Arithmétique et algèbre linéaire approfondie

Licence 2^{e} année, unité MAT301

Cours de Jean-Pierre Demailly

Année 2018-2019

Table des matières

Chapitre 1
Notions fondamentales et compléments sur les espaces vectoriels

Le but de l'algèbre linéaire est de décrire tous les phénomènes de nature vectorielle et les propriétés de linéarité qui leurs sont liées. Il est important de savoir effectuer des calculs en coordonnées, mais il est encore plus important de visualiser géométriquement les objets mis en jeu et de savoir faire le lien entre les propriétés géométriques et leur traduction algébrique.

1. Notions d'espace et de sous-espace vectoriel

1.1. Espace vectoriel

On se place ici sur un corps commutatif $(\mathbb{K}, +, \times)$, qui sera en général l'un des corps $\mathbb{K} = \mathbb{Q}$ (nombres rationnels), $\mathbb{K} = \mathbb{R}$ (nombres réels), ou $\mathbb{K} = \mathbb{C}$ (nombres complexes). Un rappelle qu'un corps commutatif $\mathbb{K}$ est une structure pour laquelle :

- l'addition $+$ est associative, commutative, dotée d'un élément neutre $0 = 0_{\mathbb{K}}$, tout élément $\lambda \in \mathbb{K}$ ayant un opposé $-\lambda \in \mathbb{K}$ tel que $\lambda + (-\lambda) = 0$.
- la multiplication $\times$ est associative, commutative, distributive par rapport à $+$, dotée d'un élément neutre $1 = 1_{\mathbb{K}}$, et telle que tout élément $\lambda \in \mathbb{K} \setminus \{0\}$ possède un inverse $\lambda' = \lambda^{-1}$ tel que $\lambda \times \lambda' = 1$.

1.1.1. Définition. *Un* $\mathbb{K}$-espace vectoriel *est une structure* $(E, +, \cdot)$ *formée d'un ensemble* E *et de deux lois* $+, \cdot$

$$\begin{array}{lll} \text{(loi interne)} & E \times E \to E, & (x, y) \mapsto x + y, \\ \text{(loi externe)} & \mathbb{K} \times E \to E, & (\lambda, x) \mapsto \lambda \cdot x, \end{array}$$

vérifiant les propriétés suivantes:

- pour l'addition, $(E, +)$ est un groupe commutatif :

A1 (*associativité de* $+$) *pour tous* $x, y, z \in E$, $x + (y + z) = (x + y) + z$;

A2 (*commutativité de* $+$) *pour tous* $x, y \in E$, $x + y = y + x$;

A3 (*élément neutre*) *il existe un élément* 0_E *tel que pour tout* $x \in E$ *on ait* $0_E + x = x + 0_E = x$;

A4 *pour tout* $x \in E$, *il existe un élément* $x' \in E$ *tel que* $x + x' = x' + x = 0_E$ (*cet élément* x' *est alors unique, il est appelé opposé de* x *et noté* $-x$) ;

- propriétés de la loi externe :

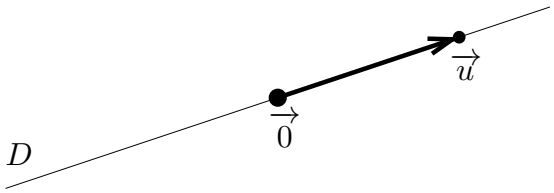
E1 *pour tout* $x \in E$, $1 \cdot x = x$,

E2 *pour tous* $\lambda, \mu \in \mathbb{K}$, $x \in E$, $(\lambda \times \mu) \cdot x = \lambda \cdot (\mu \cdot x)$;

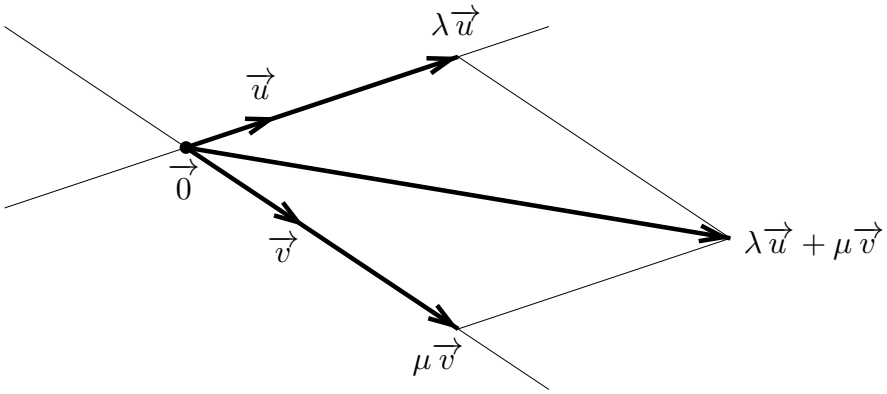
E3 *pour tous* $x, y \in E$, $\lambda \in \mathbb{K}$, $\lambda \cdot (x + y) = \lambda \cdot x + \lambda \cdot y$;

E4 *pour tous* $x \in E$, $\lambda, \mu \in \mathbb{K}$, $(\lambda + \mu) \cdot x = \lambda \cdot x + \mu \cdot x$.

Pour ne pas alourdir les notations, il est fréquent que l'on note simplement 0 au lieu de 0_E, que l'on parle d'espace vectoriel E plutôt que $(E, +, \cdot)$, et que l'on omette le $\cdot$ dans la notation de la multiplication externe : on notera ainsi λx plutôt que $\lambda \cdot x$, de même que l'on note simplement $\lambda\mu$ le produit $\lambda \times \mu$ dans le corps $\mathbb{K}$. Au contraire, pour insister sur le fait qu'on manipule des vecteurs, en particulier en Physique, il est fréquent d'utiliser des flèches et de noter un vecteur $\overrightarrow{x}$ au lieu de x.



On a représenté ci-dessus une droite vectorielle D. Chaque "point" de la droite doit être pensé comme un vecteur, c'est-à-dire que le point noir figurant $\overrightarrow{u}$ représente en réalité la flèche qui joint ce "point" à $\overrightarrow{0}$. L'addition des vecteurs "géométriques" du plan et de l'espace se fait au moyen de la règle du parallélogramme :



1.1.2. Exemples d'espaces vectoriels.

(a) L'ensemble $\mathbb{K}^n$ des n-uplets $(x_1, \ldots, x_n)$ d'éléments de $\mathbb{K}$ muni des lois

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) := (x_1 + y_1, \ldots, x_n + y_n)$$
$$\lambda \cdot (x_1, \ldots, x_n) := (\lambda x_1, \ldots, \lambda x_n)$$

est un $\mathbb{K}$-espace vectoriel.

(b) L'ensemble $\mathbb{K}[X]_n$ des polynômes à coefficients dans $\mathbb{K}$ de degré $\leqslant n$,

$$P(X) = \sum_{i=0}^{n} a_i X^i, \quad a_i \in \mathbb{K}$$

(où X est une "indéterminée", c'est-à-dire un symbole formel), avec les lois usuelles $(P,Q) \mapsto P+Q$, $(\lambda, P) \mapsto \lambda P$, $\lambda \in \mathbb{K}$, est un $\mathbb{K}$-espace vectoriel ; on notera que $\mathbb{K}[X]_n$ est (par définition) en bijection avec $\mathbb{K}^{n+1}$ par l'application

$$\mathbb{K}^{n+1} \to \mathbb{K}[X]_n, \qquad (a_0, a_1, \dots, a_n) \mapsto P(X) = \sum_{i=0}^{n} a_i X^i.$$

L'ensemble de tous les polynômes de degré non précisé, défini comme

$$\mathbb{K}[X] = \bigcup_{n \in \mathbb{N}} \mathbb{K}[X]_n$$

est lui aussi un espace vectoriel.

(c) *Espace vectoriels fonctionnels* : si A est un ensemble quelconque, l'ensemble des applications noté $\mathbb{K}^A$ ou encore $\mathcal{F}(A, \mathbb{K})$ tel que

$$\mathbb{K}^A = \mathcal{F}(A, \mathbb{K}) = \big\{\text{applications } f : A \to \mathbb{K}\big\}$$

avec les lois $(f,g) \mapsto f+g$ et $(\lambda, f) \mapsto \lambda f$ telles que

$$\forall t \in A, \quad (f+g)(t) = f(t) + g(t), \quad (\lambda f)(t) = \lambda\, f(t)$$

est un espace vectoriel. On notera que si $A = \{1, 2, \dots, n\}$, on peut identifier K^A à $\mathbb{K}^n$, puisque se donner une application $A \to \mathbb{K}$ est la même chose que se donner un n-uplet $(x_1, \dots, x_n) \in \mathbb{K}^n$.

1.1.3. Contre-exemple. *L'ensemble $\mathcal{F}(\mathbb{R}, \mathbb{R}_+)$ des fonctions $f : \mathbb{R} \to \mathbb{R}_+$ n'est pas un $\mathbb{R}$-espace vectoriel pour les lois usuelles, car la loi externe de multiplication par un réel négatif n'y est pas définie ; l'addition est bien définie comme loi de composition interne, mais seule la fonction $f = 0$ admet un symétrique $-f$ pour l'addition.*
On notera en revanche que $\mathcal{F}(\mathbb{R}_+, \mathbb{R})$ est bien un $\mathbb{R}$-espace vectoriel d'après ce qui précède.

1.2. Sous-espaces vectoriels

1.2.1. Définition. *Un sous-espace vectoriel S d'un $\mathbb{K}$-espace vectoriel E est un sous-ensemble S de E ayant les propriétés suivantes :*

(i) *S est non vide* ;

(ii) *S est stable pour l'addition* : $\forall x, y \in S$, *on a* $x + y \in S$;

(iii) *S est stable pour la loi externe* : $\forall \lambda \in \mathbb{K}$, $\forall x \in S$, *on a* $\lambda \cdot x \in S$.

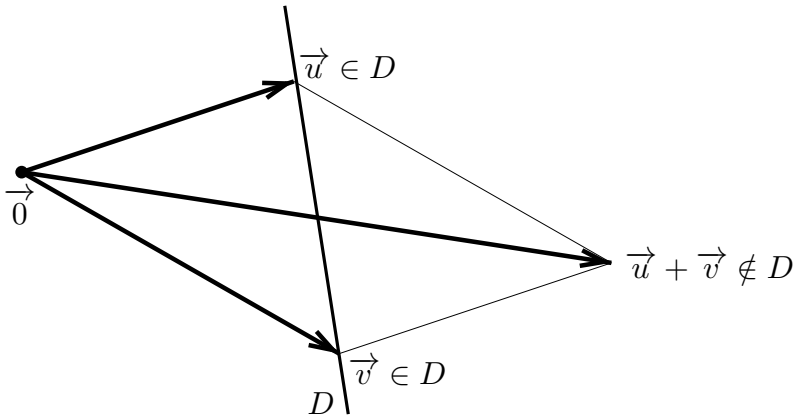
En prenant $\lambda = 0$ et un vecteur quelconque $x \in S$, on voit que l'on a nécessairement $0_E \in S$, et d'autre part on peut reformuler (ii) et (iii) en demandant que S soit stable par combinaisons linéaires :

1.2.1′. Définition équivalente. *Un sous-espace vectoriel S d'un $\mathbb{K}$-espace vectoriel E est un sous-ensemble S de E ayant les propriétés suivantes :*

(i′) $0_E \in S$;

(ii′) *S est stable par combinaisons linéaires* : $\forall \lambda, \mu \in \mathbb{K}$, $\forall x, y \in S$, *on a* $\lambda x + \mu y \in S$.

