction de F'. Pour tout sous-espace $L \subseteq E$, on a*

$$p(L) = (L + F') \cap F.$$

Preuve. Soit $x \in p(L)$. Il existe $y \in L$ tel que $x = p(y)$, et donc $y = x + x'$ pour un certain $x' \in F'$. Alors $x = y - x' \in L + F'$. D'autre part $x \in F$ puisque p est un projecteur sur F. Ceci prouve l'inclusion $p(L) \subseteq (L + F') \cap F$.

Pour montrer l'inclusion inverse, considérons $x \in (L+F') \cap F$. Il existe $y \in L$ et $x' \in F'$ tels que $x = y + x'$. Alors

$$x = p(x) = p(y + x') = p(y),$$

donc $x \in p(L)$ comme il fallait démontrer. □

Preuve alternative de la proposition 6.2.4. Soit $f \in L(E)$. Montrons par récurrence sur $r \in \mathbf{N}^*$ que s'il existe une décomposition $E = \bigoplus_{i=1}^{r} E_i$ telle que pour tout $i \in [\![1, r]\!]$, E_i est stable par f et l'endomorphisme induit f_{E_i} est simple, alors f est semi-simple-bis. Si $r = 1$ le résultat est vrai, puisque alors f est simple et donc semi-simple-bis. Supposons donc $r \geqslant 2$, et le résultat démontré pour un nombre $r' < r$ de facteurs. Posons $E_0 = \bigoplus_{i<r} E_i$; l'hypothèse de récurrence assure que l'endomorphisme $f_{E_0} \in \mathcal{L}(E_0)$ induit par f est semi-simple-bis.

Soit F un sous-espace de E stable par f. L'intersection $F_0 = F \cap E_0$ est stable par f_{E_0}, donc il existe F_0' supplémentaire de F_0 dans E_0 stable par f_{E_0}, et donc par f.

Montrons pour commencer que

$$(\star) \qquad F_0' \oplus F = F + E_0.$$

Puisque $F_0' \subseteq E_0$, on a $F_0' \cap F \subseteq F_0' \cap F_0 = \{0\}$, donc la somme est bien directe. L'inclusion $F_0' \oplus F \subseteq F + E_0$ est conséquence des inclusions $F_0' \subseteq E_0$ et $F \subseteq F$. Montrons l'inclusion inverse $F + E_0 \subseteq F_0' \oplus F$. Puisque $F \subseteq F_0' \oplus F$, il suffit de vérifier que $E_0 \subseteq F_0' \oplus F$, qui est bien vraie puisque $E_0 = F_0' \oplus F_0$ et $F_0 = F \cap E_0 \subseteq F$. Ceci conclut la preuve de $(\star)$.

Le sous-espace $F + E_0$ est stable par f, donc $(F + E_0) \cap E_r$ aussi, et puisque f_{E_r} est simple, on a ou bien $(F + E_0) \cap E_r = \{0\}$ ou bien $(F + E_0) \cap E_r = E_r$. Notons que $(F + E_0) \cap E_r$ est la projection de F sur E_r dans la direction de E_0, d'après le lemme 6.2.6.
a) Si $(F + E_0) \cap E_r = E_r$, alors $F + E_0 = E$. En effet, $F + E_0$ contient alors à la fois E_0 et E_r, qui sont supplémentaires dans E. Dans ce cas, F_0' est un supplémentaire de F dans E stable par f, et la preuve est terminée.
b) Si $(F + E_0) \cap E_r = \{0\}$, alors $F + E_0 = E_0$. En effet, F est alors contenu dans E_0 (puisque sa projection sur E_r est nulle). Dans ce cas, $F_0' \oplus F = E_0$, et $F_0' \oplus E_r$ est un supplémentaire de F dans E stable par f, ce qui conclut la preuve. □

6.2.7 Exercice. Soit $f \in \mathcal{L}(E)$.
1) Montrer que si $F \subseteq E$ est un sous-espace stable de dimension 1, alors il existe $\lambda_F \in \mathbf{k}$ tel que

$$\forall x \in F : \ f(x) = \lambda_F . x.$$

En déduire que F est contenu dans un sous-espace propre de f.
2) Montrer que si f laisse toutes les droites de E stables, alors f est une homothétie.
(*Indication :* étant donné deux droites F et F', pour montrer que $\lambda_F = \lambda_{F'}$, considérer une droite $\mathrm{Vect}(x + x')$ avec $x \in F$ et $x' \in F'$.)

6.3 – Application à la classification

6.3.1 Applications. En se ramenant à des endomorphismes atomiques, on espère pouvoir se limiter à une courte liste de briques élémentaires permettant de reconstruire n'importe

quel endomorphisme. On pourra alors analyser un endomorphisme arbitraire, par exemple en le décrivant par une matrice constituée de blocs bien compris. Une telle matrice est ce qu'on appelle une *forme normale*.

Un aspect de ce problème est la recherche d'une classification des endomorphismes "à équivalence près". La relation d'équivalence qui nous intéresse est la *similitude* : deux endomorphismes $f \in \mathcal{L}(E)$ et $f' \in \mathcal{L}(E')$ sont *semblables* s'il existe un isomorphisme $\varphi : E \simeq E'$ tel que $f' = \varphi \circ f \circ \varphi^{-1}$. Les classes d'équivalence pour la relation de similitude s'appellent les classes de similitude (c'est bien trouvé, non ?).

La recherche de formes normales consiste à donner un représentant emblématique de chaque classe de similitude. Ainsi deux endomorphismes seront semblables si et seulement si ils ont la même forme normale. Un exemple fameux est la forme normale de Jordan pour les endomorphismes trigonalisables.

On s'attachera aussi à la recherche d'*invariants de similitude*, c'est-à-dire d'objets associés à tout endomorphisme qui ne dépendent que de sa classe de similitude. Des exemples bien connus sont les polynômes caractéristique et minimal. On souhaite connaître suffisamment d'invariants pour pouvoir distinguer les classes de similitude. Les polynômes caractéristique et minimal sont notoirement insuffisants pour cela, mais nous réaliserons complètement ce souhait d'avoir suffisamment d'invariants dans la section 8.6 (voir aussi le théorème 7.8.10 pour le cas particulier des endomorphismes trigonalisables).

Avant d'aller plus loin, voyons un lemme élémentaire qui explique comment la réduction des endomorphismes peut être appliquée à la classification à similitude près.

6.3.2 Lemme. *Soit $f, g \in \mathcal{L}(E)$. On suppose qu'il existe deux décompositions $E = \bigoplus_{i=1}^r F_i$ et $E = \bigoplus_{i=1}^r G_i$ en sommes de sous-espaces stables par f et g respectivement, telles que pour tout $i = 1, \dots, r$, il existe un isomorphisme $\varphi_i : F_i \simeq G_i$ tel que $g_{G_i} = \varphi_i f_{F_i} \varphi_i^{-1}$. Alors les endomorphismes f et g sont semblables.*

Preuve. Il s'agit de construire à partir des φ_i un isomorphisme φ de E "diagonal par blocs" qui va conjuguer f et g. Précisément, on définit φ selon ses composantes $\varphi_{ij} \in \mathcal{L}(F_i, G_j)$ relativement aux décompositions $E = \bigoplus_{i=1}^r F_i$ et $E = \bigoplus_{i=1}^r G_i$, grâce à la Proposition 5.1.5 : on considère l'unique $\varphi \in \mathcal{L}(E)$ tel que pour tout $i, j \in [\![1, r]\!]$,

$$\varphi_{ij} = \begin{cases} \varphi_i & \text{si } i = j\,; \\ 0 & \text{sinon.} \end{cases}$$

On vérifie alors sans mal que φ est un isomorphisme puisque $\varphi_1, \dots, \varphi_r$ sont des isomorphismes, et $g = \varphi f \varphi^{-1}$. □

6.4 – Sous-espaces stables et matrices triangulaires par blocs

On a vu en 6.1.3 que l'existence d'une décomposition de l'espace ambiant E en somme directe de sous-espaces stables par un endomorphisme $f \in \mathcal{L}(E)$ se traduit matriciellement par l'existence d'une base de E dans laquelle la matrice de f est diagonale par blocs. Nous aurons besoin de l'énoncé un peu plus robuste ci-dessous, qui se démontre de la même façon (nous laissons donc la preuve comme exercice au lecteur).

6.4.1 Proposition. *Soit $f \in \mathcal{L}(E)$ et F un sous-espace de E stable par f. Dans une base $\mathcal{B}$ de E compatible à F, la matrice* $\mathrm{Mat}_{\mathcal{B}}(f)$ *est triangulaire supérieure par blocs*

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \tag{6.4.1.1}$$

avec $A \in \mathcal{M}_p(\mathbf{k})$ et $B \in \mathcal{M}_{n-p}(\mathbf{k})$ carréees (en notant $n = \dim(E)$ et $p = \dim(F)$).

Réciproquement, si dans une base $\mathcal{B} = (e_1, \ldots, e_n)$ la matrice de f est triangulaire supérieure par blocs comme ci-dessus, alors le sous-espace $\mathrm{Vect}(e_1, \ldots, e_p)$ *est stable par f.*

Nous allons voir que les classes de conjugaison des blocs diagonaux A et B comme en (6.4.1.1) sont canoniquement attachées à la donnée de f et du sous-espace stable F (Corollaire 6.4.7). En revanche le bloc C dépend fortement du choix de la base $\mathcal{B}$, et n'encode en général aucune information intrinsèque à la paire (f, F), comme le démontre l'exemple 6.4.2 ci-dessous.

6.4.2 Exemple. Soit f l'endomorphisme de $\mathbf{k}^2$ donné dans la base canonique (e_1, e_2) par la matrice

$$\begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

On considère le sous-espace stable $F = \mathrm{Vect}(e_1)$; la base canonique est compatible à F. La base $(e_1, e_1 + e_2)$ est elle aussi compatible à f, et dans cette base la matrice de f est

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

L'information géométrique contenue dans les blocs A et B s'exprime en fonction des objets suivants.

6.4.3 Endomorphismes induits. Soit $f \in \mathcal{L}(E)$ et F un sous-espace de E stable par f. La condition

$$\forall x \in F : f_F(x) = f(x).$$

définit un endomorphisme $f_F \in \mathcal{L}(F)$, dit *endomorphisme induit* par f sur F (déjà introduit en 6.1.5 ci-dessus).

D'autre part, la condition

$$\forall \bar{x} \in E/F : \bar{f}_F(\bar{x}) = \overline{f(x)}$$

définit un endomorphisme $\bar{f}_F \in \mathcal{L}(E/F)$, dit *endomorphisme sur le quotient induit* par f.

Il convient de vérifier que $\bar{f}_F$ est bien défini par la condition ci-dessus, ce qui revient à vérifier que $\overline{f(x)}$ ne dépend pas du choix du représentant de la classe $\overline{x}$. Soit $x' \in E$ un autre représentant de $\bar{x}$. On a $x' - x \in F$, et donc puisque F est stable et f est linéaire, $f(x') - f(x) \in F$. Ainsi $\overline{f(x')} = \overline{f(x)}$ comme il fallait démontrer.

6.4.3.1 Mise en garde. Nous recommandons d'être attentif au fait que la donnée de f_F et $\bar{f}_F$ est insuffisante pour reconstruire f, voir l'exemple 6.4.2 ci-dessus.

6.4.4 Proposition. *Soit $f \in \mathcal{L}(E)$, F stable par f, et $\mathcal{B}$ une base de E compatible à F. Alors les blocs A et B de la matrice* $\mathrm{Mat}_{\mathcal{B}}(f)$ *écrite comme en* (6.4.1.1) *représentent f_F et $\bar{f}_F$ respectivement, dans les bases de F et de E/F induites par $\mathcal{B}$.*

Avant d'attaquer la preuve, précisons ce que nous appelons les bases de F et de E/F induites par $\mathcal{B}$. Puisque $\mathcal{B}$ est compatible à F, elle est la concaténation d'une base $\mathcal{B}_F$ de F et d'une famille libre $\mathcal{B}'$; d'après le lemme 5.4.2, les classes modulo F des vecteurs de $\mathcal{B}'$ forment une base $\bar{\mathcal{B}}_F$ de E/F. Les bases induites dont il est question dans l'énoncé ci-dessus sont $\mathcal{B}_F$ et $\bar{\mathcal{B}}_F$.

Preuve. Le fait que $A = \mathrm{Mat}_{\mathcal{B}_F}(f_F)$ est contenu dans la proposition 5.2.2. Pour montrer que $B = \mathrm{Mat}_{\bar{\mathcal{B}}_F}(\bar{f}_F)$, notons $\mathcal{B}_F = (e_1, \dots, e_p)$ et $\mathcal{B}' = (e_{p+1}, \dots, e_n)$. Soit $j \in [\![1, n-p]\!]$. Notons en outre $C = (c_{ij})_{\substack{1 \leqslant i \leqslant p \\ 1 \leqslant j \leqslant n-p}}$ et $B = (b_{ij})_{1 \leqslant i,j \leqslant n-p}$ les blocs de (6.4.1.1). On a

$$(*) \qquad \begin{aligned} \bar{f}_F(\bar{e}_{p+j}) &= \overline{f(e_{p+j})} \\ &= \overline{\underbrace{\sum_{i=1}^{p} c_{ij}.e_i}_{=\bar{0}}} + \overline{\sum_{i=p+1}^{n} b_{i-p,j}.e_i} \\ &= \sum_{i=1}^{n-p} b_{ij}.\bar{e}_{p+i}, \end{aligned}$$

ce qui prouve bien que $B = \mathrm{Mat}_{\bar{\mathcal{B}}_F}(\bar{f}_F)$. (* La première somme est nulle car c'est une combinaison linéaire de $\overline{e_1}, \dots, \overline{e_p}$ qui sont tous nuls, puisque $e_1, \dots, e_p \in F$). □

6.4.5 Remarque. Il est instructif de suivre les choix faits lors de l'écriture de la matrice (6.4.1.1). Au commencement on a F stable par f. Le choix d'une base de F détermine le bloc A, qui est la matrice de f_F dans cette base. Ensuite le choix d'une base $\bar{\mathcal{B}} = (\varepsilon_1, \dots, \varepsilon_{n-p})$ de E/F détermine le bloc B, qui est la matrice de $\bar{f}_F$ dans cette base. Jusqu'ici on n'a pas eu à choisir de supplémentaire à F.

Enfin, il faut choisir des représentants $e'_1, \dots, e'_{n-p} \in E$ de $\varepsilon_1, \dots, \varepsilon_{n-p} \in E/F$ pour déterminer le bloc C. Ce choix détermine un supplémentaire $F' = \mathrm{Vect}(e'_1, \dots, e'_{n-p})$ de F dans E, qui incarne le quotient E/F dans E, voir 5.4.6. On prendra garde au fait que ce supplémentaire F' n'a aucune raison d'être stable par f (d'ailleurs on a déjà vu qu'il est tout-à-fait possible qu'un sous-espace stable ne possède aucun supplémentaire stable, voir l'exemple 6.1.8).

6.4.6 Exercice. Soit $f, g \in \mathcal{L}(E)$, et F un sous-espace stable par f et par g. Montrer que F est stable par $f \circ g$, et

$$(f \circ g)_F = f_F \circ g_F \quad \text{et} \quad (\overline{f \circ g})_F = \bar{f}_F \circ \bar{g}_F.$$

En déduire que dans un produit MN de matrices triangulaires supérieures par blocs (de tailles compatibles), les blocs diagonaux de MN sont les produits des blocs diagonaux de M et N respectivement.

6.4.7 Corollaire. *Soit $f \in \mathcal{L}(E)$ et F stable par f. Considérons les matrices*

$$\begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \quad \textit{et} \quad \begin{pmatrix} A' & C' \\ 0 & B' \end{pmatrix}$$

de f dans deux bases $\mathcal{B}$ et $\mathcal{B}'$ compatibles à F. Alors les matrices A et A' d'une part, et B et B' d'autre part, sont conjuguées.

Preuve. D'après la Proposition 6.4.4, les matrices A et A' (resp. B et B') représentent le même endomorphisme f_F (resp. $\bar{f}_F$) dans les bases de F (resp. E/F) induites par $\mathcal{B}$ et $\mathcal{B}'$ respectivement. Ceci implique le résultat. □

Chapitre 7

Réduction des endomorphismes : Analyse

7.1 – Polynôme caractéristique

7.1.1. On adopte la convention

$$\chi_f = \text{"}\det(X.\mathrm{id}_E - f)\text{"}$$

pour le polynôme caractéristique d'un endomorphisme $f \in \mathcal{L}_{\mathbf{k}}(E)$. L'alternative serait de prendre plutôt "$\det(X.\mathrm{id}_E - f)$" ; notre choix présente l'avantage d'avoir le polynôme caractéristique scindé, mais l'inconvénient de devoir écrire beaucoup de signes 'moins' lors des calculs.

Il faut prendre garde au fait que l'expression "$\det(X.\mathrm{id}_E - f)$" n'est pas vraiment bien définie, dans la mesure où $\mathcal{L}_{\mathbf{k}}(E)$ n'est pas munie d'une multiplication extérieure par $\mathbf{k}[X]$, et donc écrire "$X.\mathrm{id}_E$" est insensé. Une attitude prudente à cet égard consiste à définir χ_f de la manière suivante.

On considère une base $\mathcal{B}$ de E, et on écrit la matrice A de f dans cette base $\mathcal{B}$. *A priori* la matrice A vit dans $\mathcal{M}_n(\mathbf{k})$, avec $n = \dim(E)$. On considère $\mathcal{M}_n(\mathbf{k})$ comme sous-anneau de $\mathcal{M}_n(\mathbf{k}[X])$, ce qui permet de voir A comme élément de $\mathcal{M}_n(\mathbf{k}[X])$. Alors la matrice $X.\mathrm{Id}_n - A$ est un élément de $\mathcal{M}_n(\mathbf{k}[X])$ des plus en règle, et nous pouvons considérer son déterminant sans problème, voir la définition 2.1.3. Enfin, les formules de changement de base impliquent que ce déterminant est indépendant de la base $\mathcal{B}$ choisie pour écrire la matrice de f, et nous pouvons donc poser

$$\chi_f = \det(X.\mathrm{Id}_n - A),$$

sans guillement, et sans ambigüité de définition.

Nous introduirons la notion de valeur propre un peu plus loin, voir 7.2.4. Le résultat ci-dessous sera alors essentiel.

7.1.2 Proposition. *Soit $f \in \mathcal{L}_{\mathbf{k}}(E)$ et $\lambda \in \mathbf{k}$. Les propositions suivantes sont équivalentes :*

(i) le scalaire λ est racine de χ_f ;
(ii) l'endomorphisme $\lambda.\mathrm{id}_E - f \in \mathcal{L}_{\mathbf{k}}(E)$ n'est pas inversible ;
(iii) le sous-espace $\ker(\lambda.\mathrm{id}_E - f \in \mathcal{L}_{\mathbf{k}}(E))$ n'est pas réduit à $\{0\}$.

Preuve. L'équivalence entre (ii) et (iii) ne devrait pas poser de problème à ce stade. L'équivalence entre (i) et (ii) d'autre part est une conséquence directe de l'identité

$$\chi_f(\lambda) = \det(\lambda.\mathrm{id}_E - f), \tag{7.1.2.1}$$

valable pour tout $\lambda \in \mathbf{k}$.

Cette identité pourra sembler évidente au lecteur imprudent, mais elle ne l'est pas (à titre de comparaison, voir 7.3.3). Considérons une base $\mathcal{B}$ de E, et écrivons la matrice A de f dans cette base. Il s'agit de prouver que lorsqu'on évalue le polynôme $\chi_f(X) = \det(X.\mathrm{Id}_n - A) \in \mathbf{k}[X]$ en $\lambda \in \mathbf{k}$, le scalaire obtenu est le déterminant de la matrice $\lambda.\mathrm{Id}_n - A \in \mathcal{M}_n(\mathbf{k})$; autrement dit que calculer le déterminant de $X.\mathrm{Id}_n - A$ puis évaluer X en λ donne la même chose qu'évaluer d'abord X en λ dans la matrice $X.\mathrm{Id}_n - A$ puis calculer le déterminant.

Pour le voir, notons $a_{i,j}(X) \in \mathbf{k}[X]$ les coefficients de la matrice $X.\mathrm{Id}_n - A$. Par définition,

$$\chi_f(X) = \det(X.\mathrm{Id}_n - A) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{\sigma(1),1}(X) \cdots a_{\sigma(n),n}(X) \in \mathbf{k}[X].$$

Ainsi, l'évaluation de ce polynôme en $\lambda \in \mathbf{k}$ donne

$$\chi_f(\lambda) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{\sigma(1),1}(\lambda) \cdots a_{\sigma(n),n}(\lambda) \in \mathbf{k}.$$

D'autre part, puisque la matrice $\lambda.\mathrm{Id}_n - A$ s'obtient en évaluant X en λ dans $X.\mathrm{Id}_n - A$, les $a_{i,j}(\lambda)$ sont les coefficients de $\lambda.\mathrm{Id}_n - A$. On a donc

$$\det(\lambda.\mathrm{Id}_n - A) = \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, a_{\sigma(1),1}(\lambda) \cdots a_{\sigma(n),n}(\lambda),$$

et l'identité (7.1.2.1) est démontrée. □

7.1.3 Proposition. *Soit $f \in \mathcal{L}(E)$.*
(i) Soit F un sous-espace stable par f. Alors $\chi_f = \chi_{f_F} \cdot \chi_{\bar{f}_F}$.
(ii) On suppose qu'il existe une décomposition $E = \bigoplus_{i=1}^r F_i$ où tous les F_i sont stables par f. Alors on a l'égalité

$$\chi_f = \prod_{i=1}^r \chi_{f_{F_i}}.$$

Preuve. L'assertion (i) est une conséquence directe de la proposition 6.4.4 ci-dessus et du calcul des déterminants triangulaires par blocs 2.3.3.

L'assertion (ii) s'obtient de la même façon en observant que la matrice de f dans une base compatible à la décomposition $E = \bigoplus_{i=1}^r F_i$ est diagonale par blocs en application de 6.1.3 et 5.2.2. □

7.1.4 Application de la formule de Laplace : calcul des coefficients de χ_M. Par multilinéarité du déterminant, on a

$$\begin{aligned}
\chi_A(X) &= \det(X \cdot e_1 - C_1, \ldots, X \cdot e_n - C_n) \\
&= \sum_{k=0}^{n} \left(\sum_{i_1 < \cdots < i_k} \det(-C_1, \ldots, X.e_{i_1}, \ldots, X.e_{i_k}, \ldots, -C_n) \right) \\
&= \sum_{k=0}^{n} (-1)^{n-k} X^k \left(\sum_{i_1 < \cdots < i_k} \det(C_1, \ldots, e_{i_1}, \ldots, e_{i_k}, \ldots, C_n) \right)
\end{aligned}$$

et les déterminants dans la somme entre parenthèses à droite se calculent en un coup avec la formule de Laplace, en développant par rapport aux colonnes d'indices $i_1, \dots, i_k$.

On peut aussi calculer "directement" ce déterminant, et nous allons le faire. Les ressorts de ce calcul sont déjà en œuvre dans la preuve de la formule de Laplace, ainsi commencer par la preuve ci-dessous avant d'aborder celle de la formule de Laplace peut être une bonne idée.

On calcule

$$\det(C_1, \dots, e_{i_1}, \dots, e_{i_k}, \dots, C_n).$$

On pose $I = \{i_1 < \cdots i_k\}$, et $\bar{I} = \{\bar{\imath}_1 < \cdots < \bar{\imath}_{n-k}\}$ son complémentaire dans $[\![1, n]\!]$. On commence par effectuer des échanges de colonnes pour se ramener à

$$\det(e_{i_1}, \dots, e_{i_k}, C_{\bar{\imath}_1}, \dots, C_{\bar{\imath}_{n-k}}) \ ;$$

cette opération nécessite un certain nombre d'échanges $\varepsilon(I)$, qu'on peut calculer facilement, voir 7.1.4.1, mais qu'il est inutile de connaître explicitement pour la preuve. En effet, en opérant exactement les mêmes échanges sur les lignes on place les lignes $i_1, \dots, i_k$ en haut, obtenant ainsi le déterminant

$$\begin{vmatrix} 1 & & & a_{i_1\bar{\imath}_1} & \dots & a_{i_1\bar{\imath}_{n-k}} \\ & \ddots & & \vdots & & \vdots \\ & & 1 & a_{i_k\bar{\imath}_1} & \dots & a_{i_k\bar{\imath}_{n-k}} \\ 0 & \dots & 0 & a_{\bar{\imath}_1\bar{\imath}_1} & \dots & a_{\bar{\imath}_1\bar{\imath}_{n-k}} \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \dots & 0 & a_{\bar{\imath}_{n-k}\bar{\imath}_1} & \dots & a_{\bar{\imath}_{n-k}\bar{\imath}_{n-k}} \end{vmatrix} = \det(A^{II}).$$

7.1.4.1 Complément. Il est instructif (en particulier si on veut lire la preuve de la formule de Laplace) de calculer le nombre d'échanges nécessaires pour passer de

$$\det(C_1, \dots, e_{i_1}, \dots, e_{i_k}, \dots, C_n).$$

à

$$\det(e_{i_1}, \dots, e_{i_k}, C_{\bar{\imath}_1}, \dots, C_{\bar{\imath}_{n-k}}).$$

Il faut $i_1 - 1$ échanges pour mettre e_{i_1} en premier, ensuite $i_2 - 2$ échanges pour placer e_{i_2} juste après e_{i_1}, et ainsi de suite. Au total, le nombre d'échanges est donc $\sum_{k=1}^{p}(i_k - k)$, et nos manipulations changent le déterminant par -1 puissance cet entier.

7.1.4.2 Exercice. Étudier $\det(A + X \cdot M)$.

7.1.4.3 Corollaire. $M \mapsto \chi_M$ est continue (on verra que ce n'est pas le cas de $M \mapsto \mu_M$). Si $\mathbf{k} = \mathbf{R}$ ou $\mathbf{C}$, sinon topologie de Zariski.

7.1.5 Développement. *Soit $A, B \in \mathcal{M}_n(\mathbf{k})$. On a $\chi_{AB} = \chi_{BA}$.*

Je présente ce résultat essentiellement motivé par la méthode de preuve. Toutefois, cet énoncé a son intérêt. C'est une généralisation de la formule bien connue $\mathrm{Tr}(AB) = \mathrm{Tr}(BA)$. En fait, pour qui connaît un peu d'algèbre extérieure ce n'est pas vraiment une généralisation, puisqu'au signe près les coefficients du polynôme caractéristique χ_M sont les traces des matrices $M^{\wedge i}$, $i = 0, \dots, n$.

7.1.5.1 Cas A inversible. Dans ce cas le résultat s'obtient par un calcul qui ne dévoile pas vulgairement la raison profonde de notre énoncé. On calcule des déterminants de matrices à coefficients dans l'anneau commutatif $\mathbf{k}[X]$.

$$\begin{aligned}\chi_{AB} = \det(X.\mathrm{id} - AB) &= \det\big[A(X.A^{-1} - B)\big] \\ &= \det\big[(X.A^{-1} - B)A\big] = \det(X.\mathrm{id} - BA) = \chi_{BA}\end{aligned}$$

et c'est gagné.

Voyons maintenant comment déduire le résultat sans condition sur A par un argument de passage à la limite qui peut se formuler de plusieurs façons.

7.1.5.2 Version 1. Si $\mathbf{k}$ est infini, il existe une infinité de $\lambda \in \mathbf{k}$ tels que $A - \lambda.\mathrm{Id}$ est inversible. Pour tous ces λ, on a

$$\chi_{(A-\lambda\mathrm{Id})B} = \chi_{B(A-\lambda\mathrm{Id})}.$$

Les coefficients du polynôme en X $\chi_{(A-\lambda\mathrm{Id})B} - \chi_{B(A-\lambda\mathrm{Id})}$ sont des polynômes en λ, nuls pour une infinité de valeurs de λ, et donc uniformément nuls. Autrement dit, l'annulation de $\chi_{(A-\lambda\mathrm{Id})B} - \chi_{B(A-\lambda\mathrm{Id})} \in \mathbf{k}[X]$ pour une infinité de λ implique son annulation tout court. Pour $\lambda = 0$, ceci nous donne le résultat voulu.