Il résulte de ce qui précède qu'une droite $D : ax + by = c$ de $\mathbb{R}^2$ ne contenant pas l'origine n'est pas un sous-espace vectoriel ; on voit d'ailleurs aussitôt qu'une telle droite n'est pas stable pour l'addition des vecteurs :



(Attention : ce sont bien toujours les extrémités des vecteurs que l'on représente dans une figure vectorielle, ici pour figurer la droite D, qui est une droite affine et non une droite vectorielle).

1.2.2. Propriétés. *Soit E un $\mathbb{K}$-espace vectoriel.*

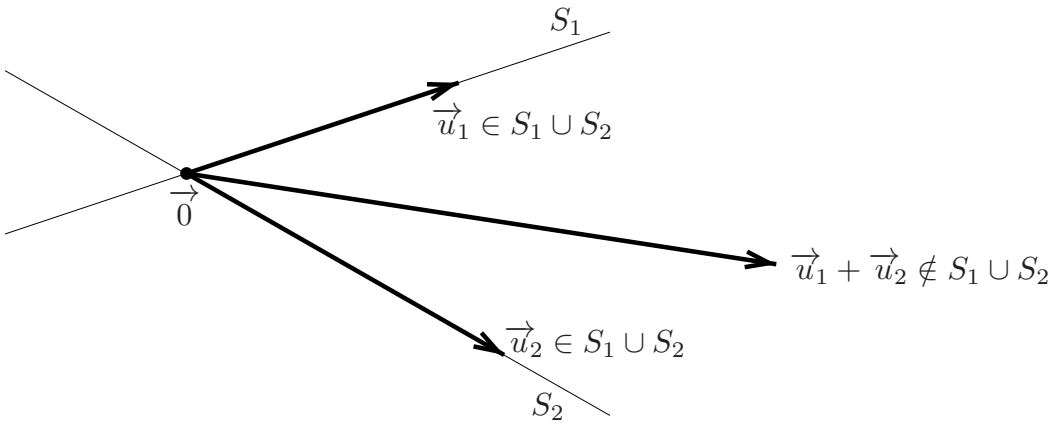
(a) *L'intersection $S_1 \cap S_2$ de deux sous-espaces vectoriels $S_1\, S_2$ de E est un sous-espace vectoriel de E, et plus généralement, toute intersection finie ou infinie d'une famille de sous-espaces vectoriels $(S_i)_{i\in I}$ de E, définie comme*

$$S = \bigcap_{i\in I} S_i = \big\{x \in E \,/\, \forall i \in I,\ x \in S_i\big\}$$

est un sous-espace vectoriel de E.

En effet, on a bien $0_E \in S$, et quels que soient $\lambda, \mu \in \mathbb{K}$ et $x, y \in S$, on a pour tout $i \in I$ que $x, y \in S_i$, donc $\lambda x + \mu y \in S_i$, et par conséquent $\lambda x + \mu y \in S = \bigcap_{i\in I} S_i$.

(b) *En revanche, la réunion $S_1 \cup S_2$ de deux sous-espaces vectoriels n'est pas, en général, un sous-espace vectoriel, comme le montre le schéma ci-dessous.*



1.2.3. Sommes de sous-espaces vectoriels. Si $(S_i)_{i\in I}$ est une famille de sous-espaces vectoriels du $\mathbb{K}$-espace vectoriel E, on définit leur somme comme étant

l'ensemble des vecteurs noté

$$\sum_{i\in I} S_i = \Big\{ \sum_{i\in I} x_i \,/\, \forall i,\ x_i \in S_i \Big\},$$

ceci, du moins, lorsque la famille I est finie. Si l'ensemble I est infini, on convient de ne prendre que les sommes où les x_i sont presque tous nuls, c'est-à-dire nuls sauf un nombre fini (de sorte que la sommation se réduise toujours à une somme finie, sinon cela n'aurait algébriquement pas de sens).

Il est clair que la somme $\sum_{i\in I} S_i$ contient 0_E et qu'elle est stable par combinaisons linéaires, puisque pour tous $\lambda, \mu \in \mathbb{K}$ et $x_i, y_i \in S_i$ presque tous nuls on peut écrire

$$\lambda \sum_{i\in I} x_i + \mu \sum_{i\in I} y_i = \sum_{i\in I} (\lambda x_i + \mu y_i) \quad \text{avec } \lambda x_i + \mu y_i \in S_i,$$

de sorte que cette combinaison linéaire appartient bien à $\sum_{i\in I} S_i$. □

1.2.4. Influence du corps. On remarquera que si E est un $\mathbb{C}$-espace vectoriel, alors c'est aussi un $\mathbb{R}$-espace vectoriel, puisqu'il suffit de ne considérer que les produits λx avec $\lambda \in \mathbb{R}$ et $x \in E$. On se méfiera cependant du fait que la dimension sur $\mathbb{R}$ et sur $\mathbb{C}$ n'est pas la même, par exemple $\mathbb{C} = \mathbb{C}^1$ est de dimension 1 comme $\mathbb{C}$ espace vectoriel, mais $\mathbb{C} \simeq \mathbb{R}^2$ est de dimension 2 comme $\mathbb{R}$-espace vectoriel. Nous reviendrons là dessus plus loin.

2. Familles génératrices, libres, bases, notion de rang

2.1. Sous-espace engendré par une partie ou une famille de vecteurs

Soit E un $\mathbb{K}$-espace vectoriel. On appelle famille de vecteurs toute collection indexée $(v_i)_{i\in I}$ d'éléments de E (et, très souvent, I sera un ensemble fini comme par exemple $I = \{1, 2, \ldots, n\}$). On appelle combinaison linéaire des vecteurs de la famille $(v_i)_{i\in I}$ toute somme *finie*

$$\sum_{i\in I} \lambda_i v_i, \quad \lambda_i \in \mathbb{K},$$

c'est-à-dire qu'on s'astreint toujours à prendre les coefficients $(\lambda_i)_{i\in I}$ presque tous nuls. De la même manière, si P est une partie de E, on appelle combinaison linéaire de vecteurs de P toute somme *finie*

$$\sum \lambda_i x_i, \quad \lambda_i \in \mathbb{K}, x_i \in P.$$

Si $P = \emptyset$, on convient par définition qu'une somme vide donne le vecteur nul 0_E.

2.1.1. Théorème et définition. *Si P est une partie de E quelconque, l'ensemble noté*

$$\mathrm{vect}(P) = \Big\{ \sum \lambda_i x_i \ (\textit{finies}) \,/\, \forall i,\ \lambda_i \in \mathbb{K},\ x_i \in P \Big\}$$

est un sous-espace vectoriel de E, appelé sous-espace vectoriel engendré par la partie P ; si on veut insister sur le fait que les coefficients sont pris dans le corps $\mathbb{K}$, on note aussi $\mathrm{vect}_{\mathbb{K}}(P)$ cet ensemble. De même, on note

$$\mathrm{vect}(v_i)_{i\in I} = \Big\{ \sum \lambda_i v_i \ (\textit{finies}) \,/\, \forall i,\ \lambda_i \in \mathbb{K} \Big\}$$

le sous-espace vectoriel engendré par la partie $\{v_i\}_{i\in I}$.

Il est en effet clair que $0_E \in \mathrm{vect}(P)$ et que $\mathrm{vect}(P)$ est stable par combinaisons linéaires, donc c'est bien un sous-espace vectoriel.

2.1.2. Propriété. *L'ensemble $\mathrm{vect}(P)$ est le plus petit sous-espace vectoriel $S \subset E$ contenant la partie P.*

En effet, $\mathrm{vect}(P)$ contient bien P (car $x = 1 \cdot x \in \mathrm{vect}(P)$ pour tout $x \in P$), et d'autre part, si un sous-espace vectoriel S contient P, il doit nécessairement contenir toute combinaison linéaire de vecteurs de P, de sorte que $S \supset \mathrm{vect}(P)$.

2.1.3. Remarque. Il est facile de voir que pour une famille de sous-espaces $(S_i)_{i\in I}$ de E, on a par définition

$$\sum_{i\in I} S_i = \mathrm{vect}\Big(\bigcup_{i\in I} S_i\Big).$$

2.2. Familles génératrices

2.2.1. Définition. *Soit E un $\mathbb{K}$-espace vectoriel. Une famille $(v_i)_{i\in I}$ de vecteurs de E est dite génératrice si $\mathrm{vect}(v_i)_{i\in I} = E$, autrement dit, si tout vecteur $x \in E$ peut s'écrire comme une combinaison linéaire finie*

$$x = \sum \lambda_i v_i \quad \textit{avec } \lambda_i \in \mathbb{K}.$$

On dit aussi alors que E est engendré par $(v_i)_{i\in I}$.

2.2.2. Exemples.

(a) La famille vide $\emptyset$ engendre l'espace vectoriel nul $\{0\}$.

(b) Les n vecteurs $(1, 0, \ldots, 0)$, $\ldots$, $(0, \ldots, 0, 1)$ engendrent $\mathbb{K}^n$.

(c) $\mathbb{K}[X]$ est engendré par la famille infinie $(X^n)_{n\in\mathbb{N}}$.

(d) Par définition, une famille quelconque $(v_i)_{i\in I}$ est toujours une famille génératrice du sous-espace $E' = \mathrm{vect}(v_i)_{i\in I}$.

2.3. Familles liées, familles libres

Soit E un $\mathbb{K}$-espace vectoriel.

2.3.1. Définition. *On dit qu'une famille $(v_i)_{i\in I}$ de vecteurs de E est liée, ou encore $\mathbb{K}$-linéairement dépendante, s'il existe une famille de scalaires $(\lambda_i)_{i\in I}$* non tous nuls (*mais tout de même presque tous nuls si I est infini*) *tels que*

$$\sum_{i\in I} \lambda_i v_i = 0.$$

En écriture formalisée, cette propriété s'écrit $(v_i)_{i\in I}$ liée si et seulement si

$$\exists (\lambda_i)_{i\in I} \in \mathbb{K}^I \text{ presque tous nuls}, \quad \sum_{i\in I} \lambda_i v_i = 0 \quad \text{et} \quad \exists i \in I,\ \lambda_i \neq 0.$$

La notion de famille libre correspond juste à la négation de cette propriété*.

2.3.2. Définition. *On dit qu'une famille $(v_i)_{i\in I}$ de vecteurs de E est libre, ou encore $\mathbb{K}$-linéairement indépendante, si on a l'implication*

$$\forall (\lambda_i)_{i\in I} \in \mathbb{K}^I \textit{ presque tous nuls}, \quad \sum_{i\in I} \lambda_i v_i = 0 \Rightarrow \forall i \in I,\ \lambda_i = 0.$$

2.3.3. Exemples.

(a) Une famille contenant 0_E est toujours liée (puisque $1 \cdot 0_E = 0_E$), autrement dit, une famille libre est nécessairement composée de vecteurs non nuls.

(b) Une famille contenant deux vecteurs identiques n'est *jamais* libre, puisqu'on peut écrire $1 \cdot v + (-1) \cdot v = 0$.

(c) La famille de fonctions (f_1, f_2, f_3) de $\mathbb{R}$ dans $\mathbb{R}$ telle que

$$f_1(x) = 1, \quad f_2(x) = \cos(2x), \quad f_2(x) = \cos^2(x)$$

n'est pas libre (*pourquoi* ?)

La proposition suivante donne une caractérisation nécessaire et suffisante pour qu'une famille $(v_i)_{i\in I}$ soit liée.

---

* La négation d'une proposition $(\forall x,\ P(x))$ est $(\exists x,\ \text{non } P(x))$, et de même, la négation de $(\exists x,\ P(x))$ est $(\forall x,\ \text{non } P(x))$. Plus généralement, la négation d'une proposition "imbriquée"

$$\forall x,\ \exists y,\ \forall z,\ \exists w,\ P(x,y,z,w)$$

(où $\exists y$, peut-être interprété comme "il existe y tel que"), est obtenue en écrivant

$$\exists x,\ \forall y,\ \exists z,\ \forall w,\ \text{non } P(x,y,z,w).$$

D'autre part, (non (P et Q)) équivaut à (non P ou non Q), et $P \Rightarrow Q$ équivaut à (non P ou Q). On prend ici P : $\sum_{i\in I} \lambda_i v_i = 0$ et Q : $\exists i \in I,\ \lambda_i \neq 0$ et on observe que non(P et Q) est la même chose que $P \Rightarrow$ non Q.

2.3.4. Caractérisation de la dépendance linéaire. *Une famille $(v_i)_{i\in I}$ est liée si et seulement s'il existe un indice i_0 tel que v_{i_0} soit combinaison linéaire des autres vecteurs : il existe des coefficients $\mu_i \in \mathbb{K}$ (presque tous nuls) tels que*

$$v_{i_0} = \sum_{i\neq i_0} \mu_i v_i.$$

Démonstration. Si $v_{i_0} = \sum_{i\neq i_0} \mu_i v_i$, on obtient la combinaison linéaire $\sum \lambda_i v_i = 0$ avec $\lambda_i = \mu_i$ pour $i \neq i_0$ et $\lambda_i = -1$, de sorte que les λ_i sont non tous nuls. Réciproquement, si $\sum \lambda_i v_i = 0$ avec $\lambda_{i_0 \leqslant 0}$, on peut transposer le terme $\lambda_{i_0} v_{i_0}$ et diviser par λ_{i_0}, ce qui donne $v_{i_0} = \sum_{i\neq i_0} \mu_i v_i$ avec $\mu_i = -\lambda_i/\lambda_{i_0}$. □

2.3.5. Exemple. Dans le $\mathbb{R}$-espace vectoriel $E = \mathbb{R}^4$, la famille

$$v_1 = (1,1,0,0), \quad v_2 = (1,1,2,-2), \quad v_3 = (0,0,-1,1)$$

est liée, car on voit que $v_2 = v_1 - 2v_3$. Il en résulte que

$$\mathrm{vect}(v_1, v_2, v_3) = \mathrm{vect}(v_1, v_3),$$

puisque v_2 est "inutile" dans les combinaisons linéaires $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3$.

2.4. Bases et dimension

Dans toute la suite, E désigne un $\mathbb{K}$-espace vectoriel.

2.4.1. Définition. *On dit qu'une famille de vecteurs $(e_i)_{i\in I}$ est une* base *de E si celle-ci est à la fois libre et génératrice. On appelle dimension de E, noté* $\dim E$ *(ou $\dim_{\mathbb{K}} E$ si on veut éviter toute ambiguïté sur le corps $\mathbb{K}$) le nombre d'éléments des bases :*

$$\dim E = \mathrm{card}\, I$$

(on montrera en effet plus loin que toutes les bases ont le même nombre d'éléments, du moins lorsque celui-ci est fini).

2.4.2. Caractérisation. *Dire que $(e_i)_{i\in I}$ est une base revient à dire que tout vecteur x de E s'écrit de manière* unique *comme combinaison linéaire $x = \sum_{i\in I} x_i e_i$ d'éléments de la famille, $x_i \in \mathbb{K}$ (la somme n'ayant qu'un nombre fini de termes $x_i \neq 0$).*

Démonstration. L'existence des coefficients x_i résulte du fait que $(e_i)_{i\in I}$ est génératrice, et l'unicité résulte par différence du fait que $(e_i)_{i\in I}$ est libre. On dit que les $x_i \in \mathbb{K}$ sont les coordonnées du vecteur x dans la base $(e_i)_{i\in I}$. □

2.4.3. Exemples.

(a) La famille vide $\emptyset$ est une base de l'espace vectoriel nul $\{0\}$, on a $\dim\{0\} = 0$.

(b) La famille de vecteurs $(1,0,\dots,0)$, … , $(0,\dots,0,1)$ forme une base de $\mathbb{K}^n$, appelée *base canonique*, et on a $\dim \mathbb{K}^n = n$.

(c) La famille $(1, X, \dots, X^n)$ forme une base de l'espace vectoriel $\mathbb{K}[X]_n$ des polynômes à coefficients dans $\mathbb{K}$ de degré au plus n. On a donc $\dim_{\mathbb{K}} \mathbb{K}[X]_n = n+1$.

(d) Plus généralement, si $P_j \in \mathbb{K}[X]$ est un polynôme de degré exactement j (c'est-à-dire de coefficient de X^j non nul), alors $(P_0, P_1, \dots, P_n)$ est une base de $\mathbb{K}[X]_n$. On démontre en effet facilement par récurrence sur n qu'il s'agit d'une famille libre, respectivement d'une famille génératrice.

(e) Si $(v_i)_{i\in I}$ est une famille libre de E, alors c'est une base du sous-espace $S = \mathrm{vect}(v_i)_{i\in I}$, et on a donc $\dim S = \mathrm{card}\, I$.

En vue de démontrer les théorèmes fondamentaux sur l'existence de bases, il est commode de démontrer d'abord les deux lemmes suivants (en mathématiques, un "lemme" est une proposition préparatoire, établie de manière préliminaire pour démontrer ensuite un énoncé plus important).

2.4.4. Lemme. *Soit $(\ell_1, \dots, \ell_s)$ une famille libre de E. Alors, pour tout $x \in E$, la famille $(\ell_1, \dots, \ell_s, x)$ est liée si et seulement si $x \in \mathrm{vect}(\ell_1, \dots, \ell_s)$.*

Démonstration. Si $x \in \mathrm{vect}(\ell_1, \dots, \ell_s)$, on sait que la famille $(\ell_1, \dots, \ell_s, x)$. Mais réciproquement, si $(\ell_1, \dots, \ell_s, x)$ est liée, avec disons $\sum_{1\leqslant i\leqslant s} \lambda_i \ell_i + \lambda_{s+1} x = 0$ et les λ_i non tous nuls, on a $\lambda_{s+1} \neq 0$ sinon $(\ell_1, \dots, \ell_s)$ serait liée. Par conséquent $x = \sum_{1\leqslant i\leqslant s} \mu_i \ell_i$ avec $\mu_i = -\lambda_i/\lambda_{s+1}$, et donc $x \in \mathrm{vect}(\ell_1, \dots, \ell_s)$. □

2.4.5. Lemme. *Supposons que $(\ell_1, \dots, \ell_s)$ soit une famille libre de E et $(g_1, \dots, g_p)$ une famille génératrice de E. Alors nécessairement $s \leqslant p$.*

Démonstration. C'est le lemme le plus délicat – on le démontre par récurrence sur p.

• Cas $p = 0$: la famille génératrice est vide, c'est-à-dire que $E = \{0\}$. Dans ce cas, puisque'une partie libre ne peut contenir de vecteur nul, on doit aussi avoir $s = 0$.

• Cas $p = 1$: il nous faut alors démontrer que $s \leqslant 1$.

Supposons au contraire $s \geqslant 2$. Comme (g_1) est génératrice, il existe des coefficients $\lambda_1, \lambda_2 \in \mathbb{K}$ tels que

$$\ell_1 = \lambda_1 g_1, \quad \ell_2 = \lambda_2 g_1,$$

et comme $\ell_i \neq 0$, on a nécessairement $\lambda_i \neq 0$. Mais il vient alors $\lambda_2 \ell_1 - \lambda_1 \ell_2 = 0$ et (ℓ_1, ℓ_2) serait liée, ce qui est contradictoire. Par conséquent on a bien $s \leqslant 1$.

• Supposons le résultat déjà démontré pour $p-1$ (avec $p \geqslant 1$), et démontrons-le pour p.

Comme $(g_1, \dots, g_p)$ est génératrice, il existe des coefficients $\lambda_{ij} \in \mathbb{K}$ tels que

$$\begin{cases} \ell_1 = \lambda_{11} g_1 + \cdots + \lambda_{1p} g_p \\ \cdots \\ \ell_i = \lambda_{i1} g_1 + \cdots + \lambda_{ip} g_p, \quad 1 \leqslant i \leqslant s \\ \cdots \\ \ell_s = \lambda_{s1} g_1 + \cdots + \lambda_{sp} g_p \end{cases}$$

Comme $\ell_s \leqslant 0$, l'un des coefficients λ_{sj} est non nul, et comme on peut toujours permuter les g_j dans ce raisonnement, il n'est pas restrictif de supposer $\lambda_{sp} \neq 0$. L'idée est d'éliminer g_p par combinaisons linéaires en calculant

$$\begin{cases} \ell'_1 = \ell_1 - \frac{\lambda_{1p}}{\lambda_{sp}}\ell_s = \lambda'_{11}g_1 + \cdots + \lambda'_{1p-1}g_{p-1}, \\ \cdots \\ \ell'_i = \ell_i - \frac{\lambda_{ip}}{\lambda_{sp}}\ell_s = \lambda'_{i1}g_1 + \cdots + \lambda'_{ip-1}g_{p-1}, \quad 1 \leqslant i \leqslant s-1. \end{cases}$$

On considère l'espace vectoriel $E' = \mathrm{vect}(g_1, \ldots, g_{p-1})$, qui contient $\ell'_1, \ldots, \ell'_{s-1}$ d'après ce qui précède. or $(\ell'_1, \ldots, \ell'_{s-1})$ est encore une famille libre, car s'il existait une combinaison linéaire $\mu_1\ell'_1 + \cdots + \mu_{s-1}\ell'_{s-1} = 0$ à coefficients $\mu_j \in \mathbb{K}$ non tous nuls, on aurait

$$0 = \mu_1\ell'_1 + \cdots + \mu_{s-1}\ell'_{s-1} = \mu_1\ell_1 + \cdots + \mu_{s-1}\ell_{s-1} + \alpha\ell_s$$

pour un certain coefficient $\alpha \in \mathbb{K}$ (inutile à expliciter), ce qui est contradictoire. L'hypothèse de récurrence appliquée à l'espace vectoriel E' entraîne alors $s-1 \leqslant p-1$, et donc on a bien $s \leqslant p$. □

2.4.6. Théorème de la dimension. *Il y a équivalence entre les propriétés* (a) *et* (b) *suivantes :*

(a) *Le nombre s d'éléments des parties libres $(\ell_1, \ldots, \ell_s)$ de E est borné ;*

(b) *E possède une famille génératrice finie $(g_1, \ldots, g_p)$.*

Si (a) *ou* (b) *est vérifié, alors E possède des bases ayant toutes un même nombre d'éléments $n = \dim E$. De plus*

(c) *Le nombre s d'éléments des familles libres $(\ell_1, \ldots, \ell_s)$ de E vérifie $s \leqslant n$ et il y a égalité si et seulement si $(\ell_1, \ldots, \ell_s)$ est une base.*

(d) *Le nombre p d'éléments des familles génératrices $(g_1, \ldots, g_p)$ de E vérifie $p \geqslant n$ et il y a égalité si et seulement si $(g_1, \ldots, g_p)$ est une base.*

Lorsque les propriétés (a) et/ou (b) ne sont pas vérifiées, on dit que E est de dimension infinie, et on écrit parfois $\dim E = +\infty$.