Si $\mathbf{k}$ n'est pas infini, on trouve une extension $\mathbf{k}'$ de $\mathbf{k}$ qui l'est, par exemple $\mathbf{k}(T)$ (fraction rationnelles en T). Par le même argument on obtient l'identité $\chi_{AB} = \chi_{BA}$ qui se fiche pas mal de savoir si on habite dans $\mathcal{M}_n(\mathbf{k})$ ou $\mathcal{M}_n(\mathbf{k}')$.

7.1.5.3 Version 2. Si $\mathbf{k} = \mathbf{R}$ ou $\mathbf{C}$, ou encore mieux si on connaît la topologie de Zariski, alors on peut arguer que

$$(A, B) \in \mathcal{M}_n(\mathbf{k}) \times \mathcal{M}_n(\mathbf{k}) \mapsto \chi_{AB} - \chi_{BA} \in \mathbf{k}[X]_n$$

est continue et nulle sur l'ouvert dense des (A, B) tels que A inversible, et donc uniformément nulle.

7.1.5.4 Version 3. On considère les deux matrices $A = (a_{ij})$ et $B = (b_{ij})$ à coefficients indéterminés. Il s'agit de démontrer l'identité $\chi_{AB} = \chi_{BA}$ entre polynômes à coefficients dans l'anneau de polynômes $\mathbf{Z}[a_{ij}, b_{ij}]$ (polynômes en $2n^2$ indéterminées, à coefficients entiers) : $\chi_{AB}, \chi_{BA} \in \mathbf{Z}[a_{ij}, b_{ij}][X]$.

On considère le corps des fractions de cet anneau intègre, qui est le corps de fractions rationnelles $\mathrm{K} := \mathbf{Q}(a_{ij}, b_{ij})$. Puisque le déterminant est un polynôme non-nul, A est inversible dans $\mathcal{M}_n(\mathrm{K})$. On en déduit par notre calcul basique l'identité $\chi_{AB} = \chi_{BA}$ dans $\mathrm{K}[X]$. Bien sûr les deux polynômes sont en fait dans le sous-anneau $\mathbf{Z}[a_{ij}, b_{ij}][X]$ de $\mathrm{K}[X]$, et l'identité $\chi_{AB} = \chi_{BA}$ est une égalité entre polynômes à coefficients dans $\mathbf{Z}[a_{ij}, b_{ij}]$: notre résultat est démontré !

Remarque. Le couple (A, B) comme ci-dessus est le point générique au sens de la géométrie algébrique de l'espace affine $\mathcal{M}_n(\mathbf{Z}) \times \mathcal{M}_n(\mathbf{Z})$.

7.2 – Polynômes d'endomorphismes

7.2.1 Définition. On considère le morphisme de $\mathbf{k}$-algèbres $\theta_f : \mathbf{k}[X] \to \mathcal{L}_{\mathbf{k}}(E)$, dit *morphisme d'évaluation* en f, défini par la condition $\theta_f(X) = f$.

Puisque X engendre $\mathbf{k}[X]$ comme $\mathbf{k}$-algèbre, cette condition détermine de manière unique le morphisme θ_f. Pour $P = a_nX^n + \cdots + a_1X + a_0$, avec $a_i \in \mathbf{k}$ pour tout $i = 0, \ldots, n$, on a

$$\begin{aligned}\theta_f(P) &= a_n\theta_f(X)^n + \cdots + a_1\theta_f(X) + a_0\theta_f(1)\\ &= a_nf^n + \cdots + a_1f + a_0\mathrm{id}_E,\end{aligned}$$

où pour tout $i \in \mathbf{N}$, $f^i = f \circ \cdots \circ f$ (i facteurs).

On notera $P(f) = \theta_f(P) \in \mathcal{L}(E)$ pour tout $P \in \mathbf{k}[X]$. On désignera par $\mathbf{k}[f]$ l'image du morphisme d'évaluation θ_f ; un *polynôme en f* est un éléments de $\mathbf{k}[f]$. On notera que $\mathbf{k}[f]$ est une sous-algèbre de $\mathcal{L}(E)$. Elle est de dimension finie puisque $\mathcal{L}(E)$ est de dimension finie égale à $\dim(E)^2$, et commutative puisque $\mathbf{k}[X]$ est une algèbre commutative. Ainsi, deux polynômes en f commutent toujours.

L'utilisation de polynômes en f sera fondamentale pour notre approche de la réduction des endomorphismes. Pour commencer, remarquons que la plupart des endomorphismes habituellement associés à f sont en fait des polynômes en f.

7.2.2 Exemples. L'inverse de f (lorsqu'il existe), l'exponentielle de f, sont des polynômes en f. Pour l'inverse voir le corollaire 7.2.8 et la remarque 7.2.8.1 ci-dessous. Pour l'exponentielle, cela vient du fait que $\exp(f)$ est une limite de polynômes en f, et que $\mathbf{k}[f]$ est fermé dans $\mathcal{L}(E)$.

On verra plus tard des exemples de projecteurs intrinsèquement liés à f (à savoir les projecteurs spectraux, voir 7.8.2) qui sont des polynômes en f.

Un premier intérêt de considérer des polynômes d'endomorphismes pour la réduction est qu'ils permettent de produire une profusion de sous-espaces stables.

7.2.3 Lemme. *Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes qui commutent, et $P \in \mathbf{k}[X]$. Alors le sous-espace vectoriel $\ker(P(f))$ est stable par g.*

En particulier, en prenant $g = f$, on obtient que $\ker(P(f))$ est stable par f. La preuve de ce lemme est élémentaire et laissée au lecteur.

7.2.3.1 Mise en garde. Même si f et g commutent, il est grossièrement faux que tout sous-espace stable par f est stable par g. Par exemple, $g = \mathrm{id}$ commute avec tout f, tout sous-espace de E est stable par id, mais bien sûr il existe en général des sous-espaces de E qui ne sont pas stables par f.

7.2.4 Valeurs propres, espaces propres. Soit $\lambda \in \mathbf{k}$. On note $E_\lambda(f) = \ker(f - \lambda.\mathrm{id}_E)$ (ou simplement E_λ s'il n'y a pas de risque de confusion), qu'on appelle *espace propre* de f pour la valeur λ. On dit que λ est *valeur propre* de f si $E_\lambda(f) \supsetneq \{0\}$. L'ensemble des valeurs propres de f est appelé le *spectre* de f, noté $\mathrm{Sp}(f)$. Un scalaire $\lambda \in \mathbf{k}$ est valeur propre de f si et seulement si λ est racine du polynôme caractéristique $\chi_f = \det(X\mathrm{id}_E - f)$, voir proposition 7.1.2.

On a $E_\lambda(f) = \ker\big((X - \lambda)(f)\big)$, donc par le lemme 7.2.3 ci-dessus les espaces propres de f sont stables par tout endomorphisme commutant avec f.

7.2.5 Polynômes annulateurs, polynôme minimal. Un *polynôme annulateur* pour f est un polynôme $P \in \mathbf{k}[X]$ tel que $P(f) = 0$. Le *polynôme minimal* de f est l'unique générateur unitaire du noyau du morphisme d'évaluation θ_f.[1] On le note μ_f, ou μ si on ne craint pas de confusion.

1. on rappelle à cet égard que l'anneau $\mathbf{k}[X]$ est principal, donc il existe $P \in \mathbf{k}[X]$ tel que $\ker(\theta_f) = (P)$. Puisque $\mathbf{k}[X]$ est de dimension infinie et $\mathcal{L}(E)$ est de dimension finie, θ_f ne peut pas être injectif, et donc

Le polynôme minimal de f est caractérisé par les deux conditions suivantes (en plus du fait d'être unitaire, qui garantit son unicité) :
(i) $\mu(f) = 0 \in \mathcal{L}(E)$;
(ii) pour tout $P \in \mathbf{k}[X]$, si $P(f) = 0$ alors μ divise P.
Ainsi μ est le polynôme unitaire annulant f dont le degré est le plus petit possible.

Par définition, on a $\ker(\theta_f) = (\mu_f)$, et donc un isomorphisme canonique $\mathbf{k}[f] \cong \mathbf{k}[X]/(\mu_f)$. En particulier, la dimension de la $\mathbf{k}$-algèbre $\mathbf{k}[f]$ est $\deg(\mu_f)$.

7.2.6 Proposition. *Le polynôme minimal est un invariant de similitude. Il est stable par extension des scalaires.*

7.2.6.1 Exercice. Montrer que

$$\deg(\mu_f) = \min\{p : (\mathrm{id}, f, \dots, f^p) \text{ est liée}\}$$

(où l'on considère $(\mathrm{id}, f, \dots, f^p)$ comme une famille de vecteurs de $\mathcal{L}(E)$).

7.2.6.2 Application. Invariance de μ par extension de corps, *via* l'invariance du rang. (La bonne preuve est que μ est un invariant de similitude, et que ceux-ci se calculent par un pivot de Gauss).

Soit $\mathbf{k} \to \mathbf{k}'$ une extension de corps. Certainement $\mu_{u^{\mathbf{k}'}}(u^{\mathbf{k}}) = 0$, donc $\mu_{u^{\mathbf{k}}} | \mu_{u^{\mathbf{k}'}}$. Ces deux polynômes ont le même degré par 7.2.6.1 et l'invariance du rang par extension de corps (voir 2.5.8 du Prologue), donc diffèrent d'une constante multiplicative. □

7.2.7 Proposition. *Soit $f \in \mathcal{L}(E)$, $P \in \mathbf{k}[X]$. Les deux propositions suivantes sont équivalentes :*
(i) $P(f)$ est inversible ;
(ii) $P(f)$ est inversible et $P(f)^{-1} \in \mathbf{k}[f]$;
(iii) P et μ_f sont premiers entre eux.

Preuve. L'implication (ii)⇒(i) est triviale. Montrons que (i) ⇒ (iii) : supposons que $P(f)$ est inversible. Soit R un facteur irréductible de P. Alors par minimalité de μ, R ne divise pas μ : en effet s'il existait Q tel que $\mu = QR$, alors puisque $R(f)$ est inversible [2] on aurait nécessairement $Q(f) = 0$. Ceci implique que P et μ sont premiers entre eux.

Montrons que (iii) ⇒ (ii) : supposons P et μ premiers entre eux. Alors il existe $U, V \in \mathbf{k}[X]$ tels que $PU + \mu V = 1$. En évaluant en f, on obtient

$$P(f)Q(f) + \underbrace{\mu(f)}_{=0} V(f) = \mathrm{id} \quad \Longleftrightarrow \quad P(f)Q(f) = \mathrm{id},$$

ainsi $P(f)$ est inversible d'inverse $Q(f)$. □

7.2.8 Corollaire. *Soit $f \in \mathcal{L}(E)$. L'endomorphisme f est inversible si et seulement si $\mu_f(0) \neq 0$. Dans ce cas $f^{-1} \in \mathbf{k}[f]$.*

Preuve. On applique la proposition avec $P = X$; les polynômes X et μ sont premiers entre eux si et seulement si X ne divise pas μ, ce qui équivaut à la condition $\mu(0) \neq 0$. □

$\ker(\theta_f) \supsetneq (0)$. Ainsi $P \neq 0$, et puisque les générateurs de l'idéal $\ker(\theta_f)$ sont les aP avec $a \in \mathbf{k}^*$, il existe un unique générateur de $\ker(\theta_f)$ qui est unitaire.

2. il existe S tel que $P = RS$, $P(f) = S(f) \circ R(f)$ est inversible, donc nécessairement $R(f)$ est inversible.

7.2.8.1 Remarque. Dans la situation du corollaire, il est facile en pratique de trouver un polynôme donnant l'inverse. Si $\mu = a_p X^p + \cdots + a_1 X + a_0$ avec $a_0 \neq 0$, alors

$$a_p f^p + \cdots + a_1 f + a_0.\mathrm{id} = f(a_p f^{p-1} + \cdots + a_1.\mathrm{id}) + a_0.\mathrm{id} = 0$$

donc $f^{-1} = -a_0^{-1}(a_p f^{p-1} + \cdots + a_1.\mathrm{id})$.

7.2.9 Proposition. *Soit $f \in \mathcal{L}(E)$ et P un polynôme annulateur de f (par exemple $P = \mu_f$ ou χ_f). Pour tout $A \in \mathbf{k}[X]$, on a*

$$A(f) = R(f)$$

où R est le reste de la division euclidienne de A par P.

La preuve est élémentaire et laissée au lecteur.

7.2.9.1 Application : calcul de puissance. La proposition ci-dessus fournit un moyen efficace de calculer f^k pour de grands entiers k. Une autre méthode consiste à utiliser la décomposition de Dunford de f, voir **??**.

7.2.10 Lemme. *Les polynômes χ et μ ont les mêmes racines.*

On verra plus loin (Prop. 8.2.1) qu'en fait, χ et μ ont toujours les mêmes facteurs irréductibles, même s'ils ne sont pas scindés.

Preuve. Soit $\lambda \in \mathbf{k}$. On a la suite d'équivalences :

$$\begin{aligned} \chi(\lambda) = 0 &\iff f - \lambda.\mathrm{id} \text{ non inversible} \\ &\iff X - \lambda \text{ et } \mu \text{ non premiers entre eux} \qquad \text{(par la proposition 7.2.7)} \\ &\iff \mu(\lambda) = 0. \end{aligned}$$

□

7.2.10.1 Remarque. On peut aussi démontrer que toute valeur propre de f est racine de μ_f par le calcul suivant. Soit $P \in \mathbf{k}[X]$, et $e \in E_\lambda(f)$. On a

$$P(f)(e) = P(\lambda).e.$$

Si λ est valeur propre, on peut supposer $e \neq 0$; le calcul précédent indique alors que λ est racine de tout polynôme annulateur de f.

Avant d'aborder le prochain résultat, il est bon de faire l'observation suivante.

7.2.11 Lemme. *Soit $f \in \mathcal{L}(E)$ et F un sous-espace stable par f. Pour tout polynôme $P \in \mathbf{k}[X]$, le sous-espace F est stable par $P(f)$, et on a*

$$P(f)_F = P(f_F) \in \mathcal{L}(F) \quad \text{et} \quad \overline{P(f)}_F = P(\bar{f}_F) \in \mathcal{L}(E/F).$$

Ce lemme est essentiellement tautologique et nous l'utiliserons sans même y penser. Il faut toutefois s'assurer qu'on sait le démontrer. Sans surprise la preuve n'est pas très instructive, mais elle est particulièrement fastidieuse.

Preuve. Pour commencer, on démontre par récurrence sur $k \in \mathbf{N}$ que F est stable par f^k, $(f^k)_F = (f_F)^k$, et $\overline{(f^k)}_F = (\bar{f}_F)^k$. Pour $k = 0$, F est bien stable par $f^0 = \mathrm{id}_E$, et on a bien $(\mathrm{id}_E)_F = \mathrm{id}_F$, et $\overline{(\mathrm{id}_E)}_F = \mathrm{id}_{E/F}$. Soit $k \in \mathbf{N}^*$, et supposons le résultat démontré pour tout $k' < k$. Montrons que F est stable par f^k : soit $x \in F$; on a $f^{k-1}(x) \in F$ par hypothèse de récurrence, donc $f^k(x) = f(f^{k-1}(x)) \in F$ puisque F est stable par f. Montrons que $(f^k)_F = (f_F)^k$: soit $x \in F$; on a $(f^k)_F(x) = f^k(x)$ par définition, donc $(f^k)_F(x) = f(f^{k-1}(x)) = f((f_F)^{k-1}(x))$ par hypothèse de récurrence, et ainsi par définition de f_F et puisque $(f_F)^{k-1}(x) \in F$, $\overline{(f^k)}_F(x) = f_F((f_F)^{k-1}(x)) = (f_F)^k(x)$ comme il fallait démontrer. La preuve du fait que $\overline{(f^k)}_F = (\bar{f}_F)^k$ est similaire, et nous ne l'infligerons à personne.

À présent considérons $P \in \mathbf{k}[X]$ que nous écrirons $P = a_k X^k + \cdots + a_0$. L'endomorphisme $P(f)$ est une combinaison linéaire de $f^k, \ldots, f, \mathrm{id}_E$, qui laissent tous F stable d'après ce qui précède, donc F est stable par $P(f)$. Montrons que $P(f)_F = P(f_F)$: soit $x \in F$; on a

$$\begin{aligned} P(f)_F(x) = P(f)(x) &= a_k f^k(x) + \cdots + a_1 f(x) + a_0 x \\ &= a_k f_F^k(x) + \cdots + a_1 f_F(x) + a_0 x = P(f_F)(x). \end{aligned}$$

La preuve de l'identité $\overline{P(f)}_F = P(\bar{f}_F)$ est similaire. □

Ce résultat semble moins abscons écrit matriciellement ; en vertu de la proposition 6.4.1, ceci prend la formule suivante : pour une matrice M triangulaire supérieure par blocs comme ci-dessous, pour tout $P \in \mathbf{k}[X]$, la matrice $P(M)$ est comme indiqué ci-dessous :

$$M = \begin{pmatrix} A & C \\ 0 & B \end{pmatrix} \; ; \qquad P(M) = \begin{pmatrix} P(A) & * \\ 0 & P(B) \end{pmatrix}. \tag{7.2.11.1}$$

Cependant pour garder les choses à l'endroit, il faut bien se souvenir que les règles de calcul pour les matrices triangulaires par blocs proviennent des règles de composition pour les endomorphismes laissant un même sous-espace stable, voir exercice 6.4.6, et donc *in fine* c'est bien la formule (7.2.11.1) qui provient du lemme 7.2.11 et pas l'inverse.

7.2.12 Proposition. *Soit $f \in \mathcal{L}(E)$ et F un sous-espace stable par f. Alors μ_{f_F} et $\mu_{\bar{f}_F}$ divisent tous les deux μ_f ; de manière équivalente,* $\mathrm{ppcm}(\mu_{f_F}, \mu_{\bar{f}_F})$ *divise* μ_f.

Si de plus F possède un supplémentaire stable, alors on a égalité $\mu_f = \mathrm{ppcm}(\mu_{f_F}, \mu_{\bar{f}_F})$.

Preuve. Il résulte des définitions des endomorphismes induits $f_F \in \mathcal{L}(F)$ et $\bar{f}_F \in \mathcal{L}(E/F)$ que tout polynôme annulateur pour f est *a fortiori* annulateur pour f_F et $\bar{f}_F$. Autrement dit on a les deux inclusions d'idéaux

$$(\mu_f) \subseteq (\mu_{f_F}) \quad \text{et} \quad (\mu_f) \subseteq (\mu_{\bar{f}_F}),$$

respectivement équivalentes aux deux relations de divisibilité $\mu_{f_F} | \mu_f$ et $\mu_{\bar{a} f_F} | \mu_f$. □

7.3 – À propos du théorème de Cayley–Hamilton

Cette section est consacrée au théorème suivant, bien connu et communément dit de Cayley–Hamilton.

Il semblerait cependant que Cayley aussi bien qu'Hamilton n'en aient démontré que des cas particuliers, la première preuve complète étant dûe à Frobenius. Pour l'instant, à

ce sujet je me contente de citer Wikipedia [3] : « A special case of the theorem was first proved by Hamilton in 1853 in terms of inverses of linear functions of quaternions. This corresponds to the special case of certain 4×4 real or 2×2 complex matrices. Cayley in 1858 stated the result for 3×3 and smaller matrices, but only published a proof for the 2×2 case. As for $n \times n$ matrices, Cayley stated "..., I have not thought it necessary to undertake the labor of a formal proof of the theorem in the general case of a matrix of any degree". The general case was first proved by Ferdinand Frobenius in 1878. ». Il me semble que la preuve de Frobenius est celle de la proposition 8.2.1 mais je dois encore aller vérifier dans l'article original en allemand.

7.3.1 Théorème. *Soit $f \in \mathcal{L}(E)$. Le polynôme caractéristique χ_f annule f : $\chi_f(f) = 0$.*

Nous allons donner ici une preuve directe de ce théorème, par le calcul. Cependant nous adopterons l'attitude suivante vis à vis de ce théorème.

7.3.2 Remarque. Il nous est apparu qu'il est possible d"établir toute la théorie de la réduction des endomorphismes sans jamais faire usage du théorème de Cayley–Hamilton. En procédant de la sorte, on obtient pour ainsi dire à chaque étape une nouvelle façon de démontrer ce théorème.

C'est la voie que nous allons suivre. En cours de route, nous soulignons les raccourcis que le théorème de Cayley–Hamilton permet de prendre, et bien sûr nous donnons explicitement les diverses preuves de ce théorème auxquelles nous avons pu penser.

Ce n'est qu'avec extrême circonspection que nous proposons cette preuve directe du théorème de Cayley–Hamilton, inquiets que nous sommes qu'elle ne pousse le lecteur imprudent dans le piège éculé suivant.

7.3.3 Mise en garde. Nous conseillons au lecteur de se demander pourquoi la preuve consistant à dire « j'évalue $\chi_A(X) = \det(X.\mathrm{Id} - A)$ en A, ainsi $\chi_A(A) = \det(A \times \mathrm{Id} - A) = \det(0) = 0$ » est irrémédiablement privée de sens. On comparera utilement cette situation à la preuve de la proposition 7.1.2.

7.3.4 Preuve directe du théorème de Cayley–Hamilton. Le point de départ est la formule de Cramer pour la matrice $X.\mathrm{Id} - A \in \mathcal{M}_n(\mathbf{k}[X])$ (voir proposition 2.5.1),

$$(X.\mathrm{Id} - A) \times \mathrm{Com}(X.\mathrm{Id} - A) = \det(X.\mathrm{Id} - A).\mathrm{Id} = \chi_A(X).\mathrm{Id}. \tag{7.3.4.1}$$

En écrivant $\mathrm{Com}(X.\mathrm{Id} - A) = C_0 + X.C_1 + \cdots + X^{n-1}.C_{n-1}$, cela donne

$$\begin{aligned}\chi_A(X).\mathrm{Id} &= (X.\mathrm{Id} - A)(C_0 + X.C_1 + \cdots + X^{n-1}.C_{n-1}) \\ &= -AC_0 + X.(C_0 - AC_1) + \cdots + X^{n-1}.(C_{n-2} - AC_{n-1}) + X^n.C_{n-1}.\end{aligned}$$

Ainsi, notant $\chi_A = X^n + a_{n-1}X^{n-1} + \cdots + a_0$, on a

$$\begin{aligned} -AC_0 &= a_0\mathrm{Id} \\ C_0 - AC_1 &= a_1\mathrm{Id} \\ &\vdots \\ C_{n-2} - AC_{n-1} &= a_{n-1}\mathrm{Id} \\ C_{n-1} &= \mathrm{Id}. \end{aligned}$$

3. https://en.wikipedia.org/wiki/Cayley-Hamilton_theorem

Finalement

$$\begin{aligned}\chi_A(A) &= A^n + a_{n-1}A^{n-1} + \cdots + a_1 A + a_0\mathrm{Id} \\ &= A^n C_{n-1} + A^{n-1}(C_{n-2} - AC_{n-1}) + \cdots + A(C_0 - AC_1) - AC_0 \\ &= 0.\end{aligned}$$

□

7.4 – Endomorphismes trigonalisables

7.4.1 Définition. Un endomorphisme $f \in \mathcal{L}(E)$ est *trigonalisable* s'il existe une base $\mathcal{B}$ de E telle que la matrice $\mathrm{Mat}_{\mathcal{B}}(f)$ est triangulaire supérieure.

7.4.1.1 Remarque.
f trigonalisable $\Leftrightarrow$ $\exists$ une base $(e_1, \ldots, e_n)$ t.q. chaque $\mathrm{Vect}(e_1, \ldots, e_p)$ est stable par f
$\Leftrightarrow$ $\exists$ une filtration $\{0\} = F_0 \subsetneq F_1 \subsetneq \ldots \subsetneq F_n = E$ t.q. chaque F_p est stable par f.

7.4.1.2 Exercice. Les deux notions de trigonalisabilité supérieure et inférieure sont elles équivalentes ?

7.4.2 Théorème. *Les trois propositions suivantes sont équivalentes :*
(i) *f trigonalisable ;*
(ii) *χ_f scindé ;*
(iii) *μ_f scindé.*

L'énoncé contient en particulier l'équivalence "χ scindé $\Leftrightarrow$ μ scindé" qui n'a rien d'évident.

L'implication (ii) $\Rightarrow$(iii) est un corollaire direct du théorème de Cayley–Hamilton. Ici nous allons démontrer directement (i) $\Leftrightarrow$ (ii) et (i) $\Leftrightarrow$ (iii), sans utiliser Cayley–Hamilton. Ceci nous permettra de donner une nouvelle preuve de ce théorème, voir 7.4.3.

Preuve. L'implication (i) $\Rightarrow$ (ii) est une conséquence directe du calcul du déterminant d'une matrice triangulaire supérieure.

Montrons (ii) $\Rightarrow$ (i) par récurrence sur la dimension. Si $\dim(E) \leqslant 1$ le résultat est trivial. Supposons $\dim(E) > 1$ et le résultat démontré pour les endomorphismes d'un espace vectoriel de dimension strictement inférieure. On suppose χ_f scindé, donc il possède au moins une racine $\lambda \in \mathbf{k}$. Cette racine λ est une valeur propre de f, considérons l'espace propre E_λ correspondant. C'est un sous-espace stable par f, et on a un endomorphisme $\bar{f}_{E_\lambda}$ induit sur le quotient E/E_λ. Le polynôme caractéristique de $\bar{f}_{E_\lambda}$ divise celui de f d'après 7.1.3, donc il est scindé lui aussi. Puisque λ est valeur propre, on a $E_\lambda \neq \{0\}$, et donc $\dim(E/E_\lambda) < \dim(E)$. L'hypothèse de récurrence s'applique donc à $\bar{f}_{E_\lambda}$, ce qui prouve qu'il est trigonalisable.

Soit $(\bar{e}_{p+1}, \ldots, \bar{e}_n)$ une base de E/E_λ trigonalisante pour $\bar{f}_{E_\lambda}$, et $(e_1, \ldots, e_p)$ n'importe quelle base pour E_λ (où $p = \dim(E_\lambda)$, et $n = \dim(E)$). Alors $(e_1, \ldots, e_n)$ est une base de E d'après le lemme 5.4.4, et la matrice de f dans cette base est triangulaire supérieure par blocs

$$\begin{pmatrix} \lambda.\mathrm{Id}_p & * \\ 0 & T \end{pmatrix},$$

où T est la matrice de $\bar{f}_{E_\lambda}$ dans la base $(\bar{e}_{p+1}, \dots, \bar{e}_n)$, qui est triangulaire supérieure. On en conclut que la base $(e_1, \dots, e_n)$ est trigonalisante pour f, ce qui achève la preuve.

Montrons (i) $\Rightarrow$ (iii). On suppose f trigonalisable ; soit $(e_1, \dots, e_n)$ une base de E dans laquelle la matrice de f est

$$\begin{pmatrix} a_1 & * & \dots & * \\ & \ddots & \ddots & \vdots \\ & & \ddots & * \\ & & & a_n \end{pmatrix},$$

et notons $F_0 = \{0\}$, et $F_p = \mathrm{Vect}(e_1, \dots, e_p)$ pour tout $p = 1, \dots, n$. Pour tout p, on a $(f - a_p.\mathrm{id})(F_p) \subseteq F_{p-1}$, donc

$$(f - a_1.\mathrm{id}) \cdots (f - a_n.\mathrm{id})(F_n = E) \subseteq F_0 = \{0\}.$$

Ainsi le polynôme $\prod_p (X - a_p)$ est annulateur pour f, et comme il est scindé μ_f est nécessairement scindé lui aussi.

Montrons (iii) $\Rightarrow$ (i) par récurrence sur la dimension. La preuve ressemble beaucoup à celle de (ii) $\Rightarrow$ (i). On suppose μ_f scindé ; soit $\lambda \in \mathbf{k}$ une de ses racines. Alors λ est racine de χ_f par le lemme 7.2.10, donc c'est une valeur propre de f. On considère E_λ l'espace propre associé à λ. C'est un sous-espace stable par f, et on peut considérer l'endomorphisme $\bar{f}_{E_\lambda}$ induit sur le quotient E/E_λ. Le polynôme $\mu_{\bar{f}_{E_\lambda}}$ divise μ_f par la proposition 7.2.12, donc il est scindé lui aussi. Puisque λ est valeur propre, E_λ est non trivial, et donc $\dim(E/E_\lambda) < \dim(E)$. Par hypothèse de récurrence, on a donc que $\bar{f}_{E_\lambda}$ est trigonalisable. On conclut alors exactement comme dans la preuve de (ii) $\Rightarrow$ (i). □

7.4.3 Preuve de Cayley–Hamilton en partant du cas trigonalisable. Dans la preuve de (i) $\Rightarrow$ (iii) du théorème 7.4.2 ci-dessus, on a en fait démontré que si f est trigonalisable alors χ_f annule f.