Démonstration. Le lemme 2.4.5 montre que $s \leqslant p$, donc (b) implique (a). Réciproquement si s est borné, soit n le maximum et $(\ell_1, \ldots, \ell_n)$ une famille libre maximale. Alors, pour tout $x \in E$, la famille $(\ell_1, \ldots, \ell_n, x)$ est liée, et le lemme 2.4.4 montre que $x \in \mathrm{vect}(\ell_1, \ldots, \ell_n)$. Par conséquent $(\ell_1, \ldots, \ell_n)$ est une famille génératrice (et donc une base), et on voit que (a) implique (b).

On a vu en passant que si une partie libre $(\ell_1, \ldots, \ell_s)$ atteint le nombre maximum n d'éléments, alors c'est une base, d'où la propriété (c). De même si une partie génératrice $(g_1, \ldots, g_p)$ possède le nombre minimal $p = n$ d'éléments, elle est nécessairement libre, sinon l'un des g_i serait combinaison linéaire des autres et on pourrait diminuer p. La propriété (d) est démontrée, et avec elle, le théorème 2.4.6. □

2.4.7. Théorème de la base incomplète. *On suppose que $(\ell_1, \ldots, \ell_s)$ est une famille libre de E et $(g_1, \ldots, g_p)$ une famille génératrice de E. Alors il existe une*

base de E de la forme

$$(\ell_1, \dots, \ell_s, g_{i_1}, \dots, g_{i_q})$$

obtenue en complétant la famille libre par certains éléments bien choisis de la famille génératrice.

Démonstration. Posons $E_0 = \mathrm{vect}(\ell_1, \dots, \ell_s) \subset E$. Si tous les éléments g_i sont combinaisons linéaires des ℓ_j, i.e. $g_i \in E_0$, alors $E = \mathrm{vect}(g_1, \dots, g_p) \subset E_0$, donc $E_0 = E$ et $(\ell_1, \dots, \ell_s)$ est une base de E. On prend $q = 0$ (aucun vecteur g_i à ajouter).

Sinon, il existe un vecteur $g_{i_1} \notin E_0$. On pose alors $E_1 = \mathrm{vect}(\ell_1, \dots, \ell_s, g_{i_1})$ qui contient strictement E_0. Si $E_1 = E$, $(\ell_1, \dots, \ell_s, g_{i_1})$ est une base de E et on a terminé avec $q = 1$. Si $E_1 \neq E$, il existe un vecteur $g_{i_2} \notin E_1$, nécessairement distinct de g_{i_1}, et on pose $E_2 = \mathrm{vect}(\ell_1, \dots, \ell_s, g_{i_1}, g_{i_2})$. On continue ainsi en prenant $g_{i_q} \notin E_{q-1}$, jusqu'à ce que

$$E_q = \mathrm{vect}(\ell_1, \dots, \ell_s, g_{i_1}, \dots, g_{i_q}) \supset E_{q-1}$$

contienne tous les vecteurs g_i (le processus a une fin puisque les g_{i_k} sont nécessairement 2 à 2 distincts et qu'il n'y a qu'un nombre fini de vecteurs g_i à ajouter). Dans ce cas $E = \mathrm{vect}(g_1, \dots, g_p) \subset E_q$ et donc $E_q = E$. Nous prétendons alors que $(\ell_1, \dots, \ell_s, g_{i_1}, \dots, g_{i_q})$ est une base de E. Sinon, on aurait une combinaison linéaire à coefficients non tous nuls

$$\lambda_1 \ell_1 + \cdots + \lambda_s \ell_s + \mu_1 g_{i_1} + \cdots + \mu_r g_{i_r} = 0, \quad \text{pour un certain } r \leqslant q.$$

Puisque $(\ell_1, \dots, \ell_s)$ est libre, les coefficients μ_j sont non tous nuls. On s'arrête à l'indice r maximum tel que $\mu_r \neq 0$. Dans ce cas g_{i_r} est combinaison linéaire de $\ell_1, \dots, \ell_s, g_{i_1}, \dots, g_{i_{r-1}}$, i.e. $g_{i_r} \in E_{r-1}$, ce qui contredit notre construction et termine la preuve de 2.4.7. □

2.4.8. Corollaire. *Si E possède une partie génératrice finie $(g_1, \dots, g_p)$, on peut en extraire une base $(g_{i_1}, \dots, g_{i_n})$ de E.*

(Cas particulier de 2.4.7 où la famille libre $(\ell_1, \dots, \ell_s)$ est vide, i.e. $s = 0$). □

2.4.9. Corollaire. *Si E est dimension finie, on peut compléter toute partie libre $(\ell_1, \dots, \ell_s)$ en une base $(\ell_1, \dots, \ell_n)$ de E.*

(Appliquer 2.4.7 avec une famille génératrice quelconque). □

2.4.10. Remarque. Si E n'est pas de dimension finie, on peut encore démontrer l'existence de bases infinies $(e_i)_{i \in I}$; il suffit pour cela de prendre une famille libre maximale, ce qui veut dire qu'on ne peut plus rajouter de nouveaux vecteurs à $(e_i)_{i \in I}$ sans que la famille devienne liée (l'existence d'une telle famille maximale est une conséquence de "l'axiome du choix" en théorie des ensembles, et n'a rien à voir avec l'algèbre linéaire).

On suppose maintenant que E est de dimension finie n. Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base de E. Alors, pour tout $x \in E$, on peut écrire de manière unique

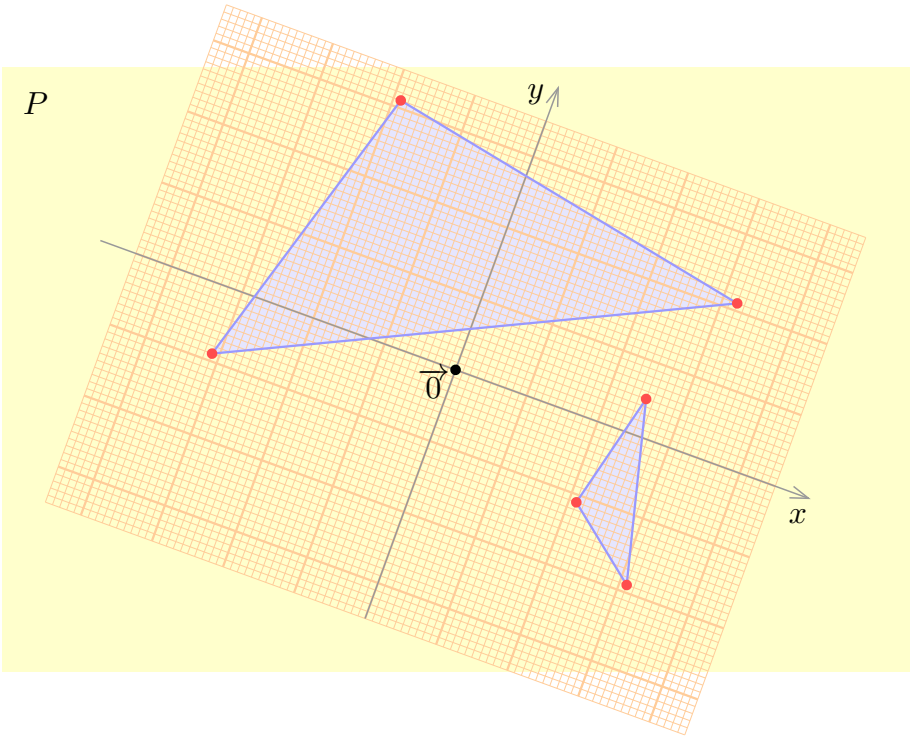
$$x = x_1 e_1 + \ldots + x_n e_n, \qquad x_i \in \mathbb{K}.$$

La matrice colonne

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$$

s'appelle la matrice des *coordonnées* de x dans la base $\mathcal{B}$.

Il convient de distinguer soigneusement le vecteur x de la matrice X qui le représente (et qui d'ailleurs dépend de la base $\mathcal{B}$ choisie). On peut imaginer ainsi un plan vectoriel P constitué d'objets géométriques (des vecteurs), les coordonnées associées au choix d'une base étant obtenues en venant glisser une feuille de papier millimétré transparent sur le plan P (le fait d'avoir un plan vectoriel impose que l'origine soit placée exactement au dessus du vecteur $\overrightarrow{0}$). Bien entendu, les coordonnées obtenues dépendent de la position de la feuille de papier millimétré.



2.4.11. Attention. Une base est une famille ordonnée ! Si on change l'ordre des vecteurs, on obtient une nouvelle base, et les coordonnées x_i sont permutées.

2.4.12. Espaces vectoriels sur $\mathbb{R}$ et sur $\mathbb{C}$. Soit E un espace vectoriel sur $\mathbb{C}$ de dimension finie $n = \dim_{\mathbb{C}} E$ et $(e_1, \ldots, e_n)$ une base de E sur $\mathbb{C}$. Tout vecteur $z \in E$ peut alors s'écrire de manière unique

$$z = z_1 e_1 + \ldots + z_n e_n, \quad z_j \in \mathbb{C},$$

et chaque complexe z_j peut s'écrire de manière unique $z_j = x_j + iy_j$, $x_j, y_j \in \mathbb{R}$, ce qui donne

$$\begin{aligned} z &= (x_1 + iy_1)e_1 + \ldots + (x_n + iy_n)e_n \\ &= x_1 e_1 + y_1(ie_1) + \ldots + x_n e_n + y_n(ie_n), \end{aligned}$$

et cette écriture en termes de coefficients réels x_j, y_j est unique. Ceci montre que $(e_1, ie_1, e_2, ie_2, \ldots, e_n, ie_n)$ est une base de E comme $\mathbb{R}$-espace vectoriel, et on en déduit

$$\dim_{\mathbb{R}} E = 2n = 2 \dim_{\mathbb{C}} E. \tag{2.4.13}$$

On peut réinterpréter ce résultat en observant qu'on a une composée de bijections

$$\begin{array}{ccccc} E & \longrightarrow & \mathbb{C}^n & \longrightarrow & \mathbb{R}^{2n} \\ z & \longmapsto & (z_1, \ldots, z_n) & \longmapsto & (x_1, y_1, \ldots, x_n, y_n). \end{array}$$

2.5. Rang d'une famille de vecteurs.

Soit E un $\mathbb{K}$-espace vectoriel.

2.5.1. Définition. *Le rang d'une famille de vecteurs $(v_i)_{i\in I}$ est par définition la dimension du sous-espace vectoriel engendré (sur le corps $\mathbb{K}$ considéré) :*

$$\operatorname{rang}_{\mathbb{K}}(v_j)_{i\in I} = \dim_{\mathbb{K}} \operatorname{vect}_{\mathbb{K}}(v_j)_{i\in I}$$

(et on omet très souvent les indices $\mathbb{K}$ s'il n'y a pas ambiguïté).

Le corollaire 2.4.8 implique la propriété suivante.

2.5.2. Propriété. *Si $(v_1, \ldots, v_p)$ est de rang r, on peut extraire une sous-famille $(v_{i_1}, \ldots, v_{i_r})$ qui est une base du sous-espace $S = \operatorname{vect}(v_1, \ldots, v_p)$.*

2.5.3. Exemple. On se place dans le $\mathbb{C}$-espace vectoriel $E = \mathbb{C}^2$ et on considère les 3 vecteurs

$$v_1 = (1, 0), \quad v_2 = (0, 1), \quad v_3 = (i, -i).$$

Sur $\mathbb{C}$, (v_1, v_2) est une base de E et $v_3 = ie_1 - ie_2$. Par conséquent

$$\operatorname{rang}_{\mathbb{C}}(v_1, v_2, v_3) = 2 = \dim_{\mathbb{C}} E.$$

Mais E est aussi un $\mathbb{R}$-espace vectoriel de dimension réelle 4, et pour tous scalaires $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ il vient

$$\begin{aligned} \lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 &= (\lambda_1 + i\lambda_3, \lambda_2 - i\lambda_2) = (0, 0) \\ &\Rightarrow \quad \lambda_1 + i\lambda_3 = \lambda_2 - i\lambda_3 = 0 \quad \Rightarrow \quad \lambda_1 = \lambda_2 = \lambda_3 = 0. \end{aligned}$$

Donc (v_1, v_2, v_3) est une famille $\mathbb{R}$-linéairement indépendante et

$$\operatorname{rang}_{\mathbb{R}}(v_1, v_2, v_3) = 3 < \dim_{\mathbb{R}} E.$$

2.6. Sommes directes de sous-espaces vectoriels.

Soient $S_1, \ldots, S_N$ des sous-espaces vectoriels d'un $\mathbb{K}$-espace vectoriel E. Rappelons que la *somme* des sous-espaces $S_1, \ldots, S_N$ est le sous-espace vectoriel S de E défini par

$$S = S_1 + \ldots + S_N = \big\{x = x_1 + \cdots + x_N \,/\, x_i \in S_i\big\}.$$

Supposons tous les sous-espaces S_i de dimension finie, avec $p_i = \dim_{\mathbb{K}} S_i$. Si on prend une base $\mathcal{B}_i = (e_{i1}, \ldots, e_{ip_i})$ de S_i, on peut écrire

$$x_i = x_{i1}e_{i1} + \cdots + x_{ip_i}e_{ip_i} = \sum_{j=1}^{p_i} x_{ij}e_{ij} \tag{2.6.1}$$

et par conséquent

$$x = \sum_{j=1}^{N} x_i = \sum_{j=1}^{N}\sum_{j=1}^{p_i} x_{ij}e_{ij}, \tag{2.6.2}$$

ce qui montre que $\mathcal{G} = (e_{ij})_{i,j}$ est une famille génératrice de S. On a donc en général $\dim_{\mathbb{K}} S \leqslant \operatorname{card} \mathcal{G} = p_1 + \cdots + p_N$, c'est-à-dire

$$\dim_{\mathbb{K}} S = \dim_{\mathbb{K}}(S_1 + \ldots + S_N) \leqslant \dim_{\mathbb{K}} S_1 + \ldots + \dim_{\mathbb{K}} S_N.$$

L'inégalité est stricte s'il on prend par exemple pour E un plan vectoriel réel, et $S_1 = D_1$, $S_2 = D_2$, $S_3 = D_3$ trois droites deux à deux distinctes de ce plan. Dans ce cas, si e_i est un vecteur directeur de D_i, on voit que (e_1, e_2, e_3) est un système générateur de $E = D_1 + D_2 + D_3$, mais ce n'est pas une famille libre (donc pas une base) et on a sur $\mathbb{K} = \mathbb{R}$

$$\dim(D_1 + D_2 + D_3) = 2 < 3 = \dim D_1 + \dim D_2 + \dim D_3.$$

2.6.3. Théorème et définition. *Soit E un $\mathbb{K}$-espace vectoriel, $S_1 \ldots, S_N$ des sous-espaces et $S = \sum_{1\leqslant i\leqslant N} S_i$. Les deux propriétés suivantes sont équivalentes :*

(i) (*absence de relation linéaire entre vecteurs non nuls des sous-espaces S_i*)

$$\forall x_i \in S_i, \quad x_1 + \cdots + x_N = 0 \quad \Rightarrow \quad x_1 = \cdots = x_N = 0\,;$$

(ii) (*unicité de la décomposition d'un vecteur de la somme S*)

$$\textit{si}\ \ x = \sum_{i=1}^{N} x_i = \sum_{i=1}^{N} x'_i \ \ \textit{avec}\ x_i, x'_i \in S_i,\ \textit{alors}\ \ \forall i \in \{1, \ldots, N\}, \quad x_i = x'_i.$$

On dit alors que $S_1, \ldots, S_N$ sont en somme directe *et on écrit*

$$S = S_1 \oplus \cdots \oplus S_N \quad \textit{ou encore} \quad S = \bigoplus_{i=1}^{N} S_i.$$

Démonstration. Il est clair que (ii) ⇒ (i) puisque le vecteur nul s'écrit $0 = 0+\cdots+0$, et que l'unicité de la décomposition du vectur 0 donne bien la propriété (i). Réciproquement, si la propriété (i) est vérifiée et si on a deux décompositions d'un vecteur $x \in S$ comme dans (ii), on peut écrire $\sum_{i=1}^{N}(x'_i - x_i) = 0$ avec $x'_i - x_i \in S_i$, donc $x'_i - x_i = 0$ pour tout i, ce qui prouve l'unicité et donc (ii). □

Maintenant, si $\mathcal{B}_i = (e_{i1}, \ldots, e_{ip_i})$ est une base de S_i, l'unicité de la décomposition en composantes x_i dans (2.6.2) montre que les coefficients x_{ij} sont uniques, par conséquent $(e_{ij})_{i,1\leqslant j\leqslant p_i}$ est une base de S et

$$S = S_1 \oplus \cdots \oplus S_N \text{ somme directe } \Rightarrow \dim S = \dim S_1 + \cdots + \dim S_N.$$

Il suffit qu'il existe une seule relation linéaire non triviale entre les $(e_{ij})_{i,1\leqslant j\leqslant p_i}$ pour que l'on ait au contraire $\dim S < \dim S_1 + \cdots + \dim S_N$. En résumé :

2.6.4. *Pour que des sous-espaces $S_1, \ldots, S_N$ de dimension finie soient en* somme directe*, il faut et il suffit que*

$$\dim(S_1 + \cdots + S_N) = \dim S_1 + \cdots + \dim S_N.$$

Si c'est le cas, on obtient une base de $S = S_1 \oplus \cdots \oplus S_N$ en concaténant des bases $\mathcal{B}_i = (e_{i1}, \ldots, e_{ip_i})$ des S_i.

2.6.5. Remarque. Pour deux sous-espaces S_1, S_2, la somme $S_1 \oplus S_2$ est directe si et seulement si $S_1 \cap S_2 = \{0\}$. En effet si $u \in S_1 \cap S_2$ avec $u \neq 0$, on obtient deux décompositions du vecteur nul $0 = 0 + 0 = u + (-u)$, et réciproquement si $x_1 + x_2 = 0$ avec $x_1 \in S_1$ et $x_2 \in S_2$, alors $x_1 = -x_2 \in S_1 \cap S_2$, donc $S_1 \cap S_2 = \{0\}$ implique $x_1 = x_2 = 0$.

2.6.6. Remarque. Il n'est pas vrai en revanche pour N sous-espaces, $N \geqslant 3$, que la condition $S_i \cap S_j = \{0\}$, pour $i \neq j$, garantisse le fait que la somme $S = \sum S_i$ soit directe. De nouveau, il suffit de prendre 3 droites deux à deux distinctes D_1, D_2, D_3 d'un plan vectoriel E pour obtenir un contre-exemple.

2.6.7. Supplémentaires. Si S est un sous-espace de E et $(e_i)_{i\in I}$ une base de S, le théorème de la base incomplète dit qu'on peut trouver $J \subset I$ tel que $(e_i)_{i\in J}$ soit une base de E (on l'a vu en dimension finie, mais en fait ceci est vrai aussi en dimension infinie, il suffit de compléter en une famille libre maximale. Si on pose $T = \mathrm{vect}(e_i)_{i\in J\smallsetminus I}$, alors

$$E = S \oplus T.$$

On dit que T est un *supplémentaire* du sous-espace S. L'argument précédent montre qu'il existe toujours des sous-espaces supplémentaires d'un sous-space S donné.

3. Applications linéaires et matrices

3.1. Principales définitions

3.1.1. Définition. *Soient E, F deux $\mathbb{K}$-espaces vectoriels. Une application $\ell : E \to F$ est dite $\mathbb{K}$-linéaire si elle vérifie les deux propriétés suivantes :*

(i) *pour tous $x, y \in E$, $\ell(x+y) = \ell(x) + \ell(y)$;*

(ii) *pour tous $\lambda \in \mathbb{K}$ et $x \in E$, $\ell(\lambda x) = \lambda f(x)$.*

Il est équivalent de vérifier (i) et (ii) ou l'unique axiome suivant relatif à l'image de combinaisons linéaires :

(iii) pour tous $\lambda, \mu \in \mathbb{K}$, $x, y \in E$, $f(\lambda x + \mu y) = \lambda f(x) + \mu f(y)$.

Sous ces hypothèses, on a nécessairement $\ell(0) = 0$ (comme on le voit en faisant $x = y = 0$ dans l'axiome (i), $\lambda = 0$ dans l'axiome (ii) et $\lambda = \mu = 0$ dans l'axiome (iii).

3.1.2. Définition.

(i) *Le* noyau *de ℓ, noté* $\operatorname{Ker} \ell$, *est l'ensemble*

$$\operatorname{Ker} \ell = \ell(E) = \big\{x \in E\,;\ \ell(x) = 0\big\} \subset E.$$

C'est un sous-espace vectoriel de E.

(ii) L'image *de ℓ, notée* $\operatorname{Im} \ell$ *ou* $\ell(E)$, *est l'ensemble*

$$\operatorname{Im} \ell = \big\{y = \ell(x) \,/\, x \in E\big\} \subset F.$$

C'est un sous-espace vectoriel de F.

Si $(e_1, \ldots, e_n)$ est une base de E, tout vecteur $x \in E$ s'écrit $x = x_1 e_1 + \cdots + x_n e_n$, $x_j \in \mathbb{K}$, et l'image est donnée par

$$y = \ell(x) = x_1 \ell(e_1) + \ldots + x_n \ell(e_n).$$

On a donc

$$\operatorname{Im}(\ell) = \operatorname{vect}(\ell(e_1), \ldots, \ell(e_n)), \tag{3.1.3}$$

et il est naturel d'introduire la définition suivante.

3.1.4. Définition. *Le rang d'une application linéaire $\ell : E \to F$ est par définition*

$$\operatorname{rang}_{\mathbb{K}}(\ell) = \dim_{\mathbb{K}} \operatorname{Im} \ell.$$

Si $\mathcal{B} = (e_1, \ldots, e_n)$ est une base de E, on a aussi

$$\operatorname{rang}_{\mathbb{K}}(\ell) = \operatorname{rang}_{\mathbb{K}}(\ell(e_1), \ldots, \ell(e_n)),$$

et en particulier $\operatorname{rang}_{\mathbb{K}}(\ell(e_1), \ldots, \ell(e_n))$ ne dépend pas de la base $\mathcal{B}$ choisie.