En général, il existe une extension $\mathbf{k}'$ de $\mathbf{k}$ telle que χ_f soit scindé sur $\mathbf{k}'$. Alors f est trigonalisable sur $\mathbf{k}'$, et on en déduit que χ_f (qui ne dépend pas du corps de base) est annulateur pour f.

On a ainsi démontré le théorème de Cayley–Hamilton pour un endomorphisme d'un espace vectoriel de dimension finie, ou de manière équivalente pour les matrices à coefficients dans un corps. Ceci permet par l'argument habituel de démontrer l'énoncé pour des coefficients dans un anneau (commutatif) arbitraire : on considère une matrice A dont les coefficients sont des indéterminées (a_{ij}). En raisonnant dans le corps $\mathbf{Q}(a_{ij})$ on peut appliquer l'énoncé démontré plus haut, et obtenir ainsi l'identité $\chi_A(A) = 0$, qui vit en fait dans $\mathcal{M}_n(\mathbf{Z}[a_{ij}])$. Par spécialisation, ceci prouve $\chi_A(A) = 0$ pour A à coefficients dans n'importe quel anneau commutatif. □

7.4.4 Proposition. *Si f et g sont trigonalisables et $fg = gf$, alors f et g sont simultanément trigonalisables.*

Preuve. La preuve suit de très près celle du fait qu'un endomorphisme à polynôme caractéristique scindé est trigonalisable. À nouveau, le résultat est clair si $\dim(E) \leqslant 1$. Supposons $\dim(E) > 1$, et le résultat démontré pour tout endomorphisme en dimension strictement inférieure à $\dim(E)$.

Puisque f est trigonalisable, son polynôme caractéristique est scindé, et donc f possède une valeur propre λ. L'espace propre associé E_λ est stable par g puisque f et g commutent

(voir lemme 7.2.3). Puisque λ est valeur propre, E_λ est non-trivial, et donc $\dim(E/E_\lambda) < \dim(E)$.

On pourra donc appliquer l'hypothèse de récurrence à $\bar{f}_{E_\lambda}$ et $\bar{g}_{E_\lambda}$ dès qu'on saura qu'ils sont trigonalisables et qu'ils commutent. Pour vérifier le premier point, on utilise le théorème 7.4.2 : χ_f est scindé, $\chi_{\bar{f}_{E_\lambda}}$ divise χ_f, donc il est scindé lui aussi, et ainsi $\bar{f}_{E_\lambda}$ est trigonalisable ; le même argument s'applique à g. Nous laissons au lecteur le soin de vérifier le second point.

Ainsi, il existe une base $\bar{\mathcal{B}} = (\bar{e}_{p+1}, \dots, \bar{e}_n)$ de E/E_λ trigonalisante pour $\bar{f}_{E_\lambda}$ et $\bar{g}_{E_\lambda}$. Considérons d'autre part une base $\mathcal{B}_{E_\lambda} = (e_1, \dots, e_p)$ de E_λ trigonalisante pour g_{E_λ} (trigonalisable lui aussi, car à polynôme caractéristique scindé). Alors $(e_1, \dots, e_p, e_{p+1}, \dots, e_n)$ est une base de E dans laquelle les matrices de f et g sont respectivement

$$\begin{pmatrix} \lambda.\mathrm{Id}_p & * \\ 0 & T \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} S_1 & * \\ 0 & S_2 \end{pmatrix},$$

où T et S_2 sont les matrices dans $\mathcal{B}_{E_\lambda}$ de $\bar{f}_{E_\lambda}$ et $\bar{g}_{E_\lambda}$ respectivement, qui sont toutes les deux triangulaires supérieures, et S_1 est la matrice dans $\bar{\mathcal{B}}$ de g_{E_λ}, elle aussi triangulaire supérieure. Ainsi les matrices de f et g sont toutes les deux triangulaires supérieures dans cette base, et le résultat est démontré. □

7.4.5 Remarque. La réciproque de 7.4.4 est fausse. Il suffit de se baisser pour ramasser un exemple. Pour ceux qui ont mal au dos, voir 7.8.6.1.

7.4.6. Si χ_f est scindé, on a la décomposition de Dunford qui nous dit que f se réduit à une partie diagonalisable et une partie nilpotente ; en particulier f est diagonalisable si et seulement si la partie nilpotente est triviale.

Ceci devrait suffire à motiver l'étude des diagonalisables à la Section 7.5 et des nilpotents à la Section 7.7. Cette étude nous permettra de donner un critère de similitude pour deux endomorphismes à polynôme caractéristique scindé (i.e. un critère de similitude pour les endomorphismes trigonalisables), sans utiliser de technique sophistiquée — croit-on.

7.5 – Lemme des noyaux

7.5.1 Proposition (Lemme des noyaux)**.** *Soit $P_1, \dots, P_r \in \mathbf{k}[X]$ des polynômes deux à deux premiers entre eux. Alors on a une décomposition en somme directe*

$$\ker\big(P_1 \cdots P_r(f)\big) = \bigoplus_{i=1}^{r} \ker\big(P_i(f)\big).$$

Les projecteurs associés à cette décomposition sont des polynômes en f que l'on peut calculer explicitement.

Il y a un léger abus de langage dans l'énoncé. Voici la version tout-à-fait rigoureuse. Notons $K = \ker\big(P_1 \cdots P_r(f)\big)$; c'est un sous-espace stable par f. Pour tout $i = 1, \dots, r$, le projecteur $p_i \in \mathcal{L}(K)$ sur $\ker\big(P_i(f)\big)$ dans la direction de $\bigoplus_{i' \neq i} \ker\big(P_{i'}(f)\big)$ est un polynôme en $f_K \in \mathcal{L}(K)$.

7.5.2 Conditions d'application du lemme des noyaux. La condition "$P_1, \dots, P_r$ deux à deux premiers entre eux" signifie "pour tout $i \neq j$, $\mathrm{pgcd}(P_i, P_j) = 1$", et est plus forte que la condition "$P_1, \dots, P_r$ premiers entre eux (dans leur ensemble)", qui signifie "$\mathrm{pgcd}(P_1, \dots, P_r) = 1$".

7.5.2.1 Exercice. Donner un exemple de trois polynômes qui sont premiers entre eux dans leur ensemble mais pas deux à deux.
(*Indication* : on peut prendre Q_1, Q_2, Q_3 construits comme en 7.5.3 ci-dessous).

7.5.2.2 Situation classique. Le cas archétypique d'application du lemme des noyaux est quand $P_i = H_i^{\alpha_i}$ avec les H_i irréductibles deux à deux distincts. On laisse au lecteur le soin de vérifier que dans ce cas les P_i sont effectivement deux à deux premiers entre eux.

7.5.3. Si $P_1, \dots, P_r \in \mathbf{k}[X]$ sont deux à deux premiers entre eux, alors les $Q_i := \prod_{j\neq i} P_j$ sont premiers entre eux dans leur ensemble.[4] Puisque $\operatorname{pgcd}(Q_1, \dots, Q_r)$ est le générateur unitaire de l'idéal $(Q_1, \dots, Q_r)$, on a dans ce cas $(Q_1, \dots, Q_r) = (1)$. Ceci implique l'existence de $U_1, \dots, U_r \in \mathbf{k}[X]$ tels que

$$1 = U_1Q_1 + \cdots + U_rQ_r \ ;$$

autrement dit on a une relation de Bezout. Cette identité peut être interprétée comme une partition de l'unité, et nous allons voir que c'est exactement le rôle qu'elle joue dans la preuve du lemme des noyaux.

7.5.3.1 Calcul pratique des U_i. Le principe est de raisonner par récurrence sur r et d'utiliser l'algorithme d'Euclide. Pour $r = 2$, il suffit d'appliquer directement l'algorithme d'Euclide à P_1 et P_2. Pour $r \geqslant 3$, supposons par récurrence avoir trouvé $V_1, \dots, V_{r-1}$ tels que

$$1 = V_1R_1 + \cdots + V_{r-1}R_{r-1},$$

où $R_i = \prod_{\substack{j\neq i\\ j<r}} P_j$ pour tout $i = 1, \dots, r-1$. En multipliant par P_r, on obtient

$$P_r = V_1Q_1 + \cdots + V_{r-1}Q_{r-1}. \tag{5.2.i}$$

Puisque P_r est premier avec chacun des P_i, $i = 1, \dots, r-1$, il est premier avec le produit $P_1 \cdots P_{r-1} = Q_r$. En appliquant l'algorithme d'Euclide, on trouve S et T tels que

$$SQ_r + TP_r = 1.$$

En remplaçant P_r par son expression dans (5.2.i), on obtient

$$SQ_r + TV_1Q_1 + \cdots + TV_{r-1}Q_{r-1} = 1,$$

et donc $U_1 = TV_1, \dots, U_{r-1} = TV_{r-1}$ et $U_r = S$ conviennent.

7.5.4 Preuve du lemme des noyaux. On note $Q_i = \prod_{j\neq i} P_j$ pour tout $i = 1, \dots, r$, et on considère des polynômes $U_1, \dots, U_r \in \mathbf{k}[X]$ tels que $1 = U_1Q_1 + \cdots + U_rQ_r$, dont l'existence est garantie par le fait que les P_i sont deux à deux premiers entre eux, voir le paragraphe 7.5.3 ci-dessus. Évaluée en f, cette identité donne

$$\mathrm{id}_E = U_1(f)Q_1(f) + \cdots + U_r(f)Q_r(f). \tag{7.5.4.1}$$

Notons $K = \ker\big(P_1 \cdots P_r(f)\big)$, et $K_i = \ker\big(P_i(f)\big)$ pour tout $i = 1, \dots, r$. Le sous-espace K est stable par f, et l'énoncé se réduit en fait à une affirmation à propos de l'endomorphisme induit $f_K \in \mathcal{L}(K)$. On a $K_i \subseteq K$ pour tout i, et donc $\sum_i K_i \subseteq K$.

4. soit par l'absurde H facteur irréductible commun à tous les Q_i. H divise Q_1, donc il doit diviser un P_i, $i > 1$, disons P_2. H divise aussi Q_2, donc nécessairement aussi un P_j avec $j \neq 2$. H serait alors diviseur irréuctible commun à P_2 et P_j, une contradiction.

Montrons que réciproquement, $K \subseteq \sum_i K_i$. Soit $x \in K$. En évaluant (7.5.4.1) en x, on obtient

$$x = U_1 Q_1(f)(x) + \cdots + U_r Q_r(f)(x), \tag{7.5.4.2}$$

et nous allons montrer que $U_i Q_i(f)(x) \in K_i$ pour tout i, ce qui prouvera bien que $x \in \sum K_i$. Soit $i \in [\![1, r]\!]$. On a

$$P_i(f)(Q_i(f)(x)) = P_i Q_i(f)(x) = P_1 \cdots P_r(f)(x) = 0$$

puisque $x \in K$, donc $Q_i(f)(x) \in K_i$, et *a fortiori* $U_i Q_i(f)(x) \in K_i$, comme on voulait.

Montrons que les K_i sont en somme directe. Il suffit de prouver que K_i et $\sum_{j \neq i} K_j$ sont en somme directe pour tout i (voir l'exercice 5.1.4). Soit $i \in [\![1, r]\!]$ et $x \in K_i \cap (\sum_{j \neq i} K_j)$. Évaluant (7.5.4.1) en x, il vient

$$x = \sum_j U_j Q_j(f)(x) = U_i Q_i(f)(x), \tag{7.5.4.3}$$

puisque $P_i | Q_j$ si $j \neq i$, et donc $Q_j(f)(x) = 0$. Observons au passage que (7.5.4.3) est en accord avec le fait que $U_i Q_i(f)$ doit être le projecteur de K sur K_i. De la même façon, $Q_i(f)$ est nul sur K_j si $j \neq i$, et donc nul sur $\sum_{j \neq i} K_j$. On a donc $Q_i(f)(x) = 0$, et ainsi $x = 0$ par (7.5.4.3).

L'identité (7.5.4.2) prouve que pour $x \in K$, les $U_i Q_i(f)(x)$ sont les composantes de x selon la décomposition $K = \bigoplus_i K_i$, donc le projecteur $p_i \in \mathcal{L}(K)$ sur K_i dans la direction de $\bigoplus_{i' \neq i} K_{i'}$ est $U_i Q_i(f_K) \in \mathcal{L}(K)$, qui est bien un polynôme en f_K. □

7.5.5 Corollaire. *Si P et Q sont premiers entre eux, alors $P(f)_{\ker Q(f)}$ est injectif.*

Preuve. Le noyau de $P(f)_{\ker(Q(f))}$ est $\ker(P(f)) \cap \ker(Q(f))$. D'après le lemme des noyaux, et puisque P et Q sont premiers entre eux, cette intersection est réduite à $\{0\}$. □

On retrouve en particulier que si P est premier à μ_f, alors $P(f)$ est injectif et donc inversible puisque c'est un endomorphisme.

7.5.6 Exercice. 1) Soit $p \in \mathcal{L}(E)$ tel que $p^2 = p$. Montrer que p est un projecteur, c'est-à-dire qu'il existe F et G sous-espaces supplémentaires de E tels que pour tout $x_F \in F$ et $x_G \in G$: $p(x_F + x_G) = x_F$ (p est alors le projecteur sur F dans la direction de G).
2) Soit $s \in \mathcal{L}(E)$ tel que $s^2 = \mathrm{id}$. Montrer que s est une symétrie, c'est-à-dire qu'il existe F et G sous-espaces supplémentaires de E tels que pour tout $x_F \in F$ et $x_G \in G$: $s(x_F + x_G) = x_F - x_G$ (s est alors la symétrie par rapport à F dans la direction de G).

7.6 – Endomorphismes diagonalisables

7.6.1 Définition. *Un endomorphisme $f \in \mathcal{L}(E)$ est* diagonalisable *si E se décompose en somme directe de sous-espaces propres de f.*

Une condition équivalente est qu'il existe une base de E constituée de vecteurs propres pour f ; la matrice de f dans une telle base est diagonale (et pas seulement diagonale par blocs).

7.6.2 Théorème. *Soit $f \in \mathcal{L}(E)$. Les trois propositions suivantes sont équivalentes :*

(i) *l'endomorphisme f est diagonalisable ;*
(ii) *il existe un polynôme annulateur de f scindé à racines simples ;*
(iii) *le polynôme minimal μ_f est scindé à racines simples.*

Preuve. Montrons "(i) $\Rightarrow$ (ii)". Si f est diagonalisable, il existe une base de E dans laquelle la matrice de f est diagonale. Notons $\alpha_1, \dots, \alpha_r$ les scalaires apparaissant sur la diagonale de cette matrice ; on autorise les répétitions, et suppose ainsi les α_i deux à deux distincts. Un calcul direct montre que $(X - \alpha_1) \cdots (X - \alpha_r)$ annule f. Puisque les α_i sont deux à deux distincts, ce polynôme est scindé à racines simples, et donc (ii) est vérifiée.

L'implication "(ii) $\Rightarrow$ (iii)" est purement algébrique. Soit P polynôme annulateur de f scindé à racines simples. Alors μ_f divise P, donc μ_f est lui-même scindé à racines simples, ce qui prouve (iii).

Nous allons maintenant prouver l'implication "(iii) $\Rightarrow$ (i)", ce qui concluera la preuve du théorème. Notons $\mu_f = \prod_{i=1}^{r}(X - \alpha_i)$, où les α_i sont des scalaires deux à deux distincts. Ainsi les polynômes $X - \alpha_1, \dots, X - \alpha_r$ sont deux à deux premiers entre eux. On peut donc leur appliquer le lemme des noyaux, ce qui donne

$$\ker(\mu_f(f)) = \bigoplus_{i=1}^{r} E_{\alpha_i}(f).$$

Puisque μ_f annule f, on a $\ker(\mu_f(f)) = E$, et donc l'égalité ci-dessus est une décomposition de E en somme de sous-espaces propres pour f, ce qui prouve (i). □

7.6.3 Proposition. *Si χ_f est scindé à racines simples, alors f est diagonalisable.*

7.6.3.1 Mise en garde. La réciproque de la proposition 7.6.3 est notoirement fausse. La diagonalisabilité de f implique que χ_f est scindé (c'est un cas particulier du théorème 7.4.2 !), mais pas nécessairement à racines simples.

La proposition 7.6.3 peut s'obtenir comme un corollaire du théorème précédent, en utilisant le théorème de Cayley–Hamilton, mais on a fait le choix de ne voir le théorème de Cayley–Hamilton que comme une conséquence de tout le reste, et donc de ne jamais l'utiliser dans les preuves.

Preuve. Si χ_f est scindé à racines simples, alors il existe $\alpha_1, \dots, \alpha_n \in \mathbf{k}$ ($n = \dim(E)$) deux à deux distincts tels que $\chi_f = \prod_{i=1}^{n}(X - \alpha_i)$; en effet, le polynôme caractéristique est de degré n. Pour tout $i = 1, \dots, n$ l'endomorphisme $f - \alpha_i.\mathrm{id}$ n'est pas inversible donc son noyau E_{α_i} est non nul. D'après le lemme des noyaux,

$$\ker(\chi_f(f)) = \bigoplus_{i=1}^{n} E_{\alpha_i},$$

donc $\ker(\chi_f(f))$ a dimension au moins n. Pour conclure, on a nécessairement $\ker(\chi_f(f)) = E$ et chaque E_{α_i} est de dimension 1. □

7.6.4 Proposition. *Un endomorphisme diagonalisable est semi-simple.*

Preuve. Soit $f \in \mathcal{L}(E)$ diagonalisable, et $(e_1, \dots, e_n)$ une base de E constituée de vecteurs propres. Posons $F_i = \mathrm{Vect}(e_i)$ pour tout $i = 1, \dots, n$. Alors chaque F_i est stable par f, et $E = \bigoplus_{i=1}^{n} F_i$. Pour tout i, on a $\dim(F_i) = 1$, donc l'endomorphisme induit f_{F_i} est simple. □

En vertu du théorème 6.2.1, les endomorphismes diagonalisables sont donc également semi-simple-bis (voir définition 6.1.10). Il est bon toutefois de savoir démontrer cette propriété directement ; c'est la proposition 7.6.6 ci-dessous.

7.6.5 Lemme. *Soit $f \in \mathcal{L}(E)$ un endomorphisme diagonalisable. Tout sous-espace stable par f se décompose selon les sous-espaces propres de f ; autrement dit, si F est stable par f, alors*

$$F = \bigoplus\nolimits_{\alpha\in \mathrm{Sp}(f)} (F \cap E_\alpha(f)).$$

7.6.5.1 Attention. Pour une décomposition en somme directe $E = \bigoplus E_i$ arbitraire et F un sous-espace de E, si les $F \cap E_i$ sont bien en somme directe, il est grossièrement faux que leur somme égale F : en général l'inclusion $\bigoplus_i (F \cap E_i) \subseteq F \cap (\bigoplus_i E_i)$ peut être stricte.. C'est le cas dans la situation (très simple) représentée ci-dessous.

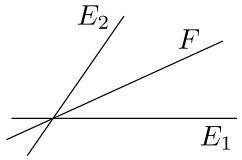


FIGURE 7.1 – Une situation dans laquelle $\bigoplus_i (F \cap E_i) \subsetneq F \cap (\bigoplus_i E_i)$

Preuve du lemme. Puisque F est stable par f, on peut considérer l'endomorphisme induit $f_F \in \mathcal{L}(F)$. On a $F \cap E_\alpha(f) = \ker(f_F - \alpha.\mathrm{id}_F)$ pour tout $\alpha \in \mathbf{k}$. Le polynôme minimal μ_f est scindé à racines simples, et annule f donc aussi f_F. On en déduit que f_F est diagonalisable, et donc

$$F = \bigoplus\nolimits_{\alpha\in\mathbf{k}} \ker(f_F - \alpha.\mathrm{id}_F) = \bigoplus\nolimits_{\alpha\in \mathrm{Sp}(f)} (F \cap E_\alpha(f)).$$

□

7.6.6 Proposition. *Si $f \in \mathcal{L}(E)$ est diagonalisable, alors tout sous-espace stable par f possède un supplémentaire dans E stable par f lui aussi.*

Preuve. Soit F stable par f. On a

$$F = \bigoplus\nolimits_{\alpha\in \mathrm{Sp}(f)} (F \cap E_\alpha(f)).$$

d'après le lemme 7.6.5. Pour chaque $\alpha \in \mathrm{Sp}(f)$, choisissons F'_α supplémentaire de $F \cap E_\alpha(f)$ dans $E_\alpha(f)$. Alors le sous-espace

$$F' = \bigoplus\nolimits_{\alpha\in \mathrm{Sp}(f)} F'_\alpha$$

est un supplémentaire de F dans E, et il est stable par f. □

7.6.7 Lemme. *Soit $f \in \mathcal{L}(E)$ un endomorphisme diagonalisable. Pour tout $\alpha \in \mathbf{k}$, la dimension de l'espace propre E_α est égale à la multiplicité de α comme racine du polynôme caractéristique χ_f.*

Preuve. Puisque f est diagonalisable, on a $E = \bigoplus_{\alpha\in \mathrm{Sp}(f)} E_\alpha$. Les sous-espaces propres étant stables par f, on a donc

$$\chi_f = \prod\nolimits_{\alpha\in \mathrm{Sp}(f)} \chi_{f_{E_\alpha}}.$$

par la proposition 7.1.3. Or pour tout scalaire α, $f_{E_\alpha} = \mathrm{id}_{E_\alpha}$, donc $\chi_{f_{E_\alpha}} = (X-\alpha)^{\dim(E_\alpha)}$, donc

$$\chi_f = \prod\nolimits_{\alpha\in \mathrm{Sp}(f)} (X-\alpha)^{\dim(E_\alpha)},$$

et le résultat est démontré. □

7.6.8 Proposition. *Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes diagonalisables. Alors les trois propositions suivantes sont équivalentes :*
(i) *f et g sont semblables ;*
(ii) $\chi_f = \chi_g$ *;*
(iii) *pour tout $\lambda \in \mathbf{k}$, $\dim(E_\lambda(f)) = \dim(E_\lambda(g))$.*

Preuve. L'implication "(i) $\Rightarrow$ (ii)" a déjà été vue ref. L'implication "(ii) $\Rightarrow$ (iii)" est une conséquence directe du lemme 7.6.7.

Nous allons montrer "(iii) $\Rightarrow$ (i)", ce qui achèvera la preuve de la proposition. La condition (i) nous dit d'ores et déjà que f et g ont le même spectre. De plus, pour toute valeur propre λ de f et g, puisque $\dim(E_\lambda(f)) = \dim(E_\lambda(g))$, il existe un isomorphisme $\varphi_\lambda : E_\lambda(f) \simeq E_\lambda(g)$. Puisque $f_{E_\lambda(f)} = \mathrm{id}_{E_\lambda(f)}$ et $g_{E_\lambda(g)} = \mathrm{id}_{E_\lambda(g)}$, on a

$$f_{E_\lambda(f)} = \varphi_\lambda \circ g_{E_\lambda(g)} \circ \varphi_\lambda^{-1}.$$

On en déduit que f et g sont semblables en appliquant le lemme 6.3.2. □

On dit que deux endomorphismes f et g sont *simultanément diagonalisables* s'il existe une base de E constituée de vecteurs qui sont à la fois propres pour f et pour g.

7.6.9 Théorème. *Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes diagonalisables. Les deux conditions suivantes sont équivalentes :*
(i) *f et g sont simultanément diagonalisables ;*
(ii) *f et g commutent.*

Preuve. L'implication "(ii) $\Rightarrow$ (i)" est une conséquence directe du fait que le produit de deux matrices diagonalisables est commutatif : si $D_1 = \mathrm{diag}(\alpha_1, \dots, \alpha_n)$ et $D_2 = \mathrm{diag}(\beta_1, \dots, \beta_n)$, alors

$$D_1 D_2 = \mathrm{diag}(\alpha_1\beta_1, \dots, \alpha_n\beta_n) = D_2 D_1.$$

Montrons la réciproque. On considère la décomposition de E selon les sous-espaces propres de f,

$$E = \bigoplus\nolimits_{\lambda \in \mathrm{Sp}(f)} E_\lambda(f).$$

Puisque g commute avec f, il résulte du lemme 7.2.3 que pour tout $\lambda \in \mathrm{Sp}(f)$, $E_\lambda(f)$ est stable par g. L'endomorphisme induit $g_{E_\lambda(f)} \in \mathcal{L}(E_\lambda(f))$ est diagonalisable, car il est annulé par μ_g qui est scindé à racines simples. Il existe donc une base $\mathcal{B}_\lambda$ de $E_\lambda(f)$ qui est constituée de vecteurs propres pour $g_{E_\lambda(f)}$, et donc aussi pour g. Ces vecteurs sont des vecteurs de $E_\lambda(f)$, donc ils sont automatiquement propres pour f. La concaténation des familles $\mathcal{B}_\lambda$, $\lambda \in \mathrm{Sp}(f)$, est une base de E, qui est constituée de vecteurs propres à la fois pour f et pour g. Ceci prouve que f et g sont simultanément diagonalisables. □

Il est aussi possible de prouver le théorème 7.6.9 directement par le calcul. Ceci prend la forme suivante.

7.6.10 Version matricielle de la simultanée diagonalisabilité. Soit $n \in \mathbf{N}^*$, et $p_1, \dots, p_r \in \mathbf{N}^*$ tels que $n = p_1 + \cdots + p_r$. Considérons les deux matrices carrées de taille n

$$A = \begin{pmatrix} \alpha_1 \mathrm{Id}_{p_1} & & \\ & \ddots & \\ & & \alpha_r \mathrm{Id}_{p_r} \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} B_{11} & \cdots & B_{1r} \\ \vdots & & \vdots \\ B_{r1} & \cdots & B_{rr} \end{pmatrix} :$$

A est diagonale, et B est une matrice par blocs où chaque B_{ij} est de taille $p_i \times p_j$. On suppose $\alpha_1, \dots, \alpha_r$ deux à deux disjoints. Alors A et B commutent si et seulement si B est diagonale par blocs,

$$B = \begin{pmatrix} B_{11} & & 0 \\ & \ddots & \\ 0 & & B_{rr} \end{pmatrix}.$$

En effet, il résulte des règles de calcul pour les matrices par blocs que

$$AB = \begin{pmatrix} \alpha_1 B_{11} & \cdots & \alpha_1 B_{1r} \\ \vdots & & \vdots \\ \alpha_r B_{r1} & \cdots & \alpha_r B_{rr} \end{pmatrix} \quad : \quad \text{et} \quad BA = \begin{pmatrix} \alpha_1 B_{11} & \cdots & \alpha_r B_{1r} \\ \vdots & & \vdots \\ \alpha_1 B_{r1} & \cdots & \alpha_r B_{rr} \end{pmatrix}.$$

Puisque les α_i sont deux à deux distincts, on a donc $AB = BA$ si et seulement si $B_{ij} = 0$ si $i \neq j$, comme il fallait démontrer. Ceci prouve que les sous-espaces propres de A sont tous stables par B.

Procédons ensuite à la seconde partie de la preuve. Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes diagonalisables qui commutent. En considérant $\mathcal{B}$ une base de E diagonalisante pour f, on a d'après ce qui précède

$$A = \begin{pmatrix} \alpha_1 \mathrm{Id}_{p_1} & & \\ & \ddots & \\ & & \alpha_r \mathrm{Id}_{p_r} \end{pmatrix} \quad \text{et} \quad B = \begin{pmatrix} B_1 & & 0 \\ & \ddots & \\ 0 & & B_r \end{pmatrix},$$

où $\{\alpha_1, \dots, \alpha_r\} = \mathrm{Sp}(f)$. Puisque g est diagonalisable, chaque bloc B_i est diagonalisable, donc il existe des matrices inversibles $P_1, \dots, P_r$ de tailles $p_1, \dots, p_r$ telles que pour tout i, $P_i^{-1} B_i P_i = \Delta_i$ est une matrice diagonale. Alors $P = \mathrm{diag}(P_1, \dots, P_r)$ est une matrice (diagonale par blocs) inversible, et

$$P^{-1}AP = \begin{pmatrix} \alpha_1 \mathrm{Id}_{p_1} & & \\ & \ddots & \\ & & \alpha_r \mathrm{Id}_{p_r} \end{pmatrix} \quad \text{et} \quad P^{-1}BP = \begin{pmatrix} \Delta_1 & & 0 \\ & \ddots & \\ 0 & & \Delta_r \end{pmatrix}$$

sont toutes les deux diagonales (pas seulement diagonales par blocs), ce qui prouve que f et g sont simultanément diagonalisables.