3.1.5. Notation. *On notera $\mathcal{L}_{\mathbb{K}}(E; F)$ l'ensemble des applications $\mathbb{K}$-linéaires de E dans F (et on se permettra d'omettre le corps $\mathbb{K}$ s'il n'y a pas d'ambiguité possible).*

3.2. Théorème du rang

3.2.1. Théorème du rang. *Soient E, F deux $\mathbb{K}$-espaces vectoriels, et soit $\ell : E \to F$ une application linéaire. Si E est de dimension finie, alors* $\operatorname{Ker}\ell$ *et* $\operatorname{Im}\ell$ *sont de dimension finie et on a*

$$\dim_{\mathbb{K}} \operatorname{Ker}\ell + \dim_{\mathbb{K}} \operatorname{Im}\ell = \dim_{\mathbb{K}} E.$$

De façon équivalente, on a

$$\operatorname{rang}_{\mathbb{K}}(\ell) = \dim_{\mathbb{K}} E - \dim_{\mathbb{K}} \operatorname{Ker}\ell.$$

Démonstration. Comme $\operatorname{Ker}\ell$ est un sous-espace vectoriel de E, il est nécessairement de dimension finie. Soit $(a_1, \dots, a_p)$ une base de $\operatorname{Ker}\ell$, avec $p = \dim_{\mathbb{K}} \operatorname{Ker}\ell$. On la complète en une base $(a_1, \dots, a_n)$ de E, où $n = \dim_{\mathbb{K}} E \geqslant p$. On a $\ell(a_1) = \cdots = \ell(a_p) = 0$ puisque ces vecteurs a_i sont dans $\operatorname{Ker}\ell$. L'image $\operatorname{Im}\ell$ est par définition l'ensemble des vecteurs images $y = \ell(x_1a_1 + \cdots + x_na_n)$. Comme ℓ est linéaire, il vient

$$y = \ell(x_1a_1 + \cdots + x_na_n) = x_{p+1}\ell(a_{p+1}) + \cdots + x_n\ell(a_n) = \ell(x_{p+1}a_{p+1} + \cdots + x_na_n),$$

et on voit déjà que la famille $\mathcal{G} = (\ell(a_{p+1}), \dots, \ell(a_n))$ est une famille génératrice de $\operatorname{Im}\ell$. Montrons que c'est une base : il reste à voir que $\mathcal{G}$ est libre. Pour cela, supposons $y = 0$. Alors $x = x_{p+1}a_{p+1} + \cdots + x_na_n \in \operatorname{Ker}\ell$, et comme $(a_1, \dots, a_p)$ est une base de $\operatorname{Ker}\ell$, il existe des scalaires $\lambda_1, \dots, \lambda_p \in \mathbb{K}$ tels que

$$x = x_{p+1}a_{p+1} + \cdots + x_na_n = \lambda_1a_1 + \cdots + \lambda_pa_p.$$

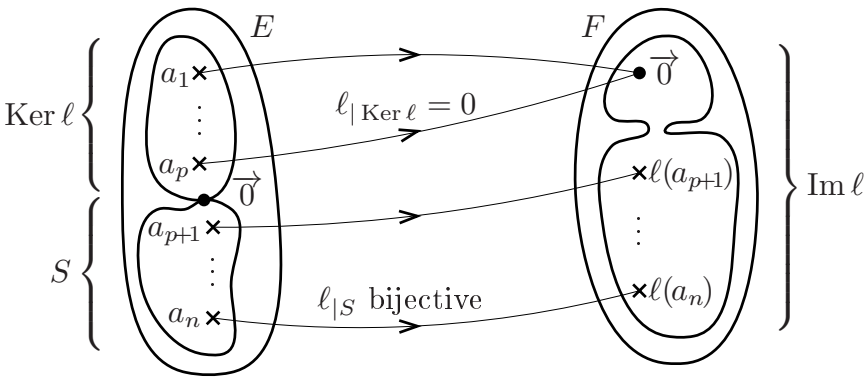
Maintenant, comme $(a_1, \dots, a_n)$ est une base de E, on en conclut par unicité de l'écriture de x que l'on a $x_{p+1} = \cdots = x_n = 0$ (et aussi $\lambda_1 = \cdots = \lambda_p = 0$). Ceci entraîne que $\mathcal{G}$ est bien libre. On a donc

$$\operatorname{rang}_{\mathbb{K}}(\ell) = \dim_{\mathbb{K}} \operatorname{Im}\ell = n - p = \dim_{\mathbb{K}} E - \dim_{\mathbb{K}} \operatorname{Ker}\ell. \qquad \square$$

3.2.2. Remarque complémentaire. Si $S = \operatorname{vect}(a_{p+1}, \dots, a_n)$, alors on a par construction la somme directe

$$E = \operatorname{Ker}\ell \oplus S,$$

et la restriction $\ell_{|S} : S \to \operatorname{Im}\ell$ est une bijection (“isomorphisme” d'espaces vectoriels, cf. plus loin), envoyant la base $(a_{p+1}, \dots, a_n)$ du sous-espace S sur la base $\mathcal{G} = (\ell(a_{p+1}), \dots, \ell(a_n))$ de $\operatorname{Im}\ell$.



Grâce au théorème du rang, il existe au moins deux méthodes pour calculer le rang d'une application linéaire ℓ : l'une d'elles est de prendre une base quelconque $(e_1,\dots,e_n)$ de E et de chercher une base de $\operatorname{Im}\ell$ en extrayant une sous-famille libre maximale $(\ell(e_{i_r}),\dots,\ell(e_{i_r}))$ de $(\ell(e_1),\dots,\ell(e_n))$ (pour cela, on peut chercher à éliminer les vecteurs $\ell(e_j)$ qui sont combinaisons des autres), ou bien on calcule $\operatorname{Ker}\ell=\{x\in E\,/\,\ell(x)=0\}$ et on en détermine une base $(a_1,\dots,a_p)$, pour aboutir à la valeur du rang $r=n-p$ de ℓ.

3.3. Représentation matricielle des applications linéaires

Soient E, F des $\mathbb{K}$-espaces vectoriels de dimensions $\dim E=p$, $\dim F=n$, et $\ell : E\to F$ une application linéaire.

3.3.1. Définition. *On suppose donnée des bases respectives $\mathcal{B}_E=(e_1,\dots,e_p)$ et $\mathcal{B}_F=(\varepsilon_1,\dots,\varepsilon_n)$ de E et F. La matrice $A=\operatorname{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(\ell)$ relativement à $\mathcal{B}_E,\mathcal{B}_F$ est par définition le tableau rectangulaire $A=(a_{ij})_{1\leqslant i\leqslant n,1\leqslant j\leqslant p}$ à coefficients $a_{ij}\in\mathbb{K}$, aussi écrit sous forme développée*

$$A=\operatorname{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(\ell)=\begin{array}{c}\begin{matrix}\ell(e_1) & \ell(e_2) & \dots & \ell(e_p)\end{matrix}\\ \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1p}\\ a_{21} & a_{22} & \dots & a_{2p}\\ \vdots & & \ddots & \vdots\\ a_{n1} & a_{n2} & \dots & a_{np}\end{pmatrix}\end{array}\begin{matrix}\varepsilon_1\\ \varepsilon_2\\ \vdots\\ \varepsilon_n\end{matrix},$$

où chaque colonne $(a_{ij})_{1\leqslant i\leqslant n}$ est la matrice des coordonnées du vecteur image $\ell(e_j)$ relativement à la base $(\varepsilon_1,\dots,\varepsilon_n)$ de F.

3.3.2. Définition. On notera $\mathcal{M}_{n\times p}(\mathbb{K})$ l'ensemble des matrices rectangulaires $n\times p$ à coefficients dans $\mathbb{K}$. C'est un $\mathbb{K}$-espace vectoriel de dimension np.

Si $A=\operatorname{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(\ell)\in\mathcal{M}_{n\times p}(\mathbb{K})$, on peut écrire par définition

$$\ell(e_j)=\sum_{i=1}^{n}a_{ij}\varepsilon_i,\quad 1\leqslant j\leqslant p. \tag{3.3.3}$$

Étant donné un vecteur $x=\sum_{j=1}^{p}x_je_j$ et $X=\begin{pmatrix}x_1\\ x_2\\ \vdots\\ x_p\end{pmatrix}$ sa matrice colonne de coordonnées, on voit par linéarité de ℓ que le vecteur image $y=\ell(x)$ est donné par

$$\begin{aligned}y=\ell(x)&=\ell\Big(\sum_{j=1}^{p}x_je_j\Big)=\sum_{j=1}^{p}x_j\ell(e_j)\\ &=\sum_{j=1}^{p}x_j\Big(\sum_{i=1}^{n}a_{ij}\varepsilon_i\Big)=\sum_{i=1}^{n}\Big(\sum_{j=1}^{p}a_{ij}x_j\Big)\varepsilon_i.\end{aligned}$$

On a donc $y = \sum_{i=1}^{n} y_i \varepsilon_i$ avec

$$y_i = \sum_{j=1}^{p} a_{ij} x_j, \quad 1 \leqslant i \leqslant n, \tag{3.3.4}$$

Compte tenu de la définition du produit des matrices, on voit que l'application linéaire ℓ se calcule comme suit.

3.3.5. Formule fondamentale. *L'application linéaire $\ell : E \to F$ se calcule matriciellement par*

$$x \mapsto y = \ell(x), \quad X \mapsto Y = AX$$

où X, Y sont les matrices colonnes de $x \in E$ et $y \in F$ relativement aux bases $\mathcal{B}_E$, $\mathcal{B}_F$, et où $A = \mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(\ell)$.

Composée d'applications linéaires. Supposons maintenant que l'on ait une composée d'applications linéaires

$$E \xrightarrow{u} F \xrightarrow{v} G$$

où E, F, G sont de dimension respectives p, n, m et munis de bases $\mathcal{B}_E = (e_k)_{1 \leqslant k \leqslant p}$, $\mathcal{B}_F = (\varepsilon_j)_{1 \leqslant j \leqslant n}$ et $\mathcal{B}_G = (\eta_i)_{1 \leqslant i \leqslant m}$. On note

$$A = \mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(u), \quad B = \mathrm{Mat}_{\mathcal{B}_F}^{\mathcal{B}_G}(v)$$

et on cherche à déterminer la matrice $C = \mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_G}(v \circ u)$ de la composée $v \circ u : E \to G$. On note

$$E \ni x \overset{u}{\longmapsto} y = u(x) \overset{v}{\longmapsto} z = v(y) = v(u(x)) \in G$$

et $X = (x_k), Y = (y_j), Z = (z_i)$ les matrices colonnes de x, y, z. On a alors les formules

$$y_j = \sum_{k=1}^{p} a_{jk} x_k, \quad z_i = \sum_{j=1}^{n} b_{ij} y_j \quad \Longrightarrow$$

$$z_i = \sum_{j=1}^{n} \left(b_{ij} \sum_{k=1}^{p} a_{jk} x_k \right) = \sum_{k=1}^{p} \left(\sum_{j=1}^{n} b_{ij} a_{jk} \right) x_k,$$

autrement dit, par l'unicité de la représentation en coordonnées, on a $z_i = \sum_{k=1}^{p} c_{ik} x_k$ avec

$$c_{ik} = \sum_{j=1}^{n} b_{ij} a_{jk}, \quad 1 \leqslant i \leqslant m, \ 1 \leqslant k \leqslant p. \tag{3.3.6}$$

Par définition, ceci correspond au produit $C = B \times A$ des matrices A, B. On peut résumer comme suit le résultat trouvé :

3.3.7. Théorème. *La matrice d'une composée $v\circ u : E \xrightarrow{u} F \xrightarrow{v} G$ d'applications linéaires relativement à des bases $\mathcal{B}_E$, $\mathcal{B}_F$, $\mathcal{B}_G$ de E, F, G est donnée par*

$$\mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_G}(v\circ u) = \mathrm{Mat}_{\mathcal{B}_F}^{\mathcal{B}_G}(v) \times \mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(u),$$

où le produit des matrices définit une opération

$$\mathcal{M}_{m\times n}(\mathbb{K}) \times \mathcal{M}_{n\times p}(\mathbb{K}) \longrightarrow \mathcal{M}_{m\times p}(\mathbb{K}), \quad (B, A) \longmapsto C = B \times A.$$

3.4. Changement de base

Soit E un $\mathbb{K}$ espace vectoriel de dimension finie n. On suppose qu'on s'est donné une première base $cB = (e_1, \ldots, e_n)$, puis une "nouvelle base" $cB' = (e'_1, \ldots, e'_n)$. Soit $x \in E$ est un vecteur quelconque et

$$X = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad X' = \begin{pmatrix} x'_1 \\ \vdots \\ x'_n \end{pmatrix}$$

les coordonnées correspondantes relativement à $\mathcal{B}, \mathcal{B}'$, de sorte que

$$x = \sum_{j=1}^{n} x_j e_j = \sum_{j=1}^{n} x'_j e'_j. \tag{3.4.1}$$

On cherche à déterminer la relation qui existe entre X et X'.