7.6.11 Exercice. Soit $f_1, \dots, f_N \in \mathcal{L}(E)$ des endomorphismes diagonalisables commutant deux à deux. Montrer qu'ils sont simultanément diagonalisables dans leur ensemble.

Cette propriété permet de démontrer que les représentations irréductibles complexes d'un groupe abélien fini sont toutes de dimension 1.

7.7 – Endomorphismes nilpotents

Dans cette section, étant donné un endomorphisme $f \in \mathcal{L}(E)$, nous noterons $K_i = \ker(f^i)$ pour tout $i \in \mathbf{N}$ (ou $K_i(f)$ s'il y a un risque d'ambiguïté), et $k_i = \dim(K_i)$ (ou $k_i(f)$ si nécessaire). La proposition suivante ne nécessite pas que f soit nilpotent.

7.7.1 Proposition. *Soit $f \in \mathcal{L}(E)$.*

(i) *La suite de sous-espaces $(K_i)_{i \geqslant 0}$ est croissante.*

(ii) *La suite* $(\dim K_{i+1} - \dim K_i)_{i\geqslant 0}$ *est décroissante.*

En particulier, il existe $i_0 \in \mathbf{N}$ tel que :
(i) pour tout $i \leqslant i_0$, $K_{i-1} \subsetneq K_i$;
(ii) pour tout $i \geqslant i_0$, $K_i = K_{i_0}$.
Autrement dit, la suite des noyaux K_i est "d'abord strictement croissante, puis constante".

Preuve. (i) On a $f^{i+1}(x) = f(f^i(x))$, donc $f^{i+1}(x) = 0$ si $f^i(x) = 0$. Ceci prouve l'inclusion $K_i \subseteq K_{i+1}$, et donc la croissance de la suite $(K_i)_{i\geqslant 0}$.

(ii) On a $f(K_{i+1}) \subseteq K_i$, donc en composant f avec la projection $K_i \to K_i/K_{i-1}$, on obtient une application linéaire $^\sharp\bar{f}_{i+1} : K_{i+1} \to K_i/K_{i-1}$. Son noyau est K_i :

$$^\sharp\bar{f}_{i+1}(x) = 0 \Leftrightarrow f(x) \in K_{i-1} \Leftrightarrow f^i(x) = 0.$$

Ainsi, $^\sharp\bar{f}_{i+1}$ induit une application linéaire injective

$$\bar{f}_{i+1} : K_{i+1}/K_i \to K_i/K_{i-1}, \tag{7.7.1.1}$$

et donc $k_{i+1} - k_i \leqslant k_i - k_{i-1}$. □

On peut aussi démontrer directement que la suite est strictement croissante puis stationnaire. Il s'agit de montrer que si $K_i = K_{i+1}$, alors $K_i = K_{i+k}$ pour tout $k \geqslant 0$. Par récurrence, il suffit de montrer que $K_{i+1} = K_{i+2}$ si $K_i = K_{i+1}$. Puisque la suite est croissante, il suffit de montrer que $K_{i+2} \subseteq K_{i+1}$. Soit $x \in K_{i+2}$. Alors $f^i(f(x)) = 0$, donc $f(x) \in K_{i+1}$; puisque $K_{i+1} = K_i$, $f(x) \in K_i$, donc $f^i(f(x)) = 0$, et ainsi $x \in K_{i+1}$, comme il fallait démontrer.

7.7.2 Remarque. On a des résultats analogues avec la suite des images des itérés de f. On peut les prouver directement comme ci-dessus, ou bien les déduire sans effort en appliquant l'énoncé donné plus haut à l'endomorphisme transposé $f^\mathsf{T} \in \mathcal{L}(E^\star)$.

7.7.3 Définition. *Un endomorphisme $f \in \mathcal{L}(E)$ est* nilpotent *s'il existe un entier $p \in \mathbf{N}$ tel que $f^p = 0$. Le plus petit entier naturel satisfaisant à cette condition s'appelle l'*indice de nilpotence *de f.*

7.7.4 Exercice. Soit $f \in \mathcal{L}(E)$. Montrer que les propositions suivantes sont équivalentes :
(i) f est nilpotent ;
(ii) il existe $k \in \mathbf{N}$ tel que X^k est un polynôme annulateur de f ;
(iii) il existe $k \in \mathbf{N}$ tel que $\mu_f = X^k$;
(iv) il existe $k \in \mathbf{N}$ tel que $\chi_f = X^k$.
Montrer que dans ces conditions, l'indice de nilpotence de f est le degré du polynôme minimal μ_f.

7.7.5 Corollaire. *Si $f \in \mathcal{L}(E)$ est nilpotent, alors son indice de nilpotence est inférieur à la dimension de E.*

Preuve. L'indice de nilpotence de f est le plus petit entier p tel que $K_p = E$, autrement dit tel que $k_p = \dim(E)$. Puisque la suite $(k_i)_{i\geqslant 0}$ est d'abord strictement croissante puis stationnaire, cet entier p est au plus $\dim(E)$. □

7.7.6 Théorème. *Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes nilpotents. Les deux propositions suivantes sont équivalentes :*
(i) f et g sont semblables ;
(ii) pour tout i, $\dim\big(\ker(f^i)\big) = \dim\big(\ker(g^i)\big)$.

Preuve. (i) ⇒ (ii) est immédiat. La stratégie pour prouver la réciproque est d'écrire tout endomorphisme nilpotent sous une forme normale ne dépendant que des dimensions des noyaux de ses itérés ; ainsi si f et g satisfont à la propriété (ii), alors ils ont la même forme normale et sont donc semblables. Autrement dit, on va construire deux bases $\mathcal{B}_f$ et $\mathcal{B}_g$ telles que

$$\mathrm{Mat}_{\mathcal{B}_f}(f) = J\big((k_i(f))_{i\geqslant 0}\big) \quad \text{et} \quad \mathrm{Mat}_{\mathcal{B}_g}(g) = J\big((k_i(g))_{i\geqslant 0}\big),$$

de sorte que si (ii) vaut, alors $J\big((k_i(f))_{i\geqslant 0}\big) = J\big((k_i(g))_{i\geqslant 0}\big)$, et donc f et g sont semblables.

Soit donc f nilpotent d'indice de nilpotence p, et notons pour tout i, $k_i = \dim(K_i) = \dim\big(\ker(f^i)\big)$ et $s_i = k_i - k_{i-1}$. On commence par choisir une base $(\bar e_1, \dots, \bar e_{s_p})$ de $K_p/K_{p-1} = E/K_{p-1}$. Puisque l'application linéaire

$$\bar f_p : \bar x \in K_p/K_{p-1} \mapsto \overline{f(x)} \in K_{p-1}/K_{p-2}$$

de (7.7.1.1) est injective, la famille $\big(\overline{f(e_1)}, \dots, \overline{f(e_{s_p})}\big)$ de vecteurs de K_{p-1}/K_{p-2} est une famille libre. On la complète en une base

$$\big(\overline{f(e_1)}, \dots, \overline{f(e_{s_p})}, \bar e_{s_p+1}, \dots, \bar e_{s_{p-1}}\big),$$

remarquant au passage que la notation est cohérente puisque $s_p \leqslant s_{p-1}$ d'après la proposition 7.7.1.

On obtient ainsi par récurrence des vecteurs

$$e_1, \dots, e_{s_p}, e_{s_p+1}, \dots\dots, e_{s_2}, e_{s_2+1}, \dots, e_{s_1} \in E$$

tels que pour tout $i = p, \dots, 1$, (i) les vecteurs $e_{s_{i+1}+1}, \dots, e_{s_i}$ sont dans K_i, et (ii) les classes des vecteurs

$$f^{p-i}(e_1), \dots, f^{p-i}(e_{s_p}), \dots, f(e_{s_{i+2}+1}), \dots, , f(e_{s_{i+1}}), e_{s_{i+1}+1}, \dots, e_{s_i}$$

constituent une base de K_i/K_{i-1}. Le lemme 5.4.4 nous assure alors que la famille

$$\begin{array}{cccccc}
e_1, \dots, e_{s_p}, & & & & & \\
f(e_1), \dots, f(e_{s_p}), & & e_{s_p+1}, \dots, e_{s_{p-1}}, & & & \\
\vdots & & \vdots & \ddots & & \\
f^{p-1}(e_1), \dots, f^{p-1}(e_{s_p}), & f^{p-2}(e_{s_p+1}), \dots, f^{p-2}(e_{s_{p-1}}), & \dots\dots, & e_{s_{1+1}+1}, \dots, e_{s_1} &
\end{array}$$

est une base de E.

Il reste à ranger ces vecteurs dans le bon ordre pour obtenir la forme normale voulue. Pour tout $i = p, \dots, 1$ et $j = s_{i+1}+1, \dots, s_i$, on pose

$$\mathcal{B}_j = (e_j, f(e_j), \dots, f^{i-1}(e_j)) \ ;$$

le sous-espace $F_j = \mathrm{Vect}(\mathcal{B}_j)$ est stable par f, et

$$\mathrm{Mat}_{\mathcal{B}_j}(f_{F_j}) = \begin{pmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{pmatrix} = J_i \in \mathcal{M}_i(\mathbf{k}). \tag{7.7.6.1}$$

Ensuite on considère la base $\mathcal{B} = (\mathcal{B}_1, \dots, \mathcal{B}_{s_1})$; dans cette base la matrice de f est diagonale par blocs,

$$\mathrm{Mat}_{\mathcal{B}}(f) = \mathrm{diag}\big(\underbrace{J_p, \dots, J_p}_{s_p \text{ blocs}}, \ \dots, \ \underbrace{J_i, \dots, J_i}_{s_i - s_{i+1} \text{ blocs}}, \ \dots, \ \underbrace{J_1, \dots, J_1}_{s_1 - s_2 \text{ blocs}}\big), \tag{7.7.6.2}$$

uniquement déterminée par la suite $\big(k_i(f)\big)_{i\geqslant 0}$ comme il fallait démontrer (noter que $s_i = 0$ pour $i > p$, donc $s_p = s_p - s_{p+1}$). □

7.7.7. La matrice J_i de (7.7.6.1) est appelée *bloc de Jordan* de taille i, et la matrice de (7.7.6.2) *forme normale de Jordan* de f (on parle aussi de *forme réduite*).

Le résultat suivant complète le théorème précédent.

7.7.8 Théorème. *Soit $f, g \in \mathcal{L}(E)$ deux endomorphismes nilpotents mis sous forme de Jordan : on suppose connaître deux bases $\mathcal{B}_f$ et $\mathcal{B}_g$ de E telles que*

$$\mathrm{Mat}_{\mathcal{B}_f}(f) = \mathrm{diag}(J_{i_1}, \dots, J_{i_s}) \quad \textit{et} \quad \mathrm{Mat}_{\mathcal{B}_g}(g) = \mathrm{diag}(J_{j_1}, \dots, J_{j_t}).$$

Les deux conditions suivantes sont équivalentes :
(i) f et g sont semblables ;
(ii) les deux suites $(i_1, \dots, i_s)$ et $(j_1, \dots, j_t)$ sont égales à permutation près. [5]

Dans la preuve du théorème 7.7.6 nous avons utilisé l'implication évidente (ii) ⇒ (i). Il est d'autre part tentant de croire que la preuve que nous avons donnée du théorème 7.7.6 montre au passage (i) ⇒ (ii), mais il n'en est rien. En effet, rien ne garantit *a priori* qu'en appliquant la construction de la preuve de 7.7.6 à un endomorphisme déjà sous forme réduite de Jordan on obtient la même forme réduite de Jordan.

La preuve de l'implication (i) ⇒ (ii) du théorème 7.7.8 ci-dessus est l'objet du problème suivant.

7.7.9 Unicité de la forme normale de Jordan. (Examen 2019 ; inspiré par [**?**]).
1) Soit $n \geqslant 1$. On considère la matrice $A_n = (a_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_n(\mathbf{Q})$ définie par

$$\forall i, j \in [\![1, n]\!] \quad a_{ij} = \begin{cases} j & \text{si } j \leqslant i \\ i & \text{si } j > i. \end{cases}$$

a) Montrer que A_3 est inversible et calculer son inverse.
b) Montrer que la matrice

$$P_n = \begin{pmatrix} 2 & -1 & & & \\ -1 & 1 & & & \\ -1 & 0 & 1 & & \\ \vdots & \vdots & & \ddots & \\ -1 & 0 & & & 1 \end{pmatrix} \in \mathcal{M}_n(\mathbf{Q})$$

est inversible, puis calculer $P_n \times A_n$.
c) Montrer que A_n est inversible pour tout $n \geqslant 1$.
2) On considère $F_1, \dots, F_r$ sous-espaces vectoriels de E tels que $E = \bigoplus_{1 \leqslant i \leqslant r} F_i$.
a) Soit $g \in \mathcal{L}(E)$ tel que pour tout $i = 1, \dots, r$, F_i est stable par g. Montrer que

$$\ker(g) = \bigoplus_{1 \leqslant i \leqslant r} (\ker(g) \cap F_i).$$

b) Soit $f \in \mathcal{L}(E)$ tel que pour tout $i = 1, \dots, r$, F_i est stable par f. Montrer que pour tout $P \in \mathbf{k}[X]$,

$$\ker\big(P(f)\big) = \bigoplus_{1 \leqslant i \leqslant r} (\ker(P(f)) \cap F_i).$$

5. autrement dit, $t = s$, et il existe une permutation $\sigma \in \mathfrak{S}_s$ telle que $j_a = i_{\sigma(a)}$ pour tout $a = 1, \dots, s$.

3) On considère l'endomorphisme f de $\mathbf{k}^r$ défini par la multiplication à gauche par la matrice

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

Calculer $\dim\big(\ker(f^i)\big)$ pour tout $i \in \mathbf{N}$.

4) Soit $f \in \mathcal{L}(E)$ et $b_1, \ldots, b_n$ ($n = \dim(E)$) des entiers. On suppose qu'il existe une décomposition

$$E = \bigoplus_{1 \leqslant r \leqslant n} \bigoplus_{1 \leqslant a \leqslant b_r} F_{r,a}$$

telle que chaque $F_{r,a}$ est de dimension r et stable par f, et $f_{F_{r,a}}$ est semblable à l'endomorphisme de $\mathbf{k}^r$ de la question 3.

a) En utilisant les questions 2b et 3, calculer $k_i := \dim\big(\ker(f^i)\big)$ pour tout $i \in \mathbf{N}$. En déduire que

$$\begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = A_n \times \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}$$

où A_n est la matrice de la question 1.

b) Conclure que s'il existe une autre décomposition

$$E = \bigoplus_{1 \leqslant r \leqslant n} \bigoplus_{1 \leqslant a \leqslant b'_r} F'_{r,a}$$

telle que chaque $F'_{r,a}$ est de dimension r et stable par f, et $f_{F'_{r,a}}$ est semblable à l'endomorphisme de $\mathbf{k}^r$ de la question 3, alors $b_i = b'_i$ pour tout $i = 1, \ldots, n$. □

7.7.10 Curiosité. Quelles valeurs la suite des k_i peut-elle prendre ? Les seules contraintes sont celles imposées par les inégalités du 7.7.1, qui se réduisent à

$$0 \leqslant k_n - k_{n-1} \leqslant \ldots \leqslant k_2 - k_1 \leqslant k_1 - k_0 = k_1.$$

En effet, ces inégalités donnent une condition nécessaire et suffisante pour que

$$A^{-1} \times \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix} = \begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & \ddots & \ddots & \ddots & \\ & & -1 & 2 & 1 \\ & & & -1 & 1 \end{pmatrix} \times \begin{pmatrix} k_1 \\ \vdots \\ k_n \end{pmatrix}$$

ait toutes ses entrées $\geqslant 0$.

On conclut par le traditionnel énoncé de simultanée propriété-ité.

7.7.11 Proposition. *Si n et n' sont nilpotents et commutent, alors $n + n'$ est nilpotent.*

7.7.12 Proposition. *Un endomorphisme nilpotent est semi-simple si et seulement si il est nul.*

Preuve. Soit $f \in L(E)$ nilpotent et semi-simple. Alors f est aussi semi-simple-bis d'après le théorème 6.2.1. Considérons le noyau de f ; il est stable par f, et doit posséder un supplémentaire F dans E, lui aussi stable par f. Alors l'endomorphisme induit $f_F \in \mathcal{L}(F)$ est injectif, puisque $\ker(f_F) = \ker(f) \cap F = \{0\}$. D'autre part, puisque pour tout entier $N \geqslant 0$, $(f_F)^N = (f^N)_F$, f_F est lui aussi nilpotent. Ceci impose $F = \{0\}$, et donc $\ker(f) = E$, autrement dit $f = 0$. □

7.8 – Endomorphismes à polynôme caractéristique scindé

En vertu du Théorème 7.4.2, cette section pourrait s'appeler *Endomorphismes trigonalisables, II.* Dans toute cette section, on suppose que χ_f est scindé, ou de manière équivalente μ_f est scindé.

7.8.1 Notation. Écrivons

$$\mu = \prod\nolimits_{\lambda\in \mathrm{Sp}} (X-\lambda)^{a_\lambda} \quad \text{et} \quad \chi = \prod\nolimits_{\lambda\in \mathrm{Sp}} (X-\lambda)^{b_\lambda}.$$

7.8.2 Définition. Les *sous-espaces caractéristiques* de f sont les $C_\lambda = \ker\big((f-\lambda.\mathrm{id})^{a_\lambda}\big)$ pour $\lambda \in \mathrm{Sp}(f)$. On a la décomposition

$$E = \bigoplus\nolimits_{\lambda\in \mathrm{Sp}} C_\lambda \tag{7.8.2.1}$$

en somme de sous-espaces stables par f (et par tout g commutant avec f). Les *projecteurs spectraux* $p_\lambda \in \mathcal{L}(E)$, qui s'écrivent

$$p_{\lambda_i}(x_{\lambda_1}, \dots, x_{\lambda_r}) = (0, \dots, x_{\lambda_i}, \dots, 0)$$

dans la décomposition (8.2.2.1), sont des polynômes en f que l'on sait calculer explicitement, voir le lemme des noyaux 7.5.1.

7.8.3 Proposition. *Soit $\lambda \in \mathrm{Sp}(f)$.*
i) L'entier a_λ est l'indice de stagnation de la suite des $K_i(\lambda) = \ker\big((f-\lambda.\mathrm{id})^i\big)$, et l'indice de nilpotence de $f_{C_\lambda} - \lambda.\mathrm{id}_{C_\lambda} \in \mathcal{L}(C_\lambda)$. En particulier, $f_{C_\lambda} - \lambda.\mathrm{id}_{C_\lambda} \in \mathcal{L}(C_\lambda)$ est nilpotent.
ii) L'entier b_λ est la dimension du sous-espace caractéristique C_λ.

On verra plus loin une version un peu plus générale de ce résultat (Proposition 8.2.3). Les preuves sont strictement identiques, et en fait la formulation pour 8.2.3 me semble plus lisible car moins polluée par les notations.

Preuve. i) D'après le lemme des noyaux, $f-\lambda.\mathrm{id}$ est inversible sur $\bigoplus_{\lambda'\neq\lambda} C_{\lambda'}$, donc $K_i(\lambda) = C_\lambda \cap \ker\big((f-\lambda.\mathrm{id})^i\big)$ par le Lemme **??**, et ce noyau s'identifie canoniquement à $\ker\big((f_{C_\lambda} - \lambda.\mathrm{id}_{C_\lambda})^i\big)$.

Par définition de C_λ, on a $K_{a_\lambda}(\lambda) = C_\lambda$. Pour $i \geqslant a_\lambda$, on a $K_{a_\lambda}(\lambda) \subseteq K_i(\lambda) \subseteq C_\lambda$ et donc $K_i(\lambda) = C_\lambda$. Il reste à démontrer que pour $i < a_\lambda$, on a $K_i(\lambda) \subsetneq C_\lambda$, ce qui est garanti par la minimalité de μ.

Précisons ce dernier point. Notons $Q = \prod_{\lambda'\neq\lambda}(X-\lambda')^{a_{\lambda'}}$. Le polynôme Q annule l'endomorphisme de $\bigoplus_{\lambda'\neq\lambda} C_{\lambda'}$ induit par f, puisque $(f-\lambda.\mathrm{id})^{a_\lambda} \circ Q(f) = 0$ et $f-\lambda.\mathrm{id}$ est inversible sur $\bigoplus_{\lambda'\neq\lambda} C_{\lambda'}$. Ainsi si on a $K_i = C_\lambda$, alors le polynôme $(X-\lambda)^i Q$ annule f, donc il est divisible par μ, et nécessairement $i \geqslant a_\lambda$.

ii) La décomposition $\bigoplus C_\lambda$ est une somme de sous-espaces stables, donc $\chi_f = \prod_\lambda \chi_{f_{C_\lambda}}$. Or pour chaque valeur propre λ, $\chi_{f_{C_\lambda}} = (X-\lambda)^{\dim C_\lambda}$, puisque $(f-\lambda.\mathrm{id})_{C_\lambda}$ est nilpotent. On a donc nécessairement $\dim(C_\lambda) = b_\lambda$. □

7.8.4 Application au théorème de Cayley–Hamilton. La Proposition 7.8.3 ci-dessus offre en corollaire le théorème de Cayley–Hamilton pour les endomorphismes trigonalisables. Pour démontrer le théorème en général, on se ramène au cas trigonalisable par un argument d'extension des scalaires comme en 7.4.3.

Soit donc $f \in \mathcal{L}(E)$ trigonalisable, et conservons les notations introduites ci-dessus. Pour chaque $\lambda \in \mathrm{Sp}(f)$, a_λ est l'indice de nilpotence de $f_{C_\lambda} - \lambda.\mathrm{id} \in \mathcal{L}(C_\lambda)$, donc $a_\lambda \leqslant \dim(C_\lambda)$. Puisque d'autre part $b_\lambda = \dim(C_\lambda)$, on a donc $a_\lambda \leqslant b_\lambda$ pour tout $\lambda \in \mathrm{Sp}(f)$, et ainsi $\mu_f | \chi_f$. □

7.8.5 Théorème (Décomposition de Jordan–Chevalley, aussi dite de Dunford)**.** *Soit $f \in \mathcal{L}(E)$ trigonalisable. Il existe une unique paire (d, n) d'endomorphismes de E satisfaisant aux quatre conditions suivantes :*

(i) *$f = d + n$;*
(ii) *d est diagonalisable ;*
(iii) *n est nilpotent ;*
(iv) *$dn = nd$.*

De plus, d et n sont tous les deux des polynômes en f, qu'on sait calculer explicitement.

La condition (iv) de commutativité est aussi importante que le reste ; elle assure que les endomorphismes d et n sont "compatibles". Dans l'exemple 7.8.6, nous allons voir qu'on perd l'unicité si on retire la condition de commutativité, et que la décomposition vérifiant cette condition est "la bonne". Dans la remarque 7.8.7, nous allons voir que la condition de commutativité assure qu'on peut trouver une base de E dans laquelle d et n sont tous les deux sous forme normale.

Nous donnerons la preuve du Théorème 7.8.5 après cet exemple et cette remarque.

7.8.6 Exemple. Considérons la matrice

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

Elle est diagonalisable puisque son polynôme caractéristique est $(X-1)(X-2)$ qui est scindé à racines simples, donc sa décomposition est

$$A = A + 0$$

où A est diagonalisable et 0 est nilpotente.

7.8.6.1 Remarque. Cet exemple donne de façon amusante un exemple de deux matrices triangulaires supérieures qui ne commutent pas. En effet, posons

$$T_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \quad \text{et} \quad T_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

On a (i) T_1 diagonalisable, (ii) T_2 nilpotente, et (iii) $A = T_1 + T_2$. Par unicité de la décomposition, $T_1 + T_2$ n'est pas la décomposition de A (puisque c'est $A + 0$). On a donc nécessairement $T_1 \times T_2 \neq T_2 \times T_1$. Bien sûr, on peut calculer explicitement les deux produits T_2T_1 et T_1T_2 et voir de ses propres yeux qu'ils sont différents.

7.8.7 Remarque. Il existe une base dans laquelle d a une matrice diagonale et n est sous forme réduite de Jordan. Une telle base est en particulier trigonalisante pour $f = d + n$.

En effet, considérons la décomposition de E selon les sous-espaces propres $E_\lambda = E_\lambda(d)$ de $d : E = \bigoplus_{\lambda\in \mathrm{Sp}(d)} E_\lambda$. Puisque d et n commutent, ces sous-espaces sont stables par n (et donc aussi par $f = d + n$). Pour chaque valeur propre λ de d, considérons une base $\mathcal{B}_\lambda$ de E_λ dans laquelle la matrice de l'endomorphisme de E_λ est sous forme réduite de Jordan. Alors $\mathcal{B} = (\mathcal{B}_\lambda)_{\lambda\in \mathrm{Sp}(d)}$ est une base de E comme on voulait. En contemplant les écritures de f, d, n dans cette base, on constate que $\mathrm{Sp}(d) = \mathrm{Sp}(f)$, et les sous-espaces propres de d sont les sous-espaces caractéristiques de f, autrement dit $E_\lambda(d) = C_\lambda(f)$.

Écrivons tout ceci un peu plus explicitement. Appelons $\lambda_1, \dots, \lambda_r$ les valeurs propres de d, et pour tout $i = 1, \dots, r$, notons E_i l'espace propre $E_{\lambda_i}(d)$, $\mathcal{B}_i$ la base $\mathcal{B}_{\lambda_i}$, f_i, d_i et n_i les endomorphismes de E_i induits respectivement par f, d et n. Par construction, on a

$$\mathrm{Mat}_{\mathcal{B}_i}(d_i) = \lambda_i.\mathrm{Id}_{b_i} \quad \text{et} \quad \mathrm{Mat}_{\mathcal{B}_i}(n_i) = \mathrm{diag}(J_{c_1^i}, \dots, J_{c_{s_i}^i}),$$

où la notation J_c désigne un bloc de Jordan de taille c. Ainsi

$$\mathrm{Mat}_{\mathcal{B}_i}(f_i) = \mathrm{diag}\big(J_{c_1^i}(\lambda_i), \dots, J_{c_{s_i}^i}(\lambda_i)\big),$$

où la notation J_c désigne $J_c + \lambda.\mathrm{Id}_c$. Puisque les E_i sont stables par f, on a

$$\begin{aligned}\mathrm{Mat}_{\mathcal{B}}(f) &= \mathrm{diag}\big(\mathrm{Mat}_{\mathcal{B}_1}(f_1), \dots, \mathrm{Mat}_{\mathcal{B}_r}(f_r)\big)\\ &= \mathrm{diag}\big(J_{c_1^1}(\lambda_1), \dots, J_{c_{s_1}^1}(\lambda_1),\ \dots,\ J_{c_1^r}(\lambda_r), \dots, J_{c_{s_r}^r}(\lambda_r)\big).\end{aligned}$$

La multiplicité de λ_i comme racine de χ_f est $b_i = \sum_{j=1}^{s_i} c_j^i$, et comme racine de μ_f est $a_i = \max(c_1^i, \dots, c_{s_i}^i)$. D'après le Théorème 7.7.6, les tailles $c_1^i, \dots, c_{s_i}^i$ des blocs de Jordan de n_i sont déterminées par les dimensions des noyaux des $n_i^k = (f - \lambda_i.\mathrm{id})_{E_i}$, $k \in \mathbf{N}$.

Les considérations de la remarque précédente contiennent tous les ingrédients nécessaires à la construction d'une décomposition de Dunford. La preuve ci-dessous en fait la synthèse. Notons aussi que la remarque précédente contient les arguments pour démontrer le Théorème 7.8.10 plus loin.

Preuve du Théorème 7.8.5. On note $p_\lambda \in \mathcal{L}(E)$ les projecteurs spectraux ; ce sont des polynômes en f. On a $\mathrm{id}_E = \sum_{\lambda\in \mathrm{Sp}(f)} p_\lambda$, et donc

$$f = f \circ \sum\nolimits_{\lambda\in \mathrm{Sp}(f)} p_\lambda = \underbrace{\sum\nolimits_\lambda \lambda.p_\lambda}_{=:d} + \underbrace{\sum\nolimits_\lambda (f - \lambda.\mathrm{id}) \circ p_\lambda}_{=:n}.$$

Manifestement, d et n satisfont aux propriétés (i)–(iv) du Théorème 7.8.5, et l'existence d'une décomposition est bien démontrée. En outre, les d et n que nous avons construits sont bien des polynômes en f.

Montrons l'unicité de la décomposition. Soit (d', n') une autre paire satisfaisant aux conditions (i)–(iv). *A priori* rien ne dit que d' et n' sont des polynômes en f. Cependant, d' commute à lui-même et à n', donc aussi à $f = d' + n'$. Puisque $d \in \mathbf{k}[f]$, d' commute aussi à d. De la même façon, n' commute à n. On en déduit que $d' - d$ est diagonalisable, et $n - n'$ nilpotent. Enfin, puisque $d + n = d' + n'$, on a $d' - d = n - n'$. Cet endomorphisme est à la fois nilpotent et diagonalisable, donc il est nul (puisqu'il est nilpotent, 0 est son unique valeur propre, et puisqu'il est diagonalisable il est donc nécessairement nul ; on peut aussi appliquer la Proposition 7.7.12). Ainsi $d = d'$ et $n = n'$, ce qui prouve l'unicité. □

7.8.8 Exemple. Considérons la matrice

$$A = \begin{pmatrix} -3 & -2 & -2 \\ -2 & 0 & -1 \\ 10 & 5 & 6 \end{pmatrix}.$$

Son polynôme caractéristique est $\chi = (X-1)^3$, donc A possède un seul sous-espace caractéristique $C_1 = \mathbf{k}^3$, et la décomposition de Jordan–Chevalley de A est

$$A = \mathrm{Id}_3 + (A - \mathrm{Id}_3),$$

où Id_3 est diagonalisable et $A - \mathrm{Id}_3$ est nilpotente (on peut vérifier par le calcul que $(A-\mathrm{Id})^2 = 0$; voir aussi l'exemple 8.6.6).

7.8.9 Proposition. *Soit $f \in \mathcal{L}(E)$ trigonalisable. Alors f est semi-simple si et seulement si il est diagonalisable.*

Le théorème suivant permet de décider en pratique si deux endomorphismes trigonalisables sont semblables, voir 7.8.11.

7.8.10 Théorème (Caractérisation des classes de similitude)**.** *Soit f et g deux endomorphismes trigonalisables. Les deux conditions suivantes sont équivalentes :*

(i) *f et g sont semblables ;*

(ii) *pour tout $\lambda \in \mathbf{k}$ et tout $s \in \mathbf{N}$,*

$$\dim\big(\ker(f-\lambda.\mathrm{id})^s\big) = \dim\big(\ker(g-\lambda.\mathrm{id})^s\big).$$

La preuve de ce théorème est la synthèse d'arguments apparaissant plus haut dans la Remarque 7.8.7.

Preuve du Théorème 7.8.10. L'implication "(i) $\Rightarrow$ (ii)" est claire ; nous allons démontrer la réciproque. Grâce au lemme 6.3.2, il suffit de savoir le faire sous-espace caractéristique par sous-espace caractéristique, et sur chaque sous-espace caractéristique nous allons appliquer le critère 7.7.6.

Soit $\lambda \in \mathbf{k}$. Les endomorphismes $(f-\lambda\mathrm{id})_{C_\lambda(f)} \in \mathcal{L}(C_\lambda(f))$ et $(g-\lambda\mathrm{id})_{C_\lambda(g)} \in \mathcal{L}(C_\lambda(g))$ sont nilpotents ; puisque

$$\ker(f-\lambda\mathrm{id})^i = C_\lambda(f) \cap \ker(f-\lambda\mathrm{id})^i \cong \ker\big((f-\lambda\mathrm{id})_{C_\lambda(f)}\big)^i,$$

et de même $\ker(g-\lambda\mathrm{id})^i \cong \ker(g-\lambda\mathrm{id})^i_{C_\lambda(g)}$, la condition (ii) nous dit que la condition (ii) du théorème 7.7.6 est vérifiée pour $(f-\lambda\mathrm{id})_{C_\lambda(f)} \in \mathcal{L}(C_\lambda(f))$ et $(g-\lambda\mathrm{id})_{C_\lambda(g)} \in \mathcal{L}(C_\lambda(g))$. On en déduit qu'il existe un isomorphisme $\varphi_\lambda : C_\lambda(f) \simeq C_\lambda(g)$ tel que

$$(f-\lambda\mathrm{id})_{C_\lambda(f)} = \varphi_\lambda \circ (g-\lambda\mathrm{id})_{C_\lambda(g)} \circ \varphi_\lambda^{-1},$$

ce qui implique que $f_{C_\lambda(f)} = \varphi_\lambda \circ g_{C_\lambda(g)} \circ \varphi_\lambda^{-1}$. On peut alors conclure grâce au lemme 6.3.2 que f et g sont semblables. □

7.8.11. Le critère de similitude du Théorème 7.8.10 permet de décider de manière algorithmique si deux endomorphismes trigonalisables sont semblables. Faisons-le directement avec des matrices.

On considère $A, B \in \mathcal{M}_n(\mathbf{k})$ et on se demande si elles sont semblables. On commence par calculer les polynômes caractéristiques χ_A et χ_B. S'ils sont distincts on peut s'arrêter tout de suite, A et B ne sont pas semblables. Si $\chi_A = \chi_B$, on décompose ce polynôme en produit de facteurs premiers. S'il n'est pas scindé, pour l'instant on ne sait pas répondre à la question et il faut utiliser des techniques plus élaborées, voir section 8.6.

Supposons donc que $\chi_A = \chi_B$ est scindé, et l'ensemble de ses racines est $\{\lambda_1, \ldots, \lambda_r\}$. Si $\lambda \in \mathbf{k} \setminus \{\lambda_1, \ldots, \lambda_r\}$, $A - \lambda\mathrm{Id}$ et $B - \lambda\mathrm{Id}$ sont inversibles, donc

$$\dim\big(\ker(A - \lambda.\mathrm{Id})^s\big) = \dim\big(\ker(B - \lambda.\mathrm{Id})^s\big) = 0$$

pour tout $s \in \mathbf{N}$. Il faut donc nous concentrer sur $\lambda_1, \ldots, \lambda_r$. Pour chaque $i = 1, \ldots, r$, on calcule successivement les

$$k_s(\lambda_i, A) = \dim\big(\ker(A - \lambda_i.\mathrm{Id})^s\big)$$

jusqu'à trouver un $s_0 \in \mathbf{N}$ tel que $k_{s_0}(\lambda_i, A) = k_{s_0+1}(\lambda_i, A)$. On sait alors par la Proposition 7.7.1 que $k_s(\lambda_i, A) = k_{s_0}(\lambda_i, A)$ pour tout $s \geqslant s_0$. On fait la même chose pour les

$$k_s(\lambda_i, B) = \dim\big(\ker(B - \lambda_i.\mathrm{Id})^s\big).$$

Il n'y a plus qu'à comparer les valeurs obtenues pour les différents $k_s(\lambda_i, A)$ et $k_s(\lambda_i, B)$ pour décider si A et B sont semblables. On retiendra qu'il suffit de calculer un nombre fini de dimensions de noyaux pour pouvoir conclure.

7.8.12 Exercice. Soit $\lambda, \mu \in \mathbf{k}$ distincts. Donner un représentant de toutes les classes de similitude de matrices de $\mathcal{M}_5(\mathbf{k})$ dont le polynôme minimal est $(X - \lambda)^2(X - \mu)$. Bien justifier que les matrices que vous donnerez sont deux à deux non semblables.

Voir aussi Exercice 8.8.6 pour des questions du même goût.

Chapitre 8

Réduction des endomorphismes : Synthèse

8.1 – Sous-espaces cycliques

8.1.1 Définition. *Soit $f \in \mathcal{L}(E)$ et $x \in E$. On appelle* sous-espace cyclique *associé à x, noté F_x, le plus petit sous-espace vectoriel de E stable par f et contenant x.*

On appelle polynôme minimal local *de f en x, noté μ_x, le générateur unitaire de l'idéal $\{P \in \mathbf{k}[X] \mid P(f)(x) = 0\}$.*

Pour commencer, notons qu'il existe effectivement un élément minimal pour l'inclusion dans l'ensemble des sous-espaces stables par f contenant x, puisque "être stable" et "contenir x" sont deux propriétés stables par intersection. Ainsi, on a

$$F_x = \bigcap_{\substack{F \text{ stable par } f \\ \text{et } x \in F}} F.$$

8.1.2 Proposition. *Soit $f \in \mathcal{L}(E)$ et $x \in E$. Le polynôme minimal $\mu_{f_{F_x}}$ de l'endomorphisme $f_{F_x} \in \mathcal{L}(F_x)$ induit par f sur le sous-espace cyclique de x est le polynôme μ_x, polynôme minimal de f en x.*

Preuve. On a $\mu_{f_{F_x}}(f_{F_x}) = 0$, donc $\mu_{f_{F_x}}(f_{F_x})(x) = \mu_{f_{F_x}}(f)(x) = 0$, et ainsi $\mu_x | \mu_{f_{F_x}}$.

Réciproquement, $\ker \mu_x(f)$ est un sous-espace stable par f qui contient x, donc $F_x \subseteq \ker \mu_x(f)$. Autrement dit, $\mu_x(f_{F_x}) = 0$, et $\mu_{f_{F_x}} | \mu_x$.

On a finalement bien $\mu_{f_{F_x}} = \mu_x$ puisque ces deux polynômes sont unitaires. □

8.1.3 Proposition. *Soit $x \in E$. On note $p = \deg(\mu_x)$ et*

$$\mu_x = X^p + a_{p-1}X^{p-1} + \cdots + a_0.$$

Alors la famille $(x, f(x), \ldots, f^{p-1}(x))$ est une base de F_x.

Preuve. On a nécessairement $x \in F_x$. Puisque F_x est stable par f, on a donc aussi $f(x) \in F_x$, puis $f(f(x)) = f^2(x) \in F_x$, et ainsi par récurrence $f^k(x) \in F_x$ pour tout $k \in \mathbf{N}$. On a donc $\mathrm{Vect}(f^k(x), k \in \mathbf{N}) \subseteq F_x$. D'autre part $\mathrm{Vect}(f^k(x), k \in \mathbf{N})$ est manifestement un sous-espace vectoriel de E, stable par f et contenant x, donc par minimalité de F_x on a aussi l'inclusion inverse $F_x \subseteq \mathrm{Vect}(f^k(x), k \in \mathbf{N})$, et finalement $F_x = \mathrm{Vect}(f^k(x), k \in \mathbf{N})$.

Montrons par récurrence sur $k \in \mathbf{N}$ que $f^k(x)$ est combinaison linéaire de $x, f(x), \dots, f^{p-1}(x)$. Si $k \leqslant p-1$ c'est trivial. Si $k \geqslant p$, on suppose $f^{k-1}(x)$ combinaison linéaire de $x, f(x), \dots, f^{p-1}(x)$ par hypothèse de récurrence. Alors $f^k(x) = f(f^{k-1}(x))$ est combinaison linéaire de $f(x), f^2(x), \dots, f^p(x)$. Puisque

$$\mu_x(f)(x) = f^p(x) + a_{p-1}f^{p-1}(x) + \cdots + a_0\mathrm{id}(x) = 0,$$

$f^p(x)$ est combinaison linéaire de $x, f(x), \dots, f^{p-1}(x)$, et finalement $f^k(x)$ est bien combinaison linéaire de $x, f(x), \dots, f^{p-1}(x)$. Ceci démontre que $(x, f(x), \dots, f^{p-1}(x))$ engendre F_x.

Montrons maintenant que cette famille est libre. Soit $\alpha_0, \dots, \alpha_{p-1} \in \mathbf{k}$ tels que

$$\alpha_0 x + \cdots + \alpha_{p-1}f^{p-1}(x) = 0.$$

Alors le polynôme $\alpha_0 + \cdots + \alpha_{p-1}X^{p-1}$ appartient à l'idéal $\{P \in \mathbf{k}[X] \mid P(f)(x) = 0\}$, donc il est divisible par μ_x. Puisque $\deg(\mu_x) = p$, ceci impose que $\alpha_0 = \cdots = \alpha_{p-1} = 0$. □

8.1.4. Écrivons la matrice C_x de $f_{F_x} \in \mathcal{L}(F_x)$ dans la base $(x, f(x), \dots, f^{p-1}(x))$ (on conserve les notations introduites ci-dessus). Pour $k < p-1$ on a $f(f^k(x)) = f^{k+1}(x)$ (pour $k \geqslant p-1$ aussi, d'ailleurs !), tandis que pour $k = p$ on a la relation observée ci-dessus

$$f^p(x) = -a_{p-1}f^{p-1}(x) - \cdots - a_0\mathrm{id}(x).$$

On a donc

$$C_x = \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -a_{p-2} \\ & & 1 & -a_{p-1} \end{pmatrix}.$$

En particulier, on a $\mu_{C_x} = \mu_{f_{F_x}}$, et donc d'après la Proposition 8.1.2,

$$\mu_{C_x} = \mu_x = X^p + a_{p-1}X^{p-1} + \cdots + a_0. \tag{8.1.4.1}$$

8.1.5 Définition. *Soit $P = X^p + a_{p-1}X^{p-1} + \cdots + a_1X + a_0 \in \mathbf{k}[X]$ un polynôme unitaire de degré p. On appelle* matrice compagnon *de P la matrice*

$$\mathrm{Comp}(P) = \begin{pmatrix} 0 & & & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -a_{p-2} \\ & & 1 & -a_{p-1} \end{pmatrix} \in \mathcal{M}_p(\mathbf{k}).$$

8.1.6 Proposition. *On a $\chi_{\mathrm{Comp}(P)} = \mu_{\mathrm{Comp}(P)} = P$.*

Preuve. L'affirmation sur le polynôme minimal a été démontrée ci-dessus, voir (8.1.4.1). Pour démontrer l'autre identité, nous allons calculer explicitement le polynôme caractéristique de la matrice $\mathrm{Comp}(P)$. On développe $\det(X.\mathrm{Id}_p - \mathrm{Comp}(P))$ par rapport à la

dernière colonne :

$$\chi_{C_x} = \sum_{k=1}^{p-1}(-1)^{k+p}a_{k-1}\cdot\begin{vmatrix} X & & & & & & \\ -1 & X & & & & & \\ & \ddots & \ddots & & & & \\ & & -1 & X & 0 & & \\ & & & -1 & X & & \\ & & & & \ddots & \ddots & \\ & & & & & -1 & X \\ & & & & & & -1 \end{vmatrix} + (X+a_{p-1})\cdot\begin{vmatrix} X & & \\ -1 & X & \\ & \ddots & \ddots \\ & & -1 \quad X \end{vmatrix}$$

$$= \sum_{k=1}^{p-1}(-1)^{k+p}a_{k-1}\cdot(-1)^{p-k}X^{k-1} + (X+a_{p-1})\cdot X^{p-1}$$

$$= \sum_{k=0}^{p-2}a_kX^k + a_{p-1}X^{p-1} + X^p$$

comme il fallait. □

Le développement par rapport à la dernière colonne effectué ci-dessus peut éventuellement nécessiter plusieurs tentatives avant d'être réussi. Il est un peu moins acrobatique de faire une récurrence sur p en développant par rapport à la première ligne. Nous recommandons plutôt cette seconde méthode en conditions de stress.

8.2 – Sous-espaces caractéristiques

Avant de définir les sous-espaces caractéristiques, nous allons démontrer deux énoncés reliant les polynômes caractéristiques et scindés.

Le premier est le fameux théorème de Cayley–Hamilton. La preuve donnée ici est directe, en ce qu'elle ne passe pas par un argument d'extension des scalaires pour se ramener au cas trigonalisable. Il me semble que cette preuve due à Frobenius est la première qui ait été donnée du théorème de Cayley–Hamilton en toute généralité, voir section 7.3.

Le second résultat a lui aussi été démontré précédemment avec l'hypothèse supplémentaire que f est trigonalisable, c'est le Lemme 7.2.10.

8.2.1 Proposition. *Soit $f \in \mathcal{L}(E)$.*
(i) *Le polynôme minimal μ_f divise le polynôme caractéristique χ_f.*
(ii) *Les polynômes μ_f et χ_f ont les mêmes facteurs irréductibles.*

Preuve. Pour montrer (i) il suffit de démontrer que χ_f annule f, ou de manière équivalente que $\chi_f(f)(x) = 0$ pour tout $x \in E$. Soit $x \in E$. On considère le sous-espace cyclique F_x associé à x. Puisqu'il est stable par f, $\chi_{f_{F_x}}$ divise χ_f (Proposition 7.1.3). D'autre part, il résulte des résultats de la section 8.1 que $\chi_{f_{F_x}} = \mu_x$, le polynôme minimal de f en x. On a donc $\chi_{f_{F_x}}(f)(x) = 0$, et par suite $\chi_f(f)(x) = 0$ comme on voulait.

Pour démontrer (ii), maintenant qu'on sait que μ_f divise χ_f, il suffit de démontrer que tout facteur irréductible de χ_f est nécessairement facteur irréductible de μ_f. Faisons-le par récurrence sur la dimension de E. Si $\dim(E) = 1$ le résultat est vrai car μ_f et χ_f sont tous les deux de degré 1. Si $\dim(E) \geqslant 2$, considérons $x \in E$ non-nul et F_x le sous-espace cyclique associé. On a $\chi_f = \chi_{f_{F_x}}\chi_{\bar{f}_{F_x}}$, et $\mu_{f_{F_x}}$ et $\mu_{\bar{f}_{F_x}}$ divisent tous les deux μ_f (Propositions 7.1.3

et 7.2.12). Soit P polynôme irréductible divisant χ_f. Puisqu'il est irréductible il divise $\chi_{f_{F_x}}$ ou $\chi_{\bar{f}_{F_x}}$. Si P divise $\chi_{f_{F_x}}$, alors puisque $\chi_{f_{F_x}} = \mu_{f_{F_x}}$ (voir section 8.1), P divise $\mu_{f_{F_x}}$, et donc μ_f comme il fallait démontrer. Si P divise $\chi_{\bar{f}_{F_x}}$, il divise $\mu_{\bar{f}_{F_x}}$ par hypothèse de récurrence (on a choisi $x \neq 0$, donc $F_x \supsetneq \{0\}$, et ainsi $\dim(E/F_x) < \dim(E)$), et donc aussi μ_f. □

8.2.2 Sous-espaces caractéristiques. On appelle $\mathrm{Irr} \subseteq \mathbf{k}[x]$ l'ensemble des facteurs irréductibles de χ et/ou μ. On note

$$\mu = \prod\nolimits_{P\in\mathrm{Irr}} P^{a_P} \quad \text{et} \quad \chi = \prod\nolimits_{P\in\mathrm{Irr}} P^{b_P}.$$

Les *sous-espaces caractéristiques* de f sont les $C_P = \ker(P^{a_P}(f))$ pour $P \in \mathrm{Irr}$. D'après le lemme des noyaux, on a la décomposition

$$E = \bigoplus\nolimits_{P\in\mathrm{Irr}} C_P \tag{8.2.2.1}$$

en somme de sous-espaces stables par f, et par tout endomorphisme commutant à f. Les *projecteurs spectraux* sont les projecteurs relativement à cette décomposition, et ce sont des polynômes en f que l'on sait calculer explicitement.

8.2.3 Proposition. *Soit P facteur irréductible de χ et/ou μ. On utilise les notations introduites ci-dessus.*
(i) *L'entier a_P est caractérisé par le fait que P^{a_P} est le polynôme minimal de l'endomorphisme induit f_{C_P} sur le sous-espace caractéristique C_P.*
(ii) *L'entier b_P est* $\dim(C_P)/\deg(P)$.

Preuve. On considère pour chaque facteur irréductible P de μ l'endomorphisme $f_{C_P} \in \mathcal{L}(C_P)$ induit sur le sous-espace caractéristique associé à P. Par définition de C_P, le polynôme P^{a_P} annule f_{C_P}. Puisque P est irréductible, on en déduit qu'il existe un entier naturel $a'_P \leqslant a_P$ tel que $\mu_{f_{C_P}} = P^{a'_P}$.

Puisque la décomposition en sous-espaces caractéristiques est une décomposition en sous-espaces stables, on a par la proposition 7.2.12

$$\begin{aligned}\mu_f &= \mathrm{ppcm}(\mu_{f_{C_P}}, P \in \mathrm{Irr}) \\ &= \prod\nolimits_{P\in\mathrm{Irr}} P^{a'_P}.\end{aligned}$$

Par unicité de la décomposition en produit de facteurs irréductibles (c'est la factorialité de $\mathbf{k}[X]$), on a donc $a'_P = a_P$ pour tout $P \in \mathrm{Irr}$, ce qui prouve (i).

Pour chaque $P \in \mathrm{Irr}$ on a $\mu_{f_{C_P}} = P^{a'_P}$, donc il existe $b'_P \in \mathbf{N}$ tel que $\chi_{f_{C_P}} = P^{b'_P}$ par la proposition 8.2.1, partie (ii). En regardant le degré, on voit que $b'_P = \dim(C_P)/\deg(C_P)$. Enfin, à nouveau grâce à la décomposition en sous-espaces caractéristiques, on a $\chi_f = \prod_{P\in\mathrm{Irr}} \chi_{f_{C_P}} = \prod_{P\in\mathrm{Irr}} P^{b'_P}$, donc à nouveau par unicité de la décomposition en produit de facteurs irréductibles, on a $b'_P = b_P$, ce qui prouve (ii). □

8.3 – Simplicité et semi-simplicité

Nous revenons ici sur les notions de simplicité et semi-simplicité (et b-semi-simplicité) introduites et étudiées en section 6.1. Nous allons notamment donner des caractérisations de la simplicité et semi-simplicité en termes des polynômes caractéristique et minimal.

Pour mémoire, nous avons convenu en section 6.1 qu'un endomorphisme est semi-simple s'il est somme directe d'endomorphismes simples, et semi-simple-bis s'il possède la propriété que tout sous-espace stable possède un supplémentaire stable. Nous avons montré que la semi-simplicité et la b-semi-simplicité sont deux propriétés équivalentes, nous allons le redémontrer ici.

8.3.1 Proposition. *Soit $f \in \mathcal{L}(E)$. Les propositions suivantes sont équivalentes :*
(i) *f simple ;*
(ii) *le polynôme caractérisique χ_f irréductible.*

Preuve. Supposons f simple, et montre que ceci impose à χ_f d'être irréductible. Soit $S, T \in \mathbf{k}[X]$ tels que $\chi_f = ST$. On a $S(f)T(f) = 0$, donc $S(f)$ ou $T(f)$ a un noyau non-nul. Disons que c'est S, et soit $x \in \ker(S(f))$ non nul. Par simplicité de f, F_x le plus petit sous-espace stable contenant x est E tout entier, donc $\deg(\mu_x) = n$ (on utilise les notations de la section 8.1). Ceci implique $\deg S \geqslant n$, et donc $T \in \mathbf{k}$. Ceci prouve que χ est irréductible.

Si au contraire f possède un sous-espace stable non-trivial F, alors χ_{f_F} est diviseur strict de χ_f, puisque $\deg(\chi_{f_F}) < \deg(\chi_f)$, donc χ n'est pas irréductible. □

8.3.2. En particulier, la proposition 8.3.1 nous dit que si $\mathbf{k}$ est algébriquement clos, alors les seuls endomorphismes simples sont les homothéties d'une droite. (En effet, si $\mathbf{k}$ est algébriquement clos les seuls polynômes irréductibles sont les polynômes de degré 1). Dans ce cas, les endomorphismes semi-simples, qui par définition sont sommes directes d'endomorphismes simples, sont nécessairement diagonalisables.

Remarquons au passage que Le caractère simple dépend du choix du corps de base. Pour l'illustrer, considérons l'endomorphisme associé à la matrice

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

($\theta \in \mathbf{R}$ non congru à 0 modulo π). L'endomorphisme de $\mathbf{R}^2$ associé à cette matrice est simple (voir exemple 6.1.7, mais l'endomorphisme de $\mathbf{C}^2$ associé à la même matrice n'est pas simple (il a deux sous-espaces propres de dimension 1).

8.3.3 Proposition. *Soit $f \in \mathcal{L}(E)$. Les propositions suivantes sont équivalentes :*
(i) *f est semi-simple-bis ;*
(ii) *aucun carré non constant ne divise le polynôme minimal μ_f.*

Preuve. Montrons "(i) ⇒ (ii)". Soit T facteur irréductible de μ_f. Il s'agit de montrer que T^2 ne divise pas μ_f. Pour cela, considérons $C_T^0 := \ker(T(f))$. [1] C'est un sous-espace stable par f. Si f est semi-simple-bis, C_T^0 possède un supplémentaire stable F. L'endomorphisme $T(f)_F \in \mathcal{L}(F)$ est injectif puisque

$$\ker\big(T(f)_F\big) = \ker\big(T(f)\big) \cap F = \{0\} \ ;$$

il est donc inversible pour raisons de dimension. On en déduit que T et μ_{f_F} sont premiers entre eux, et donc T ne divise pas μ_{f_F}. Puisque $T \cdot \mu_{f_F}$ annule f, μ_f divise $T \cdot \mu_{f_F}$, et donc T^2 ne divise pas μ_f, comme il fallait démontrer.

1. attention, en général ce n'est pas le sous-espace caractéristique associé à T ; justement ça l'est si et seulement si T^2 ne divise pas μ.

Montrons que réciproquement, "(ii) $\Rightarrow$ (i)". Écrivons $\mu_f = T_1 \cdots T_r$, où les T_i sont irréductibles deux à deux non proportionnels, et considérons la décomposition en sous-espaces caractéristiques $E = C_1 \oplus \cdots \oplus C_r$ et les projecteurs spectraux $p_1, \ldots, p_r$ (pour tout $i = 1, \ldots, r$, on note $C_i = \ker(T_i(f))$). Soit F un sous-espace stable par f. Puisque les p_i sont des polynômes en f d'après le lemme des noyaux, chacun laisse F stable. On en déduit la décomposition

$$F = (F \cap C_1) \oplus \cdots \oplus (F \cap C_r). \tag{8.3.3.1}$$

On va montrer que chaque $F \cap C_i$ possède un supplémentaire G_i dans C_i stable par f, ou de manière équivalente par f_{C_i}. Alors $G := \bigoplus G_i$ sera un supplémentaire de F dans E stable par f, et on aura gagné.

Pour montrer l'existence d'un supplémentaire stable à $F \cap C_i$ dans C_i, on utilise un argument un peu baroque. Soit $\mathbf{k}_i := \mathbf{k}[X]/(T_i)$; puisque T_i est irréductible, $\mathbf{k}_i$ est un corps. L'opération de composition externe

$$(\bar{A}, x) \in \mathbf{k}_i \times C_i \mapsto \bar{A}.x := A(f)(x)$$

munit le $\mathbf{k}$-espace vectoriel C_i d'une structure de $\mathbf{k}_i$-espace vectoriel. On observe que les $\mathbf{k}_i$-sous-espaces vectoriels de C_i correspondent ensemblistement aux $\mathbf{k}$-sous-espaces vectoriels de C_i qui sont stables par f. Ainsi $F \cap C_i$ est un $\mathbf{k}_i$-sev de C_i. On lui choisit un $\mathbf{k}_i$-supplémentaire : c'est un $\mathbf{k}$-sev de C_i supplémentaire à $F \cap C_i$ qui est stable par f, et nous avons donc trouvé notre G_i. □

8.3.4 Remarque. Il est toujours vrai, sans condition sur f, que tout sous-espace stable F se décompose selon les sous-espaces caractéristiques comme en (8.3.3.1).

On peut le voir en appliquant le même argument que dans la preuve ci-dessus, ou bien de la manière suivante. On considère $\mu_f = \prod P_i^{a_i}$ la décomposition du polynôme minimal en produit de facteurs irréductibles. Le polynôme μ_f est annulateur aussi pour l'endomorphisme induit $f_F \in \mathcal{L}(F)$, et donc on obtient par le lemme des noyaux

$$F = \bigoplus_i \ker(P_i^{a_i}(f_F)),$$

qui est exactement la décomposition cherchée puisque $\ker(P_i^{a_i}(f_F)) = F \cap \ker(P_i^{a_i}(f))$.

En revanche, il est grossièrement faux que pour une décomposition arbitraire $E = \bigoplus H_i$ on a $F = \bigoplus(F \cap H_i)$. Nous l'avons déjà observé en 7.6.5.1.