3.4.2. Définition. *On appelle* matrice de passage *de $\mathcal{B}$ à $\mathcal{B}'$ la matrice carrée dont les colonnes expriment les vecteurs $(e'_j)_{1\leqslant j\leqslant n}$ en fonction des vecteurs $(e_i)_{1\leqslant i\leqslant n}$:*

$$P = \begin{array}{c} \begin{matrix} e'_1 & e'_2 & \ldots & e'_p \end{matrix} \\ \begin{pmatrix} p_{11} & p_{12} & \ldots & p_{1n} \\ p_{21} & p_{22} & \ldots & p_{2n} \\ \vdots & & \ddots & \vdots \\ p_{n1} & p_{n2} & \ldots & p_{nn} \end{pmatrix} \end{array} \begin{matrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{matrix},$$

elle exprime la nouvelle base *en fonction de* l'ancienne base, *sous la forme*

$$e'_j = \sum_{i=1}^{n} p_{ij} e_i.$$

En substituant cette dernière relation dans (3.4.1), il vient

$$x = \sum_{j=1}^{n} x'_j e'_j = \sum_{j=1}^{n} x'_j \left(\sum_{i=1}^{n} p_{ij} e_i \right) = \sum_{i=1} \left(\sum_{j=1}^{n} p_{ij} x'_j \right) e_i = \sum_{i=1} x_i e_i,$$

et par unicité des coordonnées dans une base, on obtient la relation

$$x_i = \sum_{j=1}^{n} p_{ij} x'_j, \quad 1 \leqslant i \leqslant n.$$

Ceci implique :

3.4.3. Théorème. *Si P est la matrice de passage exprimant la* nouvelle base $\mathcal{B}' = (e'_j)$ *en fonction de* l'ancienne base $\mathcal{B} = (e_i)$, *alors les matrices colonnes respectives X, X' des coordonnées d'un vecteur $x \in E$ sont liées par la relation*

$$X = PX',$$

qui exprime les anciennes coordonnées *en fonction des* nouvelles. *Par conséquent, si on cherche à exprimer les* nouvelles coordonnées *en fonction des* anciennes, *on doit calculer*

$$X' = P^{-1}X$$

où P^{-1} est la matrice inverse de P (on verra plus tard comment calculer cette matrice en général, la matric P doit être inversible, sinon il y a erreur).

3.4.4. Exemple. Supposons, dans un $\mathbb{R}$-espace vectoriel E de dimension 3 muni d'une base (i, j, k), que l'on effectue le changement de coordonnées

$$\begin{cases} x' = 2x - y + z \\ y' = -x + 2y + 4z \\ z' = -4x + y + z \end{cases}$$

où (x, y, z) désignent les coordonnées dans la base (i, j, k). À quelle base correspondent ces nouvelles coordonnées ? Pour le trouver, on résout le système ci-dessus de la forme $X' = AX$, ce qui donne une solution unique (vérification laissée au lecteur)

$$\begin{cases} x = -\frac{1}{9}x' + \frac{1}{9}y' - \frac{1}{3}z' \\ y = -\frac{5}{6}x' + \frac{1}{3}y' - \frac{1}{2}z' \\ z = \frac{7}{18}x' + \frac{1}{9}y' + \frac{1}{6}z' \end{cases}.$$

Le changement de coordonnées est bien bijectif, les nouvelles coordonnées sont associées à la base (i', j', k') définie par la matrice de passage $P = A^{-1}$:

$$P = \begin{pmatrix} -\frac{1}{9} & \frac{1}{9} & -\frac{1}{3} \\ -\frac{5}{6} & \frac{1}{3} & -\frac{1}{2} \\ \frac{7}{18} & \frac{1}{9} & \frac{1}{6} \end{pmatrix} \quad \text{soit} \quad \begin{cases} i' = -\frac{1}{9}i - \frac{5}{6}j + \frac{7}{18}k \\ j' = \frac{1}{9}i + \frac{1}{3}j + \frac{1}{9}k \\ k' = -\frac{1}{3}i - \frac{1}{2}j + \frac{1}{6}k \end{cases}.$$

3.4.5. Formule de changement de base pour les applications linéaires. Soit $\ell : E \to F$ une application linéaire. On commence par se donner des bases $\mathcal{B}_E$, $\mathcal{B}_F$ de E, F, puis on les change en des bases $\mathcal{B}'_E$, $\mathcal{B}'_F$. On note

$$A = \mathrm{Mat}_{\mathcal{B}_E}^{\mathcal{B}_F}(\ell), \quad A' = \mathrm{Mat}_{\mathcal{B}'_E}^{\mathcal{B}'_F}(\ell),$$

et on désigne par P la matrice de passage de $\mathcal{B}_E$ à $\mathcal{B}'_E$, et par Q la matrice de passage de $\mathcal{B}_F$ à $\mathcal{B}'_F$. Notre but est de déterminer la relation qui existe entre A et A'. Par définition, ℓ est donnée en coordonnées par

$$x \mapsto y = \ell(x), \quad X \mapsto Y = AX, \quad X' \mapsto Y' = A'X'.$$

Or les "anciennes" et "nouvelles" coordonnées sont liées par les formules de changement de base

$$X = PX', \quad Y = QY' \quad \Leftrightarrow \quad Y' = Q^{-1}Y.$$

On obtient par conséquent

$$Y' = Q^{-1}Y = Q^{-1}(AX) = (Q^{-1}A)X = (Q^{-1}A)(PX') = (Q^{-1}AP)X'.$$

On retiendra (ou non) la formule de changement de base ainsi obtenue :

$$A' = Q^{-1}AP \tag{3.4.6}$$

(mais si on ne la retient pas, il faut être capable de reproduire quasi instantanément la ligne de calcul ci-dessus, ce qui est en fait bien plus important que d'apprendre par cœur la formule sans être capable de la comprendre et de la retrouver ...). □

Rappelons maintenant quelques points de terminologie.

3.4.7. Terminologie.

(i) *On appelle* endomorphisme *toute application linéaire $\ell : E \to E$ d'un espace vectoriel E dans lui-même. On note* $\mathrm{End}_{\mathbb{K}}(E)$ *l'ensemble des endomorphismes $\mathbb{K}$-linéaires de E.*

(ii) *On appelle* isomorphisme *toute application linéaire* bijective *$\ell : E \to F$ entre espaces vectoriels. On peut montrer dans ce cas que $\ell^{-1} : F \to E$ est également linéaire (exercice!). On note* $\mathrm{Isom}_{\mathbb{K}}(E;F)$ *l'ensemble des isomorphismes $\mathbb{K}$-linéaires de E sur F.*

(iii) *On appelle* automorphisme *toute application linéaire* bijective *$\ell : E \to E$ d'un espace vectoriel E dans lui-même, de sorte qu'on a un automorphisme inverse $\ell^{-1} : E \to E$. L'ensemble des automorphismes de E forme un groupe* $(\mathrm{GL}_{\mathbb{K}}(E), \circ)$ *appelé* groupe linéaire *de E.*

Pour un endomorphisme $\ell : E \to E$ et une matrice de passage P d'une base $\mathcal{B}$ vers une base $\mathcal{B}'$ de E, la formule de changement de base s'exprime par

$$A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(\ell), \quad A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(\ell) \implies A' = P^{-1}AP \tag{3.4.8}$$

(dans ce cas, on applique (3.4.6) avec $Q = P$).

3.5. Sous-espaces stables et décompositions par blocs

3.5.1. Définition. *On dit qu'un sous-espace S est stable par un endomorphisme $\ell \in \mathrm{End}_{\mathbb{K}}(E)$ si $\ell(S) \subset S$, autrement dit si $\forall x \in S$, $\ell(x) \in S$.*

3.5.2. Exemple. Dans $\mathbb{R}^3$, l'axe d'une rotation vectorielle constitue une droite stable. De manière générale, la détermination des sous-espaces stables revêt une importance cruciale dans de nombreux domaines des sciences, physique, mécanique, statistiques ... Un des objectifs majeur de ce cours sera de savoir déterminer les sous-espaces S qui sont stables pour un endomorphisme $\ell \in \mathrm{End}_{\mathbb{K}}(E)$ donné. C'est en général un problème tout à fait non trivial, et nous aurons besoin pour cela de développer plusieurs outils nouveaux comme la théorie des déterminants, cf. chapitre 3.

3.5.3. Sous-espace des invariants. Si $\ell \in \mathrm{End}_{\mathbb{K}}(E)$, l'ensemble des invariants de ℓ est par définition

$$\mathrm{Inv}\,\ell = \big\{x \in E \,/\, \ell(x) = x \,/\, x \in E\big\}.$$

C'est un sous-espace vectoriel de E, et il est évident qu'il est stable par ℓ.