8.3.5 Corollaire. *Soit $f \in \mathcal{L}(E)$ semi-simple-bis. Alors pour tout sous-espace F stable par f, l'endomorphisme induit $f_F \in \mathcal{L}(F)$ est semi-simple-bis.*

Preuve. Puisque f est semi-simple, μ_f est sans facteur carré. Comme μ_{f_F} divise μ_f, il est lui aussi sans facteur carré, et donc f_F est lui-même semi-simple-bis. □

8.3.6 Proposition. *Soit $f \in \mathcal{L}(E)$. Les propositions suivantes sont équivalentes :*
(i) *f est semi-simple ;*
(ii) *f est semi-simple-bis.*

Preuve. Supposons pour commencer que f est semi-simple. Alors par définition il existe une décomposition $E = \bigoplus_{i=1}^r F_i$ en sous-espaces stables telle que pour tout $i = 1, \ldots, r$, l'endomorphisme induit $f_{F_i} \in \mathcal{L}(F_i)$ est simple. D'après la proposition 8.3.1, le polynôme

caractéristique $\chi_{f_{F_i}}$ est un polynôme irréductible P_i. On a donc $\mu_{f_{F_i}} = P_i$ d'après le théorème de Cayley–Hamilton. Pour conclure,

$$\mu_f = \mathrm{ppcm}(\mu_{f_{F_1}}, \ldots, \mu_{f_{F_r}}) = \mathrm{ppcm}(P_1, \ldots, P_r)$$

est sans facteur carré puisque $P_1, \ldots, P_r$ sont irréductibles. D'après la proposition 8.3.3, ceci prouve que f est semi-simple-bis comme il fallait démontrer.

Réciproquement, nous allons démontrer par récurrence sur $n = \dim(E) \geqslant 0$ que tout endomorphisme de E semi-simple-bis est semi-simple. Si $n = 0$, le résultat est trivial. Supposons donc $n \geqslant 1$ et le résultat démontré pour tout $n' < n$. Soit $f \in \mathcal{L}(E)$ semi-simple-bis. Si f est simple, il n'y a rien à démontrer. Sinon il existe F sous-espace stable par f tel que

$$0 < \dim(F) < \dim(E). \tag{8.3.6.1}$$

Puisque f est semi-simple-bis, il existe un supplémentaire F' de F stable par f. Les inégalités (8.3.6.1) entraînent des inégalités identiques pour $\dim(F')$, et on peut donc appliquer l'hypothèse de récurrence à $f_F \in \mathcal{L}(F)$ et $f_{F'} \in \mathcal{L}(F')$ qui sont tous les deux semi-simples-bis d'après le corollaire 8.3.5. On en déduit qu'il existe des décompositions $F = \bigoplus_{i=1}^{r} F_i$ et $F' = \bigoplus_{i=1}^{s} F'_i$ en sous-espaces stables par f_F et $f_{F'}$ repectivement, telles que les f_{F_i} et $f_{F'_i}$ sont simples. Finalement,

$$E = (\bigoplus\nolimits_{i=1}^{r} F_i) \oplus (\bigoplus\nolimits_{i=1}^{s} F'_i)$$

est une décomposition de E en somme directe de sous-espaces stables par f tels que les endomorphismes induits sur chaque terme de la somme sont tous simples. Ceci prouve bien que f est semi-simple. □

8.3.7. Les endomorphismes qui ne sont archétypiquement pas semi-simples sont les nilpotents (non nuls). Pour ceux-là, $X^2 | \mu_f$ (sauf si $f = 0$), et le noyau est un sous-espace stable sans supplémentaire stable (voir la preuve de la proposition 7.7.12.

A contrario l'archétype de l'endomorphisme semi-simple est l'endomorphisme diagonalisable. D'ailleurs, la preuve donnée ci-dessus du fait que si μ est sans facteur carré alors f est semi-simple est parallèle à celle du fait que les diagonalisables sont semi-simples ; la différence est que dans le cas diagonalisable, on a $\mathbf{k}_i \cong \mathbf{k}$ et donc l'argument baroque est invisible (mais bien là tapis dans l'ombre).

Nous avons évoqué à plusieurs reprises le fait que les endomorphismes semi-simples sont essentiellement ceux qu'il est possible de diagonaliser quitte à étendre les scalaires. Pour que ce soit tout-à-fait vrai, il faut faire une hypothèse sur le corps de base $\mathbf{k}$.

8.3.8 Rappel de théorie des corps. Un corps $\mathbf{k}$ est *parfait* si tout polynôme irréductible de $\mathbf{k}[X]$ est scindé à racines simples dans une extension de décomposition.

Pour illustrer la notion, le mieux est sans doute de donner un exemple de corps qui n'est pas parfait (le lecteur doute peut-être du fait qu'une telle horreur puisse exister). Soit $\mathbf{k}$ un corps de caractéristique $p > 0$ tel qu'il existe $\alpha \in \mathbf{k}$ qui n'est pas une puissance p-ième : on écrit alors $\alpha \notin \mathbf{k}^p$.[2] Nous allons voir que le polynôme $X^p - \alpha \in \mathbf{k}[X]$ est irréductible, mais que c'est une puissance p-ième dans toute extension de décomposition. Considérons une extension $\mathbf{k}'$ de $\mathbf{k}$ dans laquelle α possède une racine p-ième $\beta \in \mathbf{k}'$. Alors

$$X^p - \alpha = X^p - \beta^p = (X - \beta)^p$$

2. attention à ne pas confondre avec le produit cartésien p-fois de $\mathbf{k}$, que nous notons de la même façon...

dans $\mathbf{k}'[X]$.

Reste à voir que $X^p - \alpha$ est irréductible dans $\mathbf{k}[X]$. Dans $\mathbf{k}'[X]$, $X^p - \alpha$ a un unique facteur irréductible, $X - \beta$. Soit $P \in \mathbf{k}[X]$ facteur irréductible de $X^p - \alpha$. Il existe un entier $q \in [\![1, p]\!]$ tel que, dans $\mathbf{k}'[X]$,

$$P = (X - \beta)^q = X^q - q\beta X^{q-1} + \cdots .$$

Puisque $\beta \notin \mathbf{k}$ et $P \in \mathbf{k}[X]$, on a nécessairement $q\beta = 0$, et donc $q = p$. Ainsi l'unique facteur irréductible de $X^p - \alpha$ dans $\mathbf{k}[X]$ est $X^p - \alpha$ lui-même, comme il fallait démontrer.

Un exemple explicite réalisant la situation ci-dessus est le suivant. On prend $\mathbf{k} = \mathbf{F}_p(T)$, le corps des fractions rationnelles à coefficients dans $\mathbf{F}_p = \mathbf{Z}/p\mathbf{Z}$, p premier, et $\alpha = T$.

On retiendra (nous ne le démontrerons pas ici) que les corps parfaits sont exactement les corps de caractéristique 0, et les corps de caractéristique $p > 0$ tels que $\mathbf{k}^p = \mathbf{k}$. Cette dernière catégorie inclut tous les corps finis.

8.3.9 Proposition. *Soit E un $\mathbf{k}$-espace vectoriel de dimension finie, et $f \in \mathcal{L}(E)$.*
(i) Si $\mathbf{k}$ est algébriquement clos, alors f semi-simple $\Leftrightarrow$ f diagonalisable.
(ii) Si $\mathbf{k}$ est parfait, alors $M \in \mathcal{M}_n(\mathbf{k})$ est semi-simple si et seulement si elle est diagonalisable dans une extension finie de $\mathbf{k}$.

Preuve. D'après les propositions 8.3.3 et 8.3.6, f est semi-simple si et seulement si son polynôme minimal μ_f est sans facteur constant. Si $\mathbf{k}$ est algébriquement clos, les irréductibles de $\mathbf{k}[X]$ sont les polynômes de degré 1, et cette condition équivaut à ce que μ_f soit scindé à racines simples, et donc à ce que f soit diagonalisable. Ceci prouve (i).

Montrons (ii). S'il existe $\mathbf{k}'$ extension de $\mathbf{k}$ telle que le polynôme minimal de M est scindé à racines simples sur $\mathbf{k}'$, alors μ_M est sans facteur carré dans $\mathbf{k}'[X]$ et donc *a fortiori* dans $\mathbf{k}[X]$. On en déduit que M est semi-simple.

Réciproquement, si $M \in \mathcal{M}_n(\mathbf{k})$ est semi-simple, alors il existe $P_1, \ldots, P_r \in \mathbf{k}[X]$ irréductibles deux à deux non proportionnels tels que $\mu_M = P_1 \cdots P_r$. Considérons une extension $\mathbf{k}'$ de $\mathbf{k}$ dans laquelle $P_1, \ldots, P_r$ sont tous décomposés. Puisque $\mathbf{k}$ est un corps parfait, $P_1, \ldots, P_r$ sont tous scindés à racines simples sur $\mathbf{k}'$. Comme de plus ils sont deux à deux premiers entre eux, le produit $P_1 \cdots P_r$ est lui-même scindé à racines simples sur $\mathbf{k}'$. On en déduit que M est diagonalisable dans $\mathcal{M}_n(\mathbf{k}')$, comme il fallait démontrer. □

8.3.10 Corollaire. *Si M est une matrice définie sur un corps parfait, alors le caractère semi-simple ou non est invariant par extension des scalaires.*

Preuve. D'après la proposition précédente, la matrice M est semi-simple si et seulement si elle est diagonalisable sur une clôture algébrique de $\mathbf{k}$. Cette condition ne change pas quand on passe à une extension $\mathbf{k}'$ de $\mathbf{k}$. □

8.3.11 Exemple. Si en revanche $\mathbf{k}$ n'est pas parfait, l'énoncé du corollaire ci-desssus est faux en général. Considérons par exemple la matrice compagnon

$$\mathrm{Comp}(X^p - T) \in \mathcal{M}_p(\mathbf{F}_p(T))$$

pour p premier. Son polynôme caractéristique est $X^p - T$ qui est irréductible (voir 8.3.8 ci-dessus), donc cette matrice est simple, et *a fortiori* semi-simple. En revanche, son polynôme minimal est $(X - S)^p$ dans l'extension $\mathbf{k}' = \mathbf{k}[S]/(S^p - T)$ de $\mathbf{k}$, et sur $\mathbf{k}'$ cette matrice n'est pas semi-simple, et encore moins diagonalisable.

8.4 – Structure de l'algèbre $\mathbf{k}[f]$

8.4.1 L'algèbre $\mathbf{k}[f]$. On a déjà vu en 7.2.5 qu'elle est isomorphe à $\mathbf{k}[X]/(\mu_f)$, donc de dimension finie $\deg \mu_f$ sur $\mathbf{k}$.

Parmi ses « propriétés globales », outre sa dimension il faut citer son commutant dans $\mathcal{L}(E)$, i.e. l'ensemble des endomorphismes de E qui commutent avec f. Ceci s'identifie naturellement à l'espace des endomorphismes de E comme $\mathbf{k}[X]$-module :

$$\mathrm{Com}_{\mathcal{L}(E)}(f) \cong \mathrm{End}_{\mathbf{k}[X]}(M_{E,f})$$

(voir par exemple [H2G2II, II.2]).

8.4.2. La première chose à dire c'est que, notant $\mu_f = \prod T_i^{a_i}$ la décomposition en produit de facteurs irréductible, on a par le lemme chinois

$$\mathbf{k}[f] \cong \bigoplus_i \mathbf{k}[X]/(T_i^{a_i}), \tag{8.4.2.1}$$

qui présente un parallèle évident avec la décomposition en sous-espaces caractéristiques.

Sans expliciter beaucoup plus cette décomposition, on peut déjà prouver les deux énoncés suivants.

8.4.3 Semi-simplicité et nilpotents. *L'endomorphisme f est semi-simple ssi l'algèbre $\mathbf{k}[f]$ n'a pas d'élément nilpotent non-nul.*

Preuve. Le point clef est que semi-simple $\Leftrightarrow$ aucun carré non constant ne divise μ. Ainsi, vu la décomposition (8.4.2.1), il s'agit de se convaincre que $\mathbf{k}[X]/(T^a)$, T irréductible, possède des nilpotents non-triviaux ssi $a > 1$.

Si $a > 1$, $\bar{T}$ est un nilpotent non-nul. Si $a = 1$, soit $\bar{F}$ un nilpotent : il existe $s \geqslant 1$ tel que $\bar{F}^s = 0$, i.e. $T|F^s$. Comme T irréductible, par le lemme de Gauss ceci implique $T|F$, i.e. $\bar{F} = 0$. □

8.4.4 Idempotents et projecteurs sur les sous-espaces caractéristiques. *Les idempotents de l'algèbre $\mathbf{k}[f]$ sont les (sommes de) projecteurs sur les sous-espaces caractéristiques.*

Autrement dit, pour tout $g \in \mathbf{k}[f]$ tel que $g^2 = g$, il existe $I \subseteq [\![1, r]\!]$ tel que $g = \sum_{i\in I} p_i$, où $p_1, \ldots, p_r$ sont les projecteurs sur les sous-espaces caractéristiques de f.

Preuve. Soit $g \in \mathbf{k}[f]$ idempotent. Puisque $g^2 = g$, g est un projecteur, et puisque g est un polynôme en f, il laisse stable chacun des sous-espaces caractéristiques de f. Nous allons démontrer que pour chaque T facteur irréductible de μ_f, l'endomorphisme induit g_{C_T} est 0 ou id_{C_T}, ce qui démontrera le résultat annoncé.

Ainsi, il suffit de démontrer que pour f de polynôme minimal $\mu_f = T^a$, T irréductible, tout $g \in \mathbf{k}[f]$ idempotent est 0 ou id. Il existe $A \in \mathbf{k}[X]$ tel que $g = A(f)$. Si A est premier à T, alors il est premier à $T^a = \mu_f$, donc $g = A(f)$ est un isomorphisme. Puisque c'est un projecteur, on a nécessairement $g = \mathrm{id}$. Si *a contrario* T divise A, alors μ_f divise A^a et donc $g^a = 0$. Puisque g est idempotent, $g^a = g$, et donc $g = 0$ dans ce cas. □

Il est bon néanmoins d'explorer plus en profondeur les relations entre la décomposition de $\mathbf{k}[f]$ donnée par l'isomorphisme chinois et la décomposition de E donnée par le lemme des noyaux.

8.4.5 Isomorphisme chinois. Soit $P_1, \dots, P_r$ des polynômes deux à deux premiers entre eux. L'application

$$\varphi : (A \bmod P_1 \cdots P_r) \in \frac{\mathbf{k}[X]}{(P_1 \cdots P_r)} \longmapsto (A \bmod P_i)_{1 \leqslant i \leqslant r} \in \prod_{1 \leqslant i \leqslant r} \frac{\mathbf{k}[X]}{(P_i)}$$

est un isomorphisme de $\mathbf{k}$-algèbres.

Posons $Q_i = \prod_{j \neq i} P_j$ pour tout $i = 1, \dots, r$. Les polynômes Q_i sont premiers entre eux dans leur ensemble comme on l'a vu lors de la preuve du lemme des noyaux. Ils sont donc liés par une relation de Bezout : il existe $U_1, \dots, U_r \in \mathbf{k}[X]$ tels que

$$U_1 Q_1 + \cdots + U_r Q_r = 1.$$

L'application

$$\psi : (A_i \bmod P_i)_{1 \leqslant i \leqslant r} \longmapsto U_1 Q_1 A_1 + \cdots + U_r Q_r A_r \bmod P_1 \cdots P_r$$

est le morphisme de $\mathbf{k}$-algèbres réciproque de φ.

Preuve. On laisse au lecteur le soin de vérifier que l'application φ est bien définie, et est un morphisme de $\mathbf{k}$-algèbres. Montrons que ce morphisme est injectif : Soit $\bar{A} \in \ker(\varphi)$. Pour tout $i = 1, \dots, r$, $A \bmod P_i$ est nul, autrement dit A est divisible par P_i. Puisque les P_i sont deux à deux premiers entre eux, A est donc divisible par $P_1, \dots, P_r$, autrement dit $\bar{A} = 0$. Ceci suffit pour conclure que φ est un isomorphisme, puisque les deux algèbres $\mathbf{k}[X]/(P_1 \cdots P_r)$ et $\prod_{1 \leqslant i \leqslant r} (\mathbf{k}[X]/(P_i))$ sont de dimensions égales

$$\deg(P_1 \cdots P_r) = \deg(P_1) + \cdots + \deg(P_r).$$

Le fait que ψ soit le morphisme réciproque de φ prouvera explicitement la surjectivité de φ.

à propos de l'application ψ, commençons par montrer qu'elle est bien définie : soit A_i et B_i deux polynômes congrus modulo P_i. Alors il existe $K_i \in \mathbf{k}[X]$ tel que $B_i - A_i = K_i P_i$, et donc

$$U_i Q_i B_i - U_i Q_i A_i = K_i P_i Q_i = K_i P_1 \cdots P_r,$$

autrement dit $U_i Q_i A_i$ et $U_i Q_i B_i$ sont congrus modulo $P_1 \cdots P_r$ comme on voulait démontrer. Ensuite montrons que ψ est un morphisme de $\mathbf{k}$-algèbres : toutes les vérifications sont élémentaires (donc laissées au lecteur), sauf peut-être celle de la multiplicativité, qui fonctionne comme suit. Soit $A_1, B_1, \dots, A_r, B_r \in \mathbf{k}[X]$. On a [3]

$$\begin{aligned} \psi\big((A_1, \dots, A_r) \cdot (B_1, \dots, B_r)\big) &= \psi(A_1 B_1, \dots, A_r B_r) \\ &= U_1 Q_1 A_1 B_1 + \cdots + U_r Q_r A_r B_r \end{aligned}$$

et

$$\psi(A_1, \dots, A_r) \cdot \psi(B_1, \dots, B_r) = \sum_{1 \leqslant i,j \leqslant r} U_i Q_i A_i \cdot U_j Q_j B_j.$$

Si $i \neq j$, $Q_i Q_j$ est nul modulo $P_1 \cdots P_r$, donc en fait

$$\psi(A_1, \dots, A_r) \cdot \psi(B_1, \dots, B_r) = \sum_{1 \leqslant i \leqslant r} (U_i Q_i)^2 A_i B_i.$$

3. pour alléger les notations, on confond polynômes et classes de congruences, c'est dans notre intérêt à tous

Enfin, pour tout $i = 1, \dots, r$, en multipliant la relation de Bezout par U_iQ_i il vient

$$U_iQ_i = \sum_{1\leqslant j\leqslant n} U_iQ_iU_jQ_j = (U_iQ_i)^2,$$

où la dernière égalité est donnée à nouveau par le fait que Q_iQ_j est divisible par $P_1 \cdots P_r$ si $i \neq j$. Finalement on a donc bien

$$\begin{aligned}\psi(A_1, \dots, A_r) \cdot \psi(B_1, \dots, B_r) &= \sum_{1\leqslant i\leqslant r} (U_iQ_i)^2A_iB_i = \sum_{1\leqslant i\leqslant r} U_iQ_iA_iB_i \\ &= \psi\big((A_1, \dots, A_r) \cdot (B_1, \dots, B_r)\big).\end{aligned}$$

Il reste à voir que les morphismes φ et ψ sont réciproques l'un de l'autre. Pour $A \in \mathbf{k}[X]$ on a

$$\psi \circ \varphi(A) = (U_1Q_1 + \cdots + U_rQ_r)A = A,$$

et pour $A_1, \dots, A_r \in \mathbf{k}[X]$,

$$\begin{aligned}\varphi \circ \psi(A_1, \dots, A_r) &= (U_1Q_1A_1 + \cdots U_rQ_rA_r \bmod P_i)_{1\leqslant i\leqslant r} \\ &= (U_iQ_iA_i \bmod P_i)_{1\leqslant i\leqslant r} \\ &= (A_i \bmod P_i)_{1\leqslant i\leqslant r},\end{aligned}$$

où la dernière égalité est donnée par le fait qu'en réduisant l'identité de Bezout modulo P_i, il vient $\overline{U_iQ_i} = \overline{1}$. □

8.4.6 Projecteurs spectraux. On constate que les projecteurs spectraux sont les éléments de $\mathbf{k}[f]$ qui dans la décomposition (8.4.2.1) s'écrivent $(0, \dots, 1, \dots, 0)$. Pour cela, on relit l'expression des projecteurs spectraux comme polynômes en f fournie par la preuve du lemme des noyaux 7.5.1, puis celle du morphisme réciproque ψ ci-dessus.

On en déduit une autre preuve de la caractéristation des idempotents de $\mathbf{k}[f]$.

Seconde preuve de 8.4.4. Il s'agit de montrer que pour $T \in \mathbf{k}[X]$ irréductible,les idempotents de $\mathbf{k}[X]/(T^a)$ sont 0 et 1. $\bar{A}$ est idempotent ssi

$$\begin{aligned}A^2 \equiv A \bmod T^a &\Leftrightarrow T^a | A(A-1) \\ &\Leftrightarrow T^a | A \text{ ou } T^a | (A-1)\end{aligned}$$

car T irréductible et A et $A-1$ premiers entre eux. □

8.5 – Endomorphismes cycliques

8.5.1 Définition. *On dit qu'un endomorphisme $f \in \mathcal{L}(E)$ est* cyclique *s'il existe un vecteur $x \in E$ tel que le plus petit sous-espace stable par f contenant x est E tout entier.*

On rappelle (voir section 8.1) que le plus petit sous-espace stable par f contenant x s'appelle le sous-espace cyclique associé à x, et nous le notons F_x.

D'après les résultats de la section 8.1, f est cyclique si et seulement si il existe une base de E dans laquelle f est donné par une matrice compagnon ; cette matrice compagnon est nécessairement associée au polynôme caractéristique χ_f.

Nous allons donner une condition nécessaire et suffisante pour qu'un endomorphisme soit cyclique, mais avant cela examinons à la main le cas des endomorphismes nilpotents et diagonalisables.

8.5.2 Exemple. Soit $f \in \mathcal{L}(E)$ nilpotent. Pour tout vecteur $x \in E$, soit i le plus petit entier tel que

$$x \in K_i(f) = \ker(f^i).$$

Alors $\mu_x = X^i$, et le sous-espace cyclique F_x a dimension i. On en déduit que f est cyclique si et seulement si son indice de nilpotence est égal à la dimension de E.

Le fait que $\mu_x = X^i$ résulte des arguments donnés dans la preuve du Théorème 7.7.6 : par minimalité de i, $\overline{x} \in K_i/K_{i-1}$ est non-nulle, donc $\overline{f_i(\overline{x})} = \overline{f(x)} \in K_{i-1}/K_{i-2}$ est non-nulle si $i > 1$. On montre ainsi par récurrence que $\overline{x}, \overline{f(x)}, \ldots, \overline{f^{i-1}(x)}$ sont tous non-nuls dans $K_i/K_{i-1}, K_{i-1}/K_{i-2}, \ldots, K_1/K_0$ respectivement. On en déduit par le lemme 5.4.4 que la famille $x, f(x), \ldots, f^{i-1}(x)$ est libre, et donc $\deg \mu_x \geqslant i$. Puisque $f^i(x) = 0$, on en déduit $\mu_x = X^i$ comme on voulait.

8.5.3 Exemple. Soit $f \in \mathcal{L}(E)$ diagonalisable. On note $\lambda_1, \ldots, \lambda_r$ ses valeurs propres deux à deux distinctes, et $E_1, \ldots, E_r$ les sous-espaces propres correspondants. Pour tout vecteur $x \in E$, on a la décomposition

$$x = x_1 + \cdots + x_r$$

selon la décomposition $E = \bigoplus_{i=1}^r E_i$. Alors $F_x = \mathrm{Vect}(x_1, \ldots, x_r)$ a dimension au plus r (égale au nombre de x_i non-nuls). On en déduit que f est cyclique si et seulement si tous ses sous-espaces propres sont de dimension 1, autrement dit si son polynôme caractéristique est scindé à racines simples.

Pour montrer que $F_x = \mathrm{Vect}(x_1, \ldots, x_r)$, nous proposons deux méthodes. Dans les deux cas, on commence par remarquer que $\mathrm{Vect}(x_1, \ldots, x_r)$ est un sous-espace stable contenant x, donc on a l'inclusion $F_x \subseteq \mathrm{Vect}(x_1, \ldots, x_r)$. La première méthode pour montrer l'inclusion inverse est de se souvenir que d'après la Proposition 7.6.5, tout sous-espace F stable par f est de la forme

$$F = F_1 \oplus \cdots \oplus F_r,$$

où chaque F_i est un sous-espace arbitraire de l'espace propre E_i. On en déduit que tout sous-espace stable par f contenant $\mathrm{Vect}(x_1, \ldots, x_r)$ est de la forme $\mathrm{Vect}(x_{i_1}, \ldots, x_{i_s})$, avec $1 \leqslant i_1 < \cdots < i_s \leqslant r$. Si un tel sous-espace est contenu strictement dans $\mathrm{Vect}(x_1, \ldots, x_r)$, c'est qu'il existe $i_0 \in [\![1, r]\!] \setminus \{i_1, \ldots, i_s\}$ tel que $x_{i_0} \neq 0$, et dans ce cas $\mathrm{Vect}(x_{i_1}, \ldots, x_{i_s}) \subseteq \bigoplus_{i \neq i_0} E_i$ ne peut pas contenir x, puisque $x \notin \bigoplus_{i \neq i_0} E_i$ car $x_{i_0} \neq 0$. On en déduit que le plus petit sous-espace stable par f contenant x est $\mathrm{Vect}(x_1, \ldots, x_r)$ comme il fallait démontrer.

L'autre façon de faire consiste à démontrer par un calcul explicite que

$$\mathrm{Vect}(x, f(x), \ldots, f^{r-1}(x)) = \mathrm{Vect}(x_1, \ldots, x_r),$$

ce qui implique $\mathrm{Vect}(x_1, \ldots, x_r) \subseteq F_x$ et permet donc de conclure. Le calcul est le suivant :

$$\begin{aligned} x &= x_1 + \cdots + x_r \\ f(x) &= \lambda_1 x_1 + \cdots + \lambda_r x_r \\ &\vdots \\ f^{r-1}(x) &= \lambda_1^{r-1} x_1 + \cdots + \lambda_r^{r-1} x_r. \end{aligned}$$

Il implique aussitôt l'inclusion $\mathrm{Vect}(x, f(x), \ldots, f^{r-1}(x)) \supseteq \mathrm{Vect}(x_1, \ldots, x_r)$, et l'inclusion inverse vient du fait qu'on peut inverser le système ci-dessus, la matrice de Vandermonde

$V(\lambda_1,\dots,\lambda_r)$ étant inversible dans la mesure où les valeurs propres $\lambda_1,\dots,\lambda_r$ sont deux à deux disjointes.

8.5.4 Proposition. *Pour tout $f\in\mathcal{L}(E)$, il existe un vecteur $x\in E$ tel que $\mu_x=\mu_f$.*

Preuve. On écrit la décomposition en produit de facteurs premiers $\mu_f=P_1^{a_1}\cdots P_r^{a_r}$ du polynôme minimal, et on regarde la décomposition en sous-espaces caractéristiques

$$E=C_1\oplus\cdots\oplus C_r.$$

Pour chaque $i=1,\dots,r$, il existe un $x_i\in C_i$ tel que $\mu_{x_i}=P_i^{a_i}$. En effet, pour tout $x\in C_i$, $\mu_x|P_i^{a_i}$ car $P_i^{a_i}$ annule l'endomorphisme $f_{C_i}\in\mathcal{L}(C_i)$ induit par f, donc il existe $a(x)$ tel que $\mu_x=P_i^{a(x)}$. Si pour tout $x\in C_i$ on a $a(x)<a_i$, alors $P_i^{a_i-1}(f_{C_i})=0$, et ceci contredit le fait que $P_i^{a_i}$ est le polynôme minimal de f_{C_i} (voir Proposition 8.2.3).

Ensuite $x=x_1+\cdots+x_r$ convient car pour tout $Q\in\mathbf{k}[X]$:

$$Q(f)(x)=Q(f)(x_1)+\cdots+Q(f)(x_r)=0\Leftrightarrow Q(f)(x_1)=\cdots=Q(f)(x_r)=0,$$

puisque $Q(f)(x_i)\in C_i$ pour tout i, et les sous-espaces caractéristiques sont en somme directe. On a donc

$$\begin{aligned}Q(f)(x)=0 &\iff P_1^{a_1},\dots,P_r^{a_r}\text{ divisent }Q\\ &\iff P_1^{a_1}\cdots P_r^{a_r}|Q,\end{aligned}$$

autrement dit $\mu_x=P_1^{a_1}\cdots P_r^{a_r}=\mu_f$ comme on voulait. □

8.5.5 Proposition. *Soit $f\in\mathcal{L}(E)$. Les deux propositions suivantes sont équivalentes :*
(i) *f cyclique ;*
(ii) $\chi_f=\mu_f$.