Soit $\ell \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme d'un espace E de dimension finie, muni d'une base $\mathcal{B} = (e_1, \ldots, e_n)$. Supposons que l'on connaisse un sous-espace S, par exemple au moyen d'une base $(v_1, \ldots, v_p)$ de S. Le théorème de la base incomplète dit qu'on peut compléter cette famille en une base $\mathcal{B}' = (v_1, \ldots, v_n)$ de E. On obtient alors une décomposition en somme directe

$$E = S \oplus T, \quad S = \mathrm{vect}(v_1, \ldots, v_p), \quad T = \mathrm{vect}(v_{p+1}, \ldots, v_n).$$

Notons $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(\ell)$, $A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(\ell)$. La matrice A n'a pas nécessairement de propriété particulière, mais en revanche, si on calcule $A' = (a'_{ij})_{1 \leqslant i,j \leqslant n}$, ses p premiers vecteurs colonnes sont donnés par

$$\ell(v_j) = \sum_{i=1}^{p} a'_{ij} v_i \in S, \quad 1 \leqslant j \leqslant p$$

(car un vecteur de S ne comporte pas de composantes sur les vecteurs $v_{p+1}, \ldots, v_n$). Ceci implique que A' est de la forme

$$A' = \begin{array}{c} \begin{array}{cccccc} \ell(v_1) & \ldots & \ell(v_p) & \ell(v_{p+1}) & \ldots & \ell(v_n) \end{array} \\ \left(\begin{array}{ccc|ccc} a'_{11} & \ldots & a'_{1p} & a'_{1\,p+1} & \ldots & a'_{1n} \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ a'_{p1} & \ldots & a'_{pp} & a'_{p\,p+1} & \ldots & a'_{pn} \\ \hline 0 & \ldots & 0 & a'_{p+1\,p+1} & \ldots & a'_{p+1\,n} \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & a'_{n\,p+1} & \ldots & a'_{nn} \end{array}\right) \begin{array}{c} v_1 \\ \vdots \\ v_p \\ \\ v_{p+1} \\ \vdots \\ v_n \end{array} \end{array},$$

autrement dit, la matrice de ℓ relativement à la décomposition $E = S \oplus T$ s'écrit

$$A' = \begin{array}{c} \begin{matrix} S & T \end{matrix} \\ \left(\begin{array}{c|c} U & V \\ \hline O & W \end{array}\right) \begin{matrix} S \\ T \end{matrix} \end{array} \quad \text{relativement à} \quad \mathcal{B}', \tag{3.5.4}$$

où $U \in \mathcal{M}_{p\times p}(\mathbb{K})$, $V \in \mathcal{M}_{p\times n-p}(\mathbb{K})$, $W \in \mathcal{M}_{n-p\times n-p}(\mathbb{K})$ et $O \in \mathcal{M}_{n-p\times p}(\mathbb{K})$ est la matrice nulle. Réciproquement, s'il existe une base $\mathcal{B}' = (v_1 \ldots, v_p, v_{p+1}, \ldots v_n)$ telle que, relativement à la décomposition $E = S \oplus T$ avec $S = \mathrm{vect}(v_1, \ldots, v_p)$ et $T = \mathrm{vect}(v_{p+1}, \ldots, v_n)$, la matrice A' de ℓ prenne la forme (3.5.4), alors S est un sous-espace stable de ℓ.

3.5.5. Cas d'une décomposition en N sous-espaces stables. On suppose ici qu'on a une décomposition en somme directe de N sous-espaces

$$E = S_1 \oplus S_2 \oplus \cdots \oplus S_N, \quad \dim S_j = p_j,$$

et que chaque S_j est stable par l'endomorphisme $\ell \in \mathrm{End}_{\mathbb{K}}(E)$. On a alors une restriction bien définie $\ell_j := \ell_{|S_j} : S_j \to S_j$. Considérons une base $\mathcal{B}$ de E formée d'un concaténation de bases de $\mathcal{B}_j$ des S_j, $1 \leqslant j \leqslant N$. La matrice A de ℓ dans la base $\mathcal{B}$ prend alors la forme

$$A = \begin{array}{c} \begin{matrix} S_1 & S_2 & \ldots & S_N \end{matrix} \\ \begin{pmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \vdots & & \ddots & \vdots \\ O & O & \ldots & A_n \end{pmatrix} \begin{matrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{matrix} \end{array}$$

où $A_j = \mathrm{Mat}_{\mathcal{B}_j}^{\mathcal{B}_j}(\ell_j)$ est une matrice carrée $p_j \times p_j$, et les O représentent des matrices nulles rectangulaires $p_i \times p_j$, $i \neq j$; on dit alors que A est une matrice *diagonale par blocs*. Réciproquement, si la matrice $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(\ell)$ est diagonale par blocs comme ci-dessus, on obtient une décomposition $E = S_1 \oplus S_2 \oplus \cdots \oplus S_N$ en sous-espaces S_j stables par ℓ, et si $\ell_j = \ell_{|S_j}$, et on traduira cette situation en écrivant que

$$A = A_1 \boxplus A_2 \boxplus \cdots \boxplus A_N, \quad \ell = \ell_1 \boxplus \ell_2 \boxplus \cdots \boxplus \ell_N$$

(attention : ces "sommes" n'ont rien avoir avec les sommes ordinaires qui n'ont d'ailleurs pas de sens, puisque les A_j n'ont pas nécessairement les même tailles et que les ℓ_j n'opèrent pas sur les mêmes espaces).

3.5.6. Exemple. Prenons $E = \mathbb{R}^5$ muni de la base canonique, et considérons l'endomorphisme $\ell \in \mathrm{End}_{\mathbb{R}}(\mathbb{R}^5)$ de matrice

$$A = \begin{pmatrix} -1 & 2 & 3 & 0 & 0 \\ 3 & -2 & 1 & 0 & 0 \\ 5 & 1 & -4 & 0 & 0 \\ 0 & 0 & 0 & -7 & 2 \\ 0 & 0 & 0 & 6 & 4 \end{pmatrix}$$

On voit alors qu'on a une décomposition en somme directe $\mathbb{R}^5 = S' \oplus S''$ avec

$$S' = \mathbb{R}^3 \times \{(0,0)\} = \{(x_1,x_2,x_3,0,0) \,/\, x_1,x_2,x_3 \in \mathbb{R}\} \quad \text{(isomorphe à } \mathbb{R}^3\text{)},$$
$$S'' = \{(0,0,0)\} \times \mathbb{R}^2 = \{(0,0,0,x_4,x_5) \,/\, x_4,x_5 \in \mathbb{R}\} \quad \text{(isomorphe à } \mathbb{R}^2\text{)},$$

et des endomorphismes $\ell' \in \mathrm{End}_{\mathbb{R}}(S')$, $\ell'' \in \mathrm{End}_{\mathbb{R}}(S'')$ de matrices respectives

$$A' = \begin{pmatrix} -1 & 2 & 3 \\ 3 & -2 & 1 \\ 5 & 1 & -4 \end{pmatrix}, \quad A'' = \begin{pmatrix} -7 & 2 \\ 6 & 4 \end{pmatrix}$$

de sorte que $A = A' \boxplus A''$ et $\ell = \ell' \boxplus \ell''$.

3.5.7. Exemple: rotation vectorielle en dimension 3. On considère ici $E = \mathbb{R}^3$ muni de sa base canonique (i,j,k) et de la structure euclidienne pour laquelle (i,j,k) est orthonormée. Pour la commodité des notations, on identifie $\mathbb{R}^3$ avec l'espace des matrices colonnes à 3 composantes. On se donne l'axe D constitué de la droite vectorielle D orientée par le vecteur unitaire $k' = \frac{1}{3}\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix}$. On vérifie facilement que le plan perpendiculaire $D = P^{\perp}$ admet la base orthonormée (i',j') formée des vecteurs

$$i' = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \perp k', \quad j' = k' \wedge i' = \frac{1}{3}\begin{pmatrix} 1 \\ -2 \\ 2 \end{pmatrix} \wedge \frac{1}{\sqrt{2}}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} = \frac{1}{3\sqrt{2}}\begin{pmatrix} -4 \\ -1 \\ 1 \end{pmatrix},$$

de sorte qu'on a la matrice de passage de $\mathcal{B} = (i,j,k)$ à $\mathcal{B}' = (i',j',k')$ donnée par

$$P = \frac{1}{3\sqrt{2}}\begin{pmatrix} 0 & -4 & \sqrt{2} \\ 3 & -1 & -2\sqrt{2} \\ 3 & 1 & 2\sqrt{2} \end{pmatrix}, \quad \text{et} \quad P^{-1} = {}^tP = \frac{1}{3\sqrt{2}}\begin{pmatrix} 0 & 3 & 3 \\ -4 & -1 & 1 \\ \sqrt{2} & -2\sqrt{2} & 2\sqrt{2} \end{pmatrix}.$$

Maintenant, si $\rho_{P,\theta}$ est la rotation d'angle θ dans le plan orienté $P = \mathrm{vect}(i',j')$, la rotation $R_{D,\theta}$ de $\mathbb{R}^3$ d'angle θ autour de D s'écrit

$$R_{D,\theta} = \rho_{P,\theta} \boxplus \mathrm{Id}_D, \quad \text{d'où} \quad A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(R_{D,\theta}) = \begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

D'après (3.4.8), la matrice $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(R_{D,\theta})$ dans la base $\mathcal{B} = (i,j,k)$ est $A = PA'P^{-1}$, que nous laissons au lecteur le soin d'expliciter.

3.5.8. Cas des projections. Lorsque $E = S_1 \oplus S_2 \oplus \cdots \oplus S_N$, on a pour tout vecteur $x \in E$ une décomposition unique

$$x = x_1 + x_2 + \cdots + x_N, \quad x_j \in S_j.$$

On voit facilement que l'application $\pi_j : E \to S_j \subset E$ telle que $\pi_j(x) = x_j$ (appelée projection de E sur S_j suivant la décomposition $E = \oplus S_j$) est linéaire. En effet, si $y = y_1 + y_2 + \cdots + y_N$, $y_j \in S_j$, alors pour tous $\lambda, \mu \in \mathbb{K}$ on a

$$\lambda x + \mu y = (\lambda x_1 + \mu y_1) + \cdots + (\lambda x_N + \mu y_N), \quad \lambda x_j + \mu y_j \in S_j,$$

donc

$$\pi_j(\lambda x + \mu y) = \lambda x_j + \mu y_j = \lambda\, \pi_j(x) + \mu\, \pi_j(y).$$

On notera que l'on a les relations fondamentales

$$\mathrm{Id}_E = \pi_1 + \pi_2 + \cdots + \pi_N, \quad \pi_j \circ \pi_j = \pi_j, \tag{3.5.9}$$

la première résultant de ce que pour tout $x \in E$ on a

$$\mathrm{Id}_E(x) = x = x_1 + x_2 + \cdots + x_N = \pi_1(x) + \pi_2(x) + \cdots + \pi_N(x),$$

et la deuxième du fait que

$$\pi_j \circ \pi_j(x) = \pi_j(\pi_j(x)) = \pi_j(x_j) = x_j = \pi_j(x).$$

Si $x \in S_j$, on a bien sûr $\pi_j(x) = x$ tandis que si $x \in S_k$ avec $k \neq j$, alors $\pi_j(x_k) = 0$. Ceci montre que (pour une base $\mathcal{B}$ décomposée) la matrice de π_j est une matrice diagonale par blocs

$$\mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(\pi_j) = \begin{pmatrix} O & \ldots & O & \ldots & O \\ \vdots & \ddots & \vdots & & \vdots \\ O & \ldots & I_{p_j} & \ldots & O \\ \vdots & & \vdots & \ddots & \vdots \\ O & \ldots & O & \ldots & O \end{pmatrix}$$

avec le j-ième bloc diagonal égal à la matrice unité $p_j \times p_j$, notée I_{p_j}.

3.5.10. Cas général. Soit $E = S_1 \oplus S_2 \oplus \cdots \oplus S_N$, $\dim S_j = p_j$, et $\ell \in \mathrm{End}_{\mathbb{K}}(E)$. En choisissant une base $\mathcal{B}$ de E formée de la concaténation de bases $\mathcal{B}_j$ des S_j, on peut dans tous les cas écrire la matrice $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(\ell)$ sous forme de blocs A_{ij} de tailles $p_i \times p_j$ représentant des applications linéaires $\ell_{ij} \in \mathcal{L}_{\mathbb{K}}(S_j, S_i)$:

$$A = \begin{pmatrix} A_{11} & \ldots & A_{1j} & \ldots & A_{1N} \\ \vdots & \ddots & \vdots & & \vdots \\ A_{i1} & \ldots & A_{ij