Preuve. (i) ⇒ (ii). Cette implication est une traduction du fait qu'une matrice compagnon $\mathrm{Comp}(P)$ a polynômes caractéristique et minimal tous les deux égaux à P, voir le lemme 8.1.6.
(ii) ⇒ (i). D'après la proposition 8.5.4, il existe x tel que $\mu_x=\mu_f$; on a donc $\mu_x=\chi_f$ puisque $\chi_f=\mu_f$. Du coup, $\dim F_x=\deg\mu_x=\deg\chi_f=\dim E$, et donc $E=F_x$. □

On retrouve bien avec ce critère à quelle condition un endomorphisme nilpotent ou diagonalisable est cyclique, voir les exemples 8.5.2 et 8.5.3.

8.5.6. Le fait qu'un endomorphisme à polynôme caractéristique scindé à racines simples (en particulier, diagonalisable) soit cyclique nous dit que les deux matrices ci-dessous sont semblables si et seulement si $\lambda_1,\dots,\lambda_n$ sont deux à deux distincts, ce qui ne semble pas évident *a priori*.

$$\begin{pmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{pmatrix}\quad\text{et}\quad\begin{pmatrix}0 & & & (-1)^{n-1}\sigma_n\\ 1 & \ddots & & \\ & \ddots & 0 & -\sigma_2\\ & & 1 & \sigma_1\end{pmatrix},$$

où $\sigma_1,\dots,\sigma_n$ sont les fonctions symétriques élémentaires en $\lambda_1,\dots,\lambda_n$. Plus généralement, on obtient le résultat suivant en appliquant les Propositions 8.5.4 et/ou 8.5.5.

8.5.7. Soit $Q_1, \dots, Q_r \in \mathbf{k}[X]$. Les deux matrices

$$\mathrm{diag}(\mathrm{Comp}(Q_1), \dots, \mathrm{Comp}(Q_r)) \quad \text{et} \quad \mathrm{Comp}(Q_1 \cdots Q_r)$$

(la première est une matrice diagonale par blocs, dont les blocs diagonaux sont les matrices compagnons $\mathrm{Comp}(Q_1), \dots, \mathrm{Comp}(Q_r)$) sont semblables si et seulement si $Q_1, \dots, Q_r$ sont deux à deux premiers entre eux. Un cas typique où cette condition est vérifiée est si $Q_i = P_i^{a_i}$ pour tout $i = 1, \dots, r$, où $P_1, \dots, P_r$ sont des polynômes irréductibles deux à deux non proportionnels, et $a_1, \dots, a_r \in \mathbf{N}$.

Pour prouver notre affirmation, commençons par observer qu'une matrice est semblable à la matrice compagnon $\mathrm{Comp}(Q_1 \cdots Q_r)$ si et seulement si c'est la matrice d'un endomorphisme cyclique de polynôme minimal $Q_1 \cdots Q_r$. Or la matrice de gauche est la matrice d'un endomorphisme f de $\mathbf{k}^N$, $N = \sum_{i=1}^r \deg(Q_i)$, pour lequel il existe une décomposition en sous-espaces stables $k^N = \bigoplus_{i=1}^r E_i$ telle que pour tout i, l'endomorphisme induit f_{E_i} a polynômes minimal et caractéristique tous les deux égaux à Q_i. On a donc

$$\chi_f = Q_1 \cdots Q_r \quad \text{et} \quad \mu_f = \mathrm{ppcm}(Q_1, \dots, Q_r),$$

et les deux matrices $\mathrm{diag}(\mathrm{Comp}(Q_1), \dots, \mathrm{Comp}(Q_r))$ et $\mathrm{Comp}(Q_1 \cdots Q_r)$ sont semblables si et seulement si

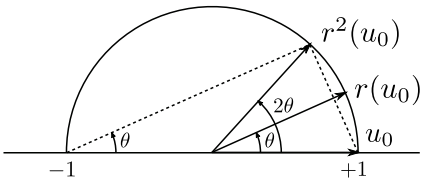
$$\mathrm{ppcm}(Q_1, \dots, Q_r) = Q_1 \cdots Q_r,$$

ce qui équivaut à ce que les polynômes $Q_1, \dots, Q_r$ soient deux à deux premiers entre eux.

8.5.8 Exemple. Illustrons 8.5.6 par un exemple permettant de faire un joli dessin. Pour tout $\theta \in \mathbf{R}$ non congru à 0 modulo π, les deux matrices

$$R_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{et} \quad \begin{pmatrix} 0 & -1 \\ 1 & 2\cos\theta \end{pmatrix}$$

sont semblables. En effet, elles ont toutes les deux $\chi = \mu = X^2 - (2\cos\theta)X + 1$; une autre façon de le dire est que la matrice de gauche vue à coefficients dans $\mathbf{C}$ est diagonalisable de polynôme caractéristique $(X - e^{i\theta})(X - e^{-i\theta}) = -(2\cos\theta)X + 1$, donc redevable de 8.5.6. Géométriquement ces matrices sont les matrices de la rotation d'angle θ dans deux bases différentes, comme le montre le dessin ci-dessous (en toute rigueur, pour parler de rotation d'angle θ il faut au préalable munir le plan $\mathbf{R}^2$ d'une structure euclidienne).



Le triangle en pointillés est rectangle puisqu'il s'appuie sur un diamètre du cercle. L'angle de gauche de ce triangle est θ à cause de la relation entre angle au centre et angle inscrit, et l'hypothénuse a longueur 2. On a donc bien

$$r^2(u_0) = -u_0 + (2\cos\theta).r(u_0).$$

Si en revanche θ est congru à 0 modulo π, la matrice R_θ n'est pas cyclique (c'est une homothétie), et n'est donc pas semblable à une matrice compagnon.

Dans la situation "orthogonale" à celle de 8.5.6 (c'est-à-dire celle où le polynôme caractéristique au lieu d'être à racines simples est une puissance de $X - \lambda$), on a le fait suivant.

8.5.9. Pour tout $\lambda \in \mathbf{k}$, les matrices

$$\begin{pmatrix} \lambda & & & \\ 1 & \ddots & & \\ & \ddots & \lambda & \\ & & 1 & \lambda \end{pmatrix}, \quad \text{et} \quad \mathrm{Comp}\big((X-\lambda)^n\big) = \begin{pmatrix} 0 & & & (-1)^{n-1}\binom{n}{n}\lambda^n \\ 1 & \ddots & & \\ & \ddots & 0 & -\binom{n}{2}\lambda^2 \\ & & 1 & \binom{n}{1}\lambda \end{pmatrix}$$

sont semblables. En effet, la matrice de gauche a polynôme minimal $(X-\lambda)^n$ où n est la taille de la matrice, et pour raisons de degré c'est aussi son polynôme caractéristique. On en déduit que c'est une matrice cyclique, et ainsi qu'elle est semblable à $\mathrm{Comp}\big((X-\lambda)^n\big)$.

8.6 – Décomposition de Frobenius et invariants de similitude

Cette section est consacrée à la preuve et à l'étude des conséquences du résultat suivant.

8.6.1 Théorème (Décomposition de Frobenius)**.** *Pour tout $f \in \mathcal{L}(E)$, il existe une unique suite de polynômes unitaires non constants $P_1, \dots, P_r \in \mathbf{k}[X]$ satisfaisant aux deux conditions :*

(i) *il existe une décomposition $E = \bigoplus_{1\leqslant i\leqslant r} F_i$ en sous-espaces stables par f, telle que pour tout $i = 1, \dots, r$ l'endomorphisme induit $f_{F_i} \in \mathcal{L}(F_i)$ est cyclique de polynôme caractéristique et minimal P_i ; et*

(ii) *$P_1 | P_2 | \cdots | P_r$.*

Nous allons démontrer à part l'existence et l'unicité de la décomposition de Frobenius, mais avant de procéder tâchons d'illustrer l'importance de ce résultat. Les points à retenir sont (i) que la suite de polynômes P_i caractérise la classe de similitude de f, et (ii) que ces polynômes peuvent se calculer algorithmiquement, en appliquant le pivot de Gauss à la matrice $X.\mathrm{Id} - A$ à coefficients dans l'anneau principal $\mathbf{k}[X]$, après avoir choisi une base pour écrire la matrice A de f.

Une décomposition $E = \bigoplus_{1\leqslant i\leqslant r} F_i$ comme dans l'énoncé du théorème est appelée une *décomposition de Frobenius* de f.

8.6.2 Définition. *Soit $f \in \mathcal{L}(E)$. Les polynômes $P_1, \dots, P_r$ associés à f comme dans le Théorème 8.6.1 sont appelés les* invariants de similitude *de f.*

Cette terminologie est justifiée par l'énoncé suivant.

8.6.3 Corollaire. *Deux endomorphismes $f, g \in \mathcal{L}(E)$ sont semblables si et seulement si ils ont les mêmes invariants de similitude.*

Preuve. Notons $(P_1, \dots, P_r)$ et $(Q_1, \dots, Q_s)$ les invariants de similitude de f et g respectivement. Si f et g sont semblables, alors toute décomposition de Frobenius pour f est aussi une décomposition de Frobenius pour g, et donc nécessairement $(P_1, \dots, P_r) = (Q_1, \dots, Q_s)$.

Réciproquement, s'il existe deux décompositions $E = \bigoplus_{1\leqslant i\leqslant r} F_i$ et $E = \bigoplus_{1\leqslant i\leqslant r} G_i$ en sous-espaces F_i stables par f et G_i stables par g telles que f_{F_i} et g_{G_i} sont cycliques de polynôme caractéristique P_i, alors il existe pour tout $i = 1, \dots, r$ un isomorphisme $\varphi_i : G_i \cong F_i$ tel que $g_{G_i} = \varphi_i^{-1} \circ f_{F_i} \circ \varphi_i$, donc f et g sont semblables par le Lemme 6.3.2. □

La partie 'unicité' du Théorème 8.6.1 est une conséquence directe de la Proposition suivante, que nous démontrerons plus loin. Pour la notion de facteurs invariants d'une matrice à coefficients dans un anneau principal, nous renvoyons à la section 3.1.

8.6.4 Proposition. *Soit $f \in \mathcal{L}(E)$, et A la matrice de f dans une base $\mathcal{B}$ de E. Les invariants de similitude de f sont les facteurs invariants de la matrice $X.\mathrm{Id} - A$ à coefficients dans l'anneau principal $\mathbf{k}[X]$ (auxquels il faut retirer les* '1' *initiaux).*

Autrement dit, les facteurs invariants de la matrice $X.\mathrm{Id} - A$ sont $(1, \dots, 1, P_1, \dots, P_r)$ où $(P_1, \dots, P_r)$ sont les invariants de similitude de f.

8.6.5. Une conséquence particulièrement intéressante est qu'on peut déterminer les invariants de similitude en opérant le pivot de Gauss (des matrices à coefficients dans $\mathbf{k}[X]$) à la matrice $X.\mathrm{Id}_n - A$, voir section 3.1.

8.6.6 Exemple. Calculons les invariants de similitude de l'endomorphisme associé à la matrice

$$A = \begin{pmatrix} -3 & -2 & -2 \\ -2 & 0 & -1 \\ 10 & 5 & 6 \end{pmatrix}.$$

On calcule $X.\mathrm{Id} - A$, puis on échange les deux premières lignes, avant d'effectuer une permutation circulaire sur les colonnes pour mettre un 1 en haut à gauche :

$$X.\mathrm{Id} - A = \begin{pmatrix} X+3 & 2 & 2 \\ 2 & X & 1 \\ -10 & -5 & X-6 \end{pmatrix} \xrightarrow[\text{ii) } C_3 \to C_1 \to C_2 \to C_3]{\text{i) } L_1 \leftrightarrow L_2} \begin{pmatrix} 1 & 2 & X \\ 2 & X+3 & 2 \\ X-6 & -10 & -5 \end{pmatrix} ;$$

ensuite on remplace L_2 par $L_2 - 2L_1$ et L_3 par $L_3 - (X-6)L_1$, on obtient

$$\begin{pmatrix} 1 & 2 & X \\ 0 & X-1 & -2(X-1) \\ 0 & -2(X-1) & -(X-1)(X-5) \end{pmatrix} ;$$

enfin on remplace L_3 par $L_3 + 2L_2$, on obtient :

$$\begin{pmatrix} 1 & 2 & X \\ 0 & X-1 & -2(X-1) \\ 0 & 0 & -(X-1)^2 \end{pmatrix}.$$

On voit déjà apparaître les facteurs invariants de la matrice, il n'y a plus qu'à faire du nettoyage en opérant sur les colonnes pour arriver à une forme diagonale : on fait tout d'abord $C_2 \leftarrow C_2 - 2C_1$ et $C_3 \leftarrow C_3 - XC_1$, ce qui donne la première matrice ci-dessous, puis $C_3 \leftarrow -C_3 - 2C_2$, ce qui donne la seconde matrice ci-dessous.

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & X-1 & -2(X-1) \\ 0 & 0 & -(X-1)^2 \end{pmatrix} ; \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & X-1 & 0 \\ 0 & 0 & (X-1)^2 \end{pmatrix}.$$

On en conclut que les invariants de similitude de la matrice A sont $X-1$ et $(X-1)^2$, donc A est semblable à la matrice

$$\begin{pmatrix} \mathrm{Comp}(X-1)) & \\ & \mathrm{Comp}((X-1)^2) \end{pmatrix} = \begin{pmatrix} 1 & & \\ & 0 & -1 \\ & 1 & 2 \end{pmatrix},$$

qui d'après 8.5.9 (voir aussi plus généralement 8.6.10 plus loin), est elle-même semblable à la matrice ci-dessous,

$$\begin{pmatrix} 1 & & \\ & 1 & 0 \\ & 1 & 1 \end{pmatrix}.$$

On laisse au lecteur le soin de vérifier qu'effectivement $\ker(A - \mathrm{Id}))$ et $\ker(A - \mathrm{Id})^2)$ ont dimension 2 et 3 respectivement.

8.6.7 Remarque. Les invariants de similitude contiennent l'information des polynômes minimal et caractéristique. Dans les notations du Théorème 8.6.1, on a

$$\mu_f = P_r \quad \text{et} \quad \chi_f = P_1 \cdots P_r$$

En effet, toujours dans les notations du Théorème 8.6.1, on a pour chaque $i = 1, \ldots, r$, $P_i = \chi_{f_{F_i}} = \mu_{f_{F_i}}$, et donc puisque tous les F_i sont stables,

$$\begin{aligned} \chi_f &= \chi_{f_{F_1}} \cdots \chi_{f_{F_r}} \qquad \text{et} \qquad \mu_f = \mathrm{ppcm}(\mu_{f_{F_1}}, \ldots, \mu_{f_{F_1}}) \\ &= P_1 \cdots P_r \qquad\qquad\qquad\quad = \mathrm{ppcm}(P_1, \ldots, P_r) = P_r. \end{aligned}$$

En particulier, on connaît désormais un algorithme pour calculer le polynôme minimal (en fait on en connaissait déjà un : calculer les puissances successives de A jusqu'à trouver le plus petit entier k tel que $\mathrm{Id}, A, \ldots, A^k$ soient liés comme vecteurs de $\mathcal{M}_n(\mathbf{k})$, puis chercher une relation de dépendance linéaire entre ces matrices ; l'algorithme du pivot de Gauss est nettement plus efficace).

8.6.8 Corollaire. *Les invariants de similitude sont invariants par extension des scalaires.*

En particulier, puisque le polynôme minimal est l'un des invariants de similitude, il est invariant par extension des scalaires. Nous renvoyons à l'exemple 8.6.10.1 plus loin pour une illustration instructive de ce résultat.

Preuve. Les invariants de similitude se calculent par l'algorithme de Gauss, qui ne fait pas sortir du corps de départ. Précisément, si on part d'une matrice $A \in \mathcal{M}_n(\mathbf{k})$, avec donc $X.\mathrm{Id} - A$ à coefficients dans $\mathbf{k}[X]$, étendre les scalaires revient à considérer A à coefficients dans $\mathbf{k}'$, une extension de corps de $\mathbf{k}$, et ainsi $X.\mathrm{Id} - A$ à coefficients dans $\mathbf{k}'[X]$. Mais si on opère le pivot de Gauss sur $X.\mathrm{Id} - A \in \mathcal{M}_n(\mathbf{k}'[X])$, on fait les mêmes opérations que si l'on opérait le pivot de Gauss sur $X.\mathrm{Id} - A \in \mathcal{M}_n(\mathbf{k}[X])$, et donc le résultat final est le même ! □

Une autre façon de prouver le corollaire est de dire directement que les facteurs invariants sont eux-mêmes invariants par extension des scalaires. Dans notre contexte, cela s'écrit de la manière suivante. Si $X.\mathrm{Id} - A$ est équivalente à $\mathrm{diag}(1, \ldots, 1, P_1, \ldots, P_r)$ dans $\mathcal{M}_n(\mathbf{k}[X])$, alors *a fortiori* ces deux matrices sont équivalentes dans $\mathcal{M}_n(\mathbf{k}'[X])$, donc les facteurs invariants de $X.\mathrm{Id} - A$ vue dans $\mathcal{M}_n(\mathbf{k}'[X])$ sont bien $1, \ldots, 1, P_1, \ldots, P_r$.

8.6.9 Remarque. Dans le cas des endomorphismes trigonalisables, nous avions déjà des invariants suffisants pour distinguer les classes de similitude (Théorème 7.8.10) et calculables algorithmiquement, ainsi qu'une forme normale représentant chacune de ces classes (Remarque 7.8.7). Ce que l'on gagne dans ce cas, c'est donc une forme un peu plus compacte pour organiser les invariants, et un algorithme lui aussi plus compact pour les calculer.

Justement, voyons comment déterminer les invariants de similitude d'un endomorphisme trigonalisable à partir de sa forme normale de Jordan.

8.6.10 Exemple. On considère la matrice diagonale par blocs

$$A = \operatorname{diag}\big(J_{a_{1,1}}(\lambda_1), \dots, J_{a_{1,k_1}}(\lambda_1),\ \dots\ , J_{a_{r,1}}(\lambda_r), \dots, J_{a_{1,k_r}}(\lambda_r)\big),$$

avec r valeurs propres $\lambda_1, \dots, \lambda_r$ deux à deux distinctes, et pour tout $i = 1, \dots, r$, k_i blocs de Jordan associés à la valeur propre λ_i, de tailles respectives $a_{i,1} \leqslant \dots \leqslant a_{i,k_i}$.

Quitte à rajouter formellement des blocs de taille 0, on peut supposer qu'il y a autant de blocs pour chaque valeur propre, autrement dit $k_1 = \dots = k_r$; nous noterons k ce nombre. En permutant les blocs, on obtient la matrice ci-dessous, semblable à A :

$$A' = \operatorname{diag}\big(J_{a_{1,1}}(\lambda_1), \dots, J_{a_{r,1}}(\lambda_r),\ \dots\ , J_{a_{1,k}}(\lambda_1), \dots, J_{a_{r,k}}(\lambda_r)\big),$$

D'après 8.5.9, le bloc de Jordan $J_a(\lambda)$ est semblable à la matrice compagnon de $(X-\lambda)^a$; pour tout $s = 1, \dots, k$, les deux matrices ci-dessous sont donc semblables,

$$\operatorname{diag}\big(J_{a_{1,s}}(\lambda_1), \dots, J_{a_{r,s}}(\lambda_r)\big) \quad \text{et} \quad A_s = \operatorname{diag}\big(\operatorname{Comp}\big((X-\lambda_1)^{a_{1,s}}\big), \dots, \operatorname{Comp}\big((X-\lambda_r)^{a_{r,s}}\big)\big).$$

Puisque $\lambda_1, \dots, \lambda_r$ sont deux à deux distinctes, les polynômes $(X-\lambda_1)^{a_{1,s}}, \dots, (X-\lambda_r)^{a_{r,s}}$ sont deux à deux premiers entre eux, donc d'après 8.5.7 la matrice A_s est cyclique associée au polynôme $\prod_{i=1}^r (X-\lambda_i)^{a_{i,s}}$ pour tout $s = 1, \dots, k$.

En conclusion, la matrice A est semblable à la matrice diagonale par blocs dont les blocs diagonaux sont les matrices compagnons associées aux polynômes $\prod_{i=1}^r (X-\lambda_i)^{a_{i,s}}$ pour $s = 1, \dots, k$. Puisque pour tout $i = 1, \dots, r$, on a choisi $a_{i,1} \leqslant \dots \leqslant a_{i,k_i}$, ces polynômes vérifient les relations de divisibilité requises de sorte que

$$\prod_{i=1}^r (X-\lambda_i)^{a_{i,1}},\ \dots\ ,\ \prod_{i=1}^r (X-\lambda_i)^{a_{i,k}}$$

sont les invariants de similitude de la matrice A.

8.6.10.1. Le cas particulier ci-dessous de 8.6.10 est intéressant du point de vue de l'invariance par extension des scalaires. On considère la matrice

$$R = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \in \mathcal{M}_2(\mathbf{R}).$$

Dans $\mathcal{M}_2(\mathbf{C})$, cette matrice est diagonalisable, semblable à la matrice diagonale $\operatorname{diag}(e^{i\theta}, e^{-i\theta})$. D'après ce qui précède, elle a un seul invariant de similitude,

$$(X - e^{i\theta})(X - e^{-i\theta})$$

qui est bien un polynôme à coefficients réels.

8.6.11 Exercice. Montrer qu'un endomorphisme trigonalisable est cyclique si et seulement si tous ses sous-espaces propres sont des droites.

8.6.12 Exercice. 1) Soit $\alpha_0 \in \mathbf{C}$. Donner un représentant de chacune des classes de similitude dans $\mathcal{M}_4(\mathbf{C})$ constituées de matrices ayant α comme unique valeur propre.

2) Pour chaque classe de similitude comme dans la question précédente, représentée par une matrice A, calculer :
a) les polynômes caractéristique et minimal χ_A et μ_A ;
b) pour tout $\alpha \in \mathbf{C}$ et $i \in \mathbf{N}$, la dimension du noyau de $(\alpha \mathrm{Id}_4 - A)^i$;
c) les invariants de similitude de A.

3) On suppose à présent $\alpha_0 \in \mathbf{R}$. Déterminer deux matrices $A', A'' \in \mathcal{M}_4(\mathbf{R})$ qui ne sont pas semblables, mais qui sont telles que pour $A = A', A''$:

$$\text{(i) } \dim\big(\ker(\alpha_0 \mathrm{Id}_4 - A)^i\big) = \begin{cases} i & \text{si } i \leqslant 1 \\ 2 & \text{sinon} \end{cases} ; \qquad \text{et (ii) } \dim\big(\ker(\alpha \mathrm{Id}_4 - A)^i\big) = 0$$

pour tout $\alpha \in \mathbf{R} - \{\alpha_0\}$ et tout $i \in \mathbf{N}$.
(*Indication :* nous remercions nos aimables lecteurs de chercher A' et A'' non-trigonalisables, pour éviter de trouver un contre-exemple à notre théorème 7.8.10).

8.6.1 – Existence d'une décomposition de Frobenius

Le point difficile est le lemme 8.6.14. Une fois ce point acquis, la preuve est une récurrence sans grand mystère. La preuve du lemme 8.6.14 fait intervenir la dualité très élégamment.

8.6.13 Preuve de la partie existence du théorème 8.6.1. On raisonne par récurrence sur la dimension de E. Si $\dim(E) \leqslant 1$, le résultat est trivial. Considérons donc E de dimension au moins 2, et supposons l'existence démontrée en dimension plus petite.

D'après la proposition 8.5.4 il existe $x \in E$ tel que $\mu_x = \mu_f$. Soit F_x le sous-espace cyclique associé à x. D'après le lemme 8.6.14 ci-dessous, il existe un supplémentaire G_x de F_x dans E qui est stable par f. On a $\dim(G_x) < \dim(E)$, donc par hypothèse de récurrence appliquée à l'endomorphisme induit $f_{G_x} \in \mathcal{L}(G_x)$ il existe des polynômes $P_1 | \cdots | P_r$ et une décomposition $G_x = \bigoplus_{i=1}^r G_i$ telle que chaque f_{G_i} soit cyclique de polynôme P_i. On a alors $E = F_x \oplus \bigoplus_{i=1}^r G_i$, et f_{F_x} est cyclique de polynôme $\mu_x = \mu_f$. On a $P_r = \mu_{f_{G_x}}$, voir remarque 8.6.7, donc $P_r | \mu_f$ (c'est la proposition 7.2.12). Ainsi $(P_1, \ldots, P_r, \mu_f)$ est une suite de polynômes comme on voulait. □

8.6.14 Lemme. *Soit $x \in E$ tel que $\mu_x = \mu_f$. Alors le sous-espace cyclique associé à x possède un supplémentaire dans E stable par f.*

8.6.14.1 Attention. En général il est faux que le sous-espace cyclique associé à x quelconque possède un supplémentaire stable.

Par exemple, si f est nilpotent cyclique, c'est-à-dire nilpotent d'indice $n = \dim(E)$, pour $x \in \ker(f^i) \setminus \ker(f^{i-1})$, $i < n$, le sous-espace cyclique F_x n'a pas de supplémentaire stable. En effet, on a $0 < \dim(F_x) < n$, donc si on avait une décomposition $E = F_x \oplus G$ avec G stable lui aussi, l'indice de nilpotence serait inférieur à $\max(\dim(F_x), \dim(G))$, donc strictement inférieur à n. Dans ce cas, $\mu_x = X^i$ et $\mu_f = X^n$.

Preuve. Notons p le degré de $\mu_f = \mu_x$. Nous allons définir explicitement un supplémentaire stable à F_x par p équations linéaires indépendantes. La famille $(x, f(x), \ldots, f^{p-1}(x))$ est une base de F_x. En particulier c'est une famille libre, et nous pouvons donc licitement considérer une forme linéaire $\ell \in E^\star$ telle que $\ell(f^i(x)) = 0$ si $i < p-1$ et $\ell(f^{p-1}(x)) = 1$.

Montrons que les formes linéaires $\ell\circ f^i$, $i=0,\dots,p-1$ sont linéairement indépendantes. Il suffit de vérifier que leurs restrictions à F_x le sont. Or la matrice de $(\ell,\dots,\ell\circ f^{p-1})\in\mathcal{L}(F_x,\mathbf{k}^p)$ dans la base $(x,f(x),\dots,f^{p-1}(x))$ est de la forme

$$\begin{pmatrix} 0 & \dots & 0 & 1 \\ \vdots & \iddots & 1 & * \\ 0 & \iddots & \iddots & \vdots \\ 1 & * & \dots & * \end{pmatrix}, \tag{8.6.14.1}$$

donc elle est inversible, et ainsi les restrictions des $\ell\circ f^i$ $(i=0,\dots,p-1)$ à F_x sont linéairement indépendantes comme il fallait.

On définit $G=(\ell,\dots,\ell\circ f^{p-1})^\perp$. Puisque $\ell,\dots,\ell\circ f^{p-1}$ sont linéairement indépendantes, c'est un sous-espace de dimension $n-p$ de E. Montrons que c'est un supplémentaire de F_x. Il suffit de montrer que $F_x\cap G=\{0\}$. Or un vecteur $z\in F_x$ est dans $(\ell,\dots,\ell\circ f^{p-1})^\perp$ si et seulement si ses coordonnées dans la base $(x,f(x),\dots,f^{p-1}(x))$ sont dans le noyau de la matrice (8.6.14.1). Celle-ci étant inversible, on a bien $F_x\cap G=\{0\}$.

Il reste à montrer que G est stable par f. Soit $z\in G$. On a

$$\ell(z)=\ell(f(z))=\dots=\ell(f^{p-1}(z))=0.$$

Puisque $\deg(\mu_f)=p$, $f^p(z)$ est une combinaison linéaire de $z,f(z),\dots,f^{p-1}(z))$. Ceci implique que $\ell(f^p(z))$ est une combinaison linéaire de $\ell(z),\ell(f(z)),\dots,\ell(f^{p-1}(z)))$, et donc que $\ell(f^p(z))=0$. Finalement on a donc

$$\begin{aligned}\ell(f(z))=\dots=\ell(f^{p-1}(z))=\ell(f^p(z))=0\\ \iff\quad \ell(f(z))=\dots=\ell(f^{p-2}(f(z)))=\ell(f^{p-1}(f(z)))=0\end{aligned}$$

et ainsi $f(z)\in G$ comme il fallait démontrer. □

8.6.15 Remarque. On peut comprendre la preuve ci-dessus un peu plus conceptuellement. En fait, le $\ell\in E^\star$ qu'on considère est tel que $\mu_\ell=\mu_{f^\mathsf{T}}$: le polynôme minimal local de $f^\mathsf{T}\in\mathcal{L}(E^\star)$ égale le polynôme minimal de f^T. En effet f et f^T ont même polynôme minimal, [4] de degré p, et on montre dans la preuve que $\ell,\ell\circ f=f^\mathsf{T}(\ell),\dots,\ell\circ f^{p-1}=(f^\mathsf{T})^{p-1}(\ell)$ sont linéairement indépendantes, donc $\deg\mu_\ell\geqslant p$, et finalement $\mu_\ell=\mu_{f^\mathsf{T}}$.

Alors

$$\mathrm{Vect}(\ell,\dots,\ell\circ f^{p-1})=\mathrm{Vect}(\ell,f^\mathsf{T}(\ell),\dots,({}^\mathsf{T}f)^{p-1}(\ell))=F_\ell$$

est le sous-espace cyclique de f^T associé à ℓ. Il est donc automatiquement stable par f, et de dimension p.

Enfin ℓ a été choisi de sorte qu'en plus $F_x\subseteq E$ et $F_\ell\subseteq E^\star$ sont naturellement en dualité, au sens où par le morphisme de restriction $\varphi\in E^\star\mapsto\varphi|_F\in F_x^\star$, F_ℓ est isomorphe à $F_x^\star$. Si on préfère (mais j'en doute), F_ℓ est un supplémentaire de $F_x^\perp$, donc il est isomorphe au quotient $E^\star/F_x^\perp$, qui lui est canoniquement isomorphe à $F_x^\star$ (voir 4.4.7).

8.6.15.1 Attention. La famille des restrictions de $\ell,\dots,\ell\circ f^{p-1}$ à F_x n'est en général pas la base duale de $(f^{p-1}(x),\dots,f(x),x)$ (et encore moins celle de $(x,f(x),\dots,f^{p-1}(x))$). à titre d'exercice, on pourra écrire explicitement la matrice (8.6.14.1) en fonction des coefficients du polynôme minimal $\mu_x=\mu_f$.

4. on peut s'en convaincre directement à titre d'exercice ; sinon les matrices de f et f^T dans des bases duales l'une de l'autre sont transposées l'une de l'autre, donc elles ont même polynôme minimal.

8.6.2 – Unicité de la décomposition de Frobenius

Comme nous l'avons annoncé plus haut, la partie unicité du théorème 8.6.1 est une conséquence directe de l'identification des invariants de similitude de f aux facteurs invariants non triviaux de la matrice $X.\mathrm{Id} - A \in \mathcal{M}_n(\mathbf{k}[X])$, où A est la matrice de f dans une base arbitraire (c'est la proposition 8.6.4). Nous allons donc ici démontrer cette proposition. Avant de le faire, nous attirons l'attention du lecteur sur une subtilité de l'énoncé du théorème 8.6.1.

8.6.16 Mise en garde. Si les invariants de similitude de f sont bien uniques, en général la décomposition en somme de sous-espaces cycliques elle ne l'est pas.

Par exemple, si f est une homothétie n'importe quelle décomposition de E en somme directe de droites est une décomposition de Frobenius de f.

Un autre exemple : si f est nilpotent avec k blocs de Jordan tous de la même taille a, de manière équivalente si f est nilpotent d'indice a tel que $\dim(\ker f^i) - \dim(\ker f^{i-1}) = k$ pour tout $i = 1, \ldots, a$, alors pour toute famille de vecteurs $x_1, \ldots, x_k$ telle que $(\bar{x}_1, \ldots, \bar{x}_k)$ est une base de $E/\ker(f^{a-1})$, les sous-espaces cycliques $F_{x_1}, \ldots, F_{x_k}$ fournissent une décomposition de Frobenius de f.

Nous prouvons la proposition 8.6.4 par une mise en oeuvre explicite du pivot de Gauss sur $X.\mathrm{Id} - A$ pour une matrice A sous forme décomposée de Frobenius.

8.6.17 Preuve de la proposition 8.6.4. Soit $P_1, \ldots, P_r$ des polynômes comme dans le théorème 8.6.1. Alors il existe une base de E dans laquelle la matrice de f est diagonale par blocs

$$A = \mathrm{diag}(\mathrm{Comp}(P_1), \ldots, \mathrm{Comp}(P_r)) \in \mathcal{M}_n(\mathbf{k}).$$

D'après le lemme 8.6.18 ci-dessous, pour chaque $i = 1, \ldots, r$ la matrice $X.\mathrm{Id}_{n_i} - \mathrm{Comp}(P_i)$ est équivalente dans $\mathcal{M}_{n_i}(\mathbf{k}[X])$, $n_i = \deg(P_i)$, à la matrice diagonale $\mathrm{diag}(1, \ldots, 1, P_i)$. On en déduit que la matrice $X.\mathrm{Id}_n - A$ est équivalente dans $\mathcal{M}_n(\mathbf{k}[X])$ à la matrice diagonale

$$\mathrm{diag}(1, \ldots, \ldots, 1, P_1, \ldots, P_r) \in \mathcal{M}_n(\mathbf{k}[X]).$$

□

8.6.18 Lemme. *Soit $P \in \mathbf{k}[X]$ unitaire non constant. Les facteurs invariants de la matrice $X.\mathrm{Id} - \mathrm{Comp}(P)$ sont $(1, \ldots, 1, P)$.*

Preuve. On écrit $P = X^n + a_{n-1}X + \cdots + a_0$. On effectue les opérations élémentaires suivantes sur les lignes de la matrice

$$X.\mathrm{Id} - \mathrm{Comp}(P) = \begin{pmatrix} X & & & & a_0 \\ -1 & \ddots & & & a_1 \\ & \ddots & \ddots & & \vdots \\ & & -1 & X & a_{n-2} \\ & & & -1 & X + a_{n-1} \end{pmatrix},$$

qu'on appelle $L_1, \ldots, L_n$: on effectue la permutation circulaire $L_n \to L_{n-1} \to \cdots \to L_1 \to L_n$, on obtient

$$\begin{pmatrix} -1 & X & & & a_1 \\ & \ddots & \ddots & & \vdots \\ & & -1 & X & a_{n-2} \\ & & & -1 & X + a_{n-1} \\ X & & & & a_0 \end{pmatrix},$$

puis on remplace L_n par

$$L_n + XL_1 + X^2L_2 + \cdots + X^{n-1}L_{n-1},$$

ce qui donne

$$\begin{pmatrix} -1 & X & & & a_1 \\ & \ddots & \ddots & & \vdots \\ & & -1 & X & a_{n-2} \\ & & & -1 & X + a_{n-1} \\ & & & & P(X) \end{pmatrix}.$$

Il ne reste alors plus qu'à faire du nettoyage automatique en opérant sur les colonnes pour arriver à la matrice diagonale $\mathrm{diag}(1, \ldots, 1, P(X))$.

Précisément, on remplace C_2 par $C_2 + XC_1$ et C_n par $C_n + a_1C_1$ pour mettre des 0 sur la première ligne, puis C_3 par $C_3 + XC_2$ et C_n par $C_n + a_2C_2$ pour mettre des 0 sur la seconde ligne, et ainsi de suite jusqu'à remplacer C_n par $C_n + (X + a_{n-1})C_{n-1}$ pour mettre un 0 sur l'avant-dernière ligne. On a alors obtenu la matrice $\mathrm{diag}(-1, \ldots, -1, P(X))$ qui convient à notre bonheur, qu'on peut transformer en $\mathrm{diag}(1, \ldots, 1, P(X))$ en multipliant les $n-1$ premières lignes par -1. □

8.7 – Interprétation en termes de $\mathbf{k}[X]$-modules

8.7.1 Modules. Définition d'un module. Ce qui change par rapport aux espaces vectoriels : il n'est pas toujours possible de résoudre les systèmes d'équations linéaires, en particulier ciao la théorie de la dimension. Dans la catégorie des espaces vectoriels, (i) E de dimension n possède des sev de toutes les dimensions $\leqslant n$, (ii) F sev possède toujours un supplémentaire, et (iii) toute suite exacte est scindée. Dans la catégorie des modules ces trois énoncés sont faux (on verra qu'ils le sont aussi dans la catégorie des groupes).

8.7.2 $\mathbf{k}[X]$-module associé à un endomorphisme. Le $\mathbf{k}[X]$-module $M_{E,f}$. Ceci revient à considérer que la donnée de $f \in \mathcal{L}(E)$ induit une représentation de l'algèbre $\mathbf{k}[X]$. La recherche de décompositions de ces représentations est exactement la théorie de la réduction des endomorphismes.

8.7.3 Proposition. *Les $\mathbf{k}[X]$-modules isomorphes à $M_{E,f}$ sont exactement les $M_{E',f'}$ pour lesquels qu'il existe $\varphi : E \cong E'$ tel que $f = \varphi^{-1} \circ f' \circ \varphi$..*

Preuve. Soit $\varphi : M_{E,f} \cong M$ isomorphisme de $\mathbf{k}[X]$-modules. Je définis $E' := M$ muni de la structure de $\mathbf{k}$-ev sous-jacente, et $f' := m_X \in \mathcal{L}(E')$. La relation $f = \varphi^{-1} \circ f' \circ \varphi$ est offerte par le fait que φ est un isomorphisme de $\mathbf{k}[X]$-modules, donc commute à la multiplication par $X \in \mathbf{k}[X]$. □

8.7.4 Sous-modules. Les sous modules de $M_{E,f}$ sont les M_{F,f_F}, F stable par f. Ceci dicte les définitions :

(i) f simple s'il n'a pas de sev stable non trivial ;
(ii) f semi-simple si tout stable possède un supplémentaire stable.

Invariants de similitude : on les a définis plus haut, et on sait qu'on peut les trouver de la manière suivante : on prend une base, on fabrique $X \cdot \mathrm{Id}_n - M$, et on la met sous forme réduite.

8.7.5 Théorème. *Deux matrices sont semblables ssi elles ont les mêmes invariants de similitude.*

8.7.5.1. *Le $\mathbf{k}[X]$-module $M_{E,f}$ est isomorphe au quotient $E[X]/\mathrm{im}(X\cdot\mathrm{id}-f)$ via*

$$\pi : \sum e_i X^i \mapsto \sum f^i(e_i).$$

Preuve. on vérifie que $\pi((X\mathrm{id}-f)(\sum e_i X^i))=0$, et réciproquement si $\sum e_i X^i \in \ker\pi$ alors

$$\begin{aligned}\sum e_i X^i &= \sum e_i X^i - 0\\ &= \sum e_i X^i - \sum f^i(e_i)\\ &= \sum (X^i\mathrm{id}-f^i)(e_i),\end{aligned}$$

et chaque $X^i\mathrm{id}-f^i=(X\mathrm{id})^i-f^i$ se factorise par $X\mathrm{id}-f$. □

8.7.5.2 Conclusion.

$$\frac{E[X]}{\mathrm{im}(X\cdot\mathrm{id}-f)} \;\cong\; \frac{\mathbf{k}[X]}{(P_1)}\oplus\cdots\oplus\frac{\mathbf{k}[X]}{(P_n)}.$$

□

8.7.6 Remarque. On a démontré que les invariants de similitude peuvent être définis à partir du $\mathbf{k}[X]$-module $M_{E,f}$, en utilisant le fait qu'il est de type fini.

8.7.7 Décomposition de Frobenius. Chaque $E_i=\mathbf{k}[X]/(P_i)$ est un sev stable par f et $f_i:=f_{E_i}$ est cyclique avec $\chi_{f_i}=\mu_{f_i}=P_i$. La décomposition

$$M_{E,f}=E_1\oplus\cdots\oplus E_n$$

est une somme directe de sous-espaces cycliques pour f.

On en déduit $\chi_f=\chi_{f_1}\cdots\chi_{f_n}=P_1\cdots P_n$, et $\mu_f=P_n$. Ceci dévoile le Théorème de Cayley–Hamilton (je ne dirais pas que ça le trivialise...), y compris sa version améliorée "χ et μ ont les mêmes facteurs irréductibles".

8.7.7.1 Exercice. f cyclique $\Leftrightarrow$ tous les P_i sauf le dernier sont égaux à 1.
[En particulier, on fait remarquer qu'en général il y a un certain nombre d'invariants égaux à 1, et que ceux-ci donnent des facteurs triviaux dans la décomposition cyclique.]

8.7.8 Décomposition de Dunford–Jordan d'un bloc cyclique.

$$\frac{\mathbf{k}[X]}{((X-\lambda_1)^{a_1}\cdots(X-\lambda_r)^{a_r})} \;\cong\; \frac{\mathbf{k}[X]}{(X-\lambda_1)^{a_1}}\oplus\cdots\oplus\frac{\mathbf{k}[X]}{(X-\lambda_r)^{a_r}}$$

par le lemme chinois ; pour chaque morceau de la décomposition de droite, la multiplication par X s'écrit

$$\begin{pmatrix}\lambda & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1\\ & & & \lambda\end{pmatrix} \tag{8.7.8.1}$$

dans la base $\big((X-\lambda)^{a-1},\dots,X-\lambda,1\big)$, puisque

$$X\cdot(X-\lambda)^i=(X-\lambda)^{i+1}+\lambda(X-\lambda)^i.$$

(8.7.8.1) est un bloc de Dunford–Jordan de taille a.

8.7.9 Exercice. Soit $P \in \mathbf{k}[X]$. On considère le $\mathbf{k}$-espace vectoriel $E_P = \mathbf{k}[X]/(P)$ (il se trouve que c'est aussi une $\mathbf{k}[X]$-algèbre), et l'endomorphisme $m_P \in \mathcal{L}(E_P)$ de multiplication par X :

$$\forall H \in \mathbf{k}[X], \quad m_P(\bar{H}) = \overline{X \cdot H}.$$

1) On note $n = \deg(P)$. Montrer que E_P est de dimension n, muni de la base canonique $(\bar{1}, \bar{X}, \ldots, \bar{X}^{n-1})$.

2) Démontrer que $\mu_{m_P} = P$.

3) Écrire la matrice de m_P dans la base canonique. En conclure que le polynôme minimal de la matrice compagnon associée à P est P lui-même.

8.8 – Invariants de similitude

8.8.1 – Principe

Soit E un espace vectoriel de dimension finie sur un corps $\mathbf{k}$. Lorsque $\mathbf{k}$ est algébriquement clos, on sait déjà donner la liste complète des classes de similitude dans $\mathcal{L}(E)$ (c'est-à-dire les orbites de $\mathcal{L}(E)$ sour l'action de $\mathrm{GL}(E)$ par conjugaison), et étant donné $f \in \mathcal{L}(E)$ on sait déterminer algorithmiquement à quelle classe il appartient : il faut commencer par déterminer ses valeurs propres, puis déterminer sa forme normale de Jordan en calculant pour chaque $\lambda \in \mathrm{Sp}(f)$ la suite $\big(\dim(\ker(f - \lambda.\mathrm{id})^i\big)_{i \geqslant 0}$.

Ici nous allons expliquer comment tout ceci peut s'obtenir en effectuant un seul pivot de Gauss à coefficients dans $\mathbf{k}[X]$. Moyennant le choix d'une base de E, on identifie f à une matrice carrée $A \in \mathcal{M}_n(\mathbf{k})$. On considère alors la matrice $A - X.\mathrm{Id} \in \mathcal{M}_n(\mathbf{k}[x])$, et on sait d'après le théorème des facteurs invariants 3.1.1 qu'il existe une unique suite de polynômes unitaires $P_1, \ldots, P_n$ tels que les matrices

$$A - X.\mathrm{Id}_n \quad \text{et} \quad \begin{pmatrix} P_1 & & \\ & \ddots & \\ & & P_n \end{pmatrix}$$

soient équivalentes dans $\mathcal{M}_n(\mathbf{k}[X])$. Ces polynômes $P_1, \ldots, P_n$ se calculent directement en appliquant l'algorithme du pivot de Gauss à $A - X.\mathrm{Id}$ comme il est expliqué en 3.2. On les appelle les *invariants de similitude* de f (ce sont les facteurs invariants de la matrice $A - X.\mathrm{Id}$), et en effet ils caractérisent la classe de similitude de f en vertu du résultat suivant.

8.8.1 Théorème. *Il existe un isomorphisme de $\mathbf{k}$-espaces vectoriels*

$$\varphi : E \xrightarrow{\simeq} \prod_{i=1}^{n} \Big(\mathbf{k}[X]/(P_i)\Big) \tag{8.8.1.1}$$

tel que $f = \varphi^{-1} m_X \varphi$, où m_X est l'endomorphisme du $\mathbf{k}$-espace vectoriel à droite de (8.8.1.1) *défini par*

$$\forall\ (\overline{Q_1}, \ldots, \overline{Q_n}) \in \prod_{i=1}^{n} \Big(\mathbf{k}[X]/(P_i)\Big) : \quad m_X(\overline{Q_1}, \ldots, \overline{Q_n}) = (\overline{XQ_1}, \ldots, \overline{XQ_n}).$$

Ainsi deux endomorphismes ayant les mêmes invariants de similitude sont semblables. Réciproquement, si f et g sont deux endomorphismes semblables, alors leurs matrices A et B dans la base qu'on a choisie sont semblables dans $\mathcal{M}_n(\mathbf{k})$, et ceci implique que les matrices $A - X.\mathrm{Id}$ et $B - X.\mathrm{Id}$ sont semblables dans $\mathcal{M}_n(\mathbf{k}[X])$. Elles sont donc *a fortiori* équivalentes, et ont donc les mêmes facteurs invariants. Finalement, deux endomorphismes sont semblables si et seulement si ils ont les mêmes invariants de similitudes.

On verra que $P_1 = \mu_f$ (polynôme minimal de f) et $P_1 \cdots P_n = \chi_f$ (polynôme caractéristique de f). On donne dans la partie 8.8.2 la traduction du Théorème 8.8.1 dans le langage habituel de l'algèbre linéaire.

Décrivons synthétiquement les idées qui vont nous permettre d'arriver à ce résultat. Premièrement, on va associer de manière naturelle un $k[X]$-module $M_{E,f}$ à l'endomorphisme f, de sorte que deux endomorphismes sont semblables si et seulement si leurs $\mathbf{k}[X]$-modules associés sont isomorphes. Ce $\mathbf{k}[X]$-module est de type fini, et même de torsion. Il est donc redevable du théorème de classification des $\mathbf{k}[X]$-modules de type fini, qui est exactement analogue au Théorème **??** pour les groupes abéliens de type fini. Ainsi, $M_{E,f}$ est isomorphe comme $\mathbf{k}[X]$-module à un produit de $\mathbf{k}[X]$-modules $\mathbf{k}[X]/(P_i)$. On démontrera que $M_{E,f}$ s'identifie au quotient de $k[X]^n$ ($n = \dim(E)$) par l'image de l'endomorphisme

$$f - X.\mathrm{id} \in \mathcal{L}_{\mathbf{k}[X]}(\mathbf{k}[X]^n),$$

ce qui permettra d'identifier les P_i à ceux du Théorème 8.8.1.

D'une certaine manière, voir f à travers le prisme du $\mathbf{k}[X]$-module $M_{E,f}$ revient à associer à f une représentation de l'algèbre $\mathbf{k}[X]$ sur E. De ce point de vue, la décomposition 8.8.1.1 n'est autre que la décomposition isotypique de cette représentation.

8.8.2 – Exemples fondamentaux

8.8.2 Bloc cyclique. Soit $P \in \mathbf{k}[X]$ un polynôme unitaire de degré d. On écrit $P = X^d + a_1X^{d-1} + \cdots + a_d$. Le quotient $\mathbf{k}[X]/(P)$ est un $\mathbf{k}$-espace vectoriel de dimension d, et la famille $(\overline{1}, \overline{X}, \ldots, \overline{X^{d-1}})$ en est une base. L'application

$$m_X : \overline{Q} \mapsto \overline{XQ}$$

est un endomorphisme du $\mathbf{k}$-espace vectoriel $\mathbf{k}[X]/(P)$. Sa matrice dans la base $(\overline{1}, \overline{X}, \ldots, \overline{X^{d-1}})$ est

$$C(P) = \begin{pmatrix} 0 & & & -a_d \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & -a_2 \\ & & 1 & -a_1 \end{pmatrix},$$

on reconnaît la *matrice compagnon* du polynôme P.

On peut démontrer de manière limpide que le polynôme minimal de m_X (et donc celui de la matrice compagnon $C(P)$) est P. Soit $F \in \mathbf{k}[X]$. L'endomorphisme $F(m_X)$ associe à toute classe $\overline{Q}$ la classe $\overline{FQ}$. Ainsi $F(m_X)$ est l'endomorphisme nul de $\mathbf{k}[X]/(P)$ si et seulement

$$F(m_X) = 0_{\mathcal{L}_{\mathbf{k}}(\mathbf{k}[X]/(P))} \quad \Longleftrightarrow \quad \forall \overline{Q} \in \mathbf{k}[X]/(P) : \ \overline{FQ} = 0_{\mathbf{k}[X]/(P)}.$$

Cette dernière condition équivaut à $F \in (P)$: clairement tout multiple de P vérifie cette condition, et réciproquement si F la vérifie, alors pour $Q = 1$ on a $\overline{F} = 0$, c'est-à-dire $F \in (P)$.

D'autre part on a déjà calculé que le polynôme caractéristique de $C(P)$ est P lui-même. Ainsi les polynômes minimal et caractéristique de m_X sont égaux : on dit que m_X est un endomorphisme *cyclique*. C'est un exercice classique de vérifier que réciproquement, tout endomorphisme cyclique est donné dans une base adaptée par la matrice compagnon de son polynôme caractéristique et minimal.

Le Théorème 8.8.1 fournit une *décomposition de Frobenius* de f : chacun des facteurs $\mathbf{k}[X]/(P_i)$ est un sous-espace stable par m_X, sur lequel l'endomorphisme se restreint à un endomorphisme cyclique de polynôme P.

8.8.3 Bloc de Jordan. On considère ici le cas particulier du paragraphe 8.8.2 où $P = (X-\lambda)^d$, $\lambda \in \mathbf{k}$. Dans ce cas, il et judicieux de considérer $\mathbf{k}[X]/(P)$ muni de la base $(\overline{1}, \dots, \overline{X-\lambda}^{d-1})$. Dans cette base, puisque

$$X(X-\lambda)^i = (X-\lambda)(X-\lambda)^i + \lambda(X-\lambda)^i = (X-\lambda)^{i+1} + \lambda(X-\lambda)^i,$$

la matrice de l'endomorphisme de multiplication par X est

$$J_d(\lambda) = \begin{pmatrix} \lambda & & & \\ 1 & \lambda & & \\ & \ddots & \ddots & \\ & & 1 & \lambda \end{pmatrix},$$

on reconnaît un bloc de Jordan de taille d pour la valeur propre λ.

8.8.4 Isomorphisme chinois. Considérons à présent le cas où

$$P = (X-\lambda_1)^{d_1} \cdots (X-\lambda_r)^{d_r},$$

avec les λ_i deux à deux différents. Alors les $(X-\lambda_i)^{d_i}$ sont deux à deux premiers entre eux, donc on a un isomorphisme de $\mathbf{k}$-algèbres

$$\overline{Q} \in \mathbf{k}[X]/(P) \xmapsto{\cong} (Q \bmod P_1, \dots, Q \bmod P_r) \in \prod_{i=1}^{r} \mathbf{k}[X]/(P_i).$$

Via cet isomorphisme, multiplier par $\overline{X}$ dans le membre de gauche revient à multiplier par $(\overline{X}, \dots, \overline{X})$ dans le membre de droite. Ainsi chaque

$$\{0\} \times \cdots \times \mathbf{k}[X]/(P_i) \times \cdots \times \{0\}$$

(qu'on notera plus légèrement $\mathbf{k}[X]/(P_i)$ par abus de notation) est un sous-espace stable par m_X, sur lequel m_X agit par multiplication par X : « m_X de $\mathbf{k}[X]/(P)$ se décompose en m_X des $\mathbf{k}[X]/(P_i)$ ».

D'après le paragraphe précédent 8.8.3, m_X de $\mathbf{k}[X]/(P_i)$ s'écrit dans une base adaptée comme un bloc de Jordan $J_{d_i}(\lambda_i)$. On conclut donc que dans une base adaptée, m_X de $\mathbf{k}[X]/(P)$ s'écrit comme une matrice diagonale par blocs

$$\begin{pmatrix} J_{d_1}(\lambda_1) & & \\ & \ddots & \\ & & J_{d_r}(\lambda_r) \end{pmatrix}.$$

8.8.5 Facteurs invariants et décomposition de Dunford–Jordan. La condition (i) $P_1|\cdots|P_n$ sert à assurer l'unicité des facteurs invariants. Pour avoir bien compris, il faut s'assurer de savoir démontrer l'affirmation suivante. Soit $f \in \mathcal{L}(E)$ à polynôme caractéristique scindé, de valeurs propres $\lambda_1, \ldots, \lambda_r$ deux à deux différentes. On suppose que dans la décomposition de Dunford–Jordan de f, pour tout $i = 1, \ldots, r$, la valeur propre λ_i apparaît dans des blocs de Jordan de tailles $0 \leqslant d_{i,1} \leqslant \ldots \leqslant d_{i,n}$ (on autorise des blocs de Jordan de taille 0). Alors les invariants de similitude de f sont les polynômes $P_1, \ldots, P_n$ avec pour tout $j = 1, \ldots, n$

$$P_j = (X - \lambda_1)^{d_{1,j}} \cdots (X - \lambda_r)^{d_{r,j}}.$$

On remarquera qu'en général les premiers invariants de similitude peuvent être égaux à 1, auquel cas les facteurs correspondants dans la décomposition (8.8.1.1) sont triviaux.

8.8.6 Exercice.

1) Soit $\alpha \in \mathbf{C}$. Donner un représentant de chacune des classes de similitude dans $\mathcal{M}_4(\mathbf{C})$ constituées de matrices ayant α comme unique valeur propre.

2) Pour chaque classe $\overline{A}$ comme dans la question précédente, calculer :
a) les polynômes caractéristique et minimal χ_A et μ_A ;
b) pour tout $\beta \in \mathbf{C}$ et $i \in \mathbf{N}$, la dimension du noyau de $(\beta \mathrm{Id}_4 - A)^i$;
c) les invariants de similitude de A.

3) Soit $\alpha \in \mathbf{R}$. Déterminer deux matrices $A', A'' \in \mathcal{M}_4(\mathbf{R})$ qui ne sont pas semblables, mais qui sont telles que pour $A = A', A''$:

$$\text{(i) } \dim\big(\ker(\alpha \mathrm{Id}_4 - A)^i\big) = \begin{cases} i & \text{si } i \leqslant 1 \\ 2 & \text{sinon} \end{cases} ; \qquad \text{et (ii) } \dim\big(\ker(\beta \mathrm{Id}_4 - A)^i\big) = 0$$

pour tout $\beta \in \mathbf{R} - \{\alpha\}$ et tout $i \in \mathbf{N}$.

Bibliographie

[H2G2I] P. Caldero & J. Germoni. *Histoires hédonistes de groupes et de géométries, Tome premier*. Calvage & Mounet, 2013.

[H2G2II] P. Caldero & J. Germoni. *Histoires hédonistes de groupes et de géométries, Tome second*. Calvage & Mounet, 2014.

[Cia82] P. G. Ciarlet. *Introduction à l'analyse numérique matricielle et à l'optimisation*. Masson, 1982.

[Deb] O. Debarre. Réduction des endomorphismes. Notes de cours à l'ENS.

[HP54] W. V. D. Hodge & D. Pedoe. *Methods of algebraic geometry*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 1994. Reprint of the 1954 original.

[Lasz] Y. Laszlo. Introduction à l'algèbre commutative et homologique. Cours de Maîtrise 2003–2004.

[Per96] D. Perrin. *Cours d'algèbre*. Ellipses, 1996.

[Ser89] E. Sernesi. *Geometria 1 e 2*. Bollati Boringhieri, 1989.

[Szp09] A. Szpirglas, editor. *Mathématiques L3 Algèbre*. Pearson Education, 2009. Lien vers version électronique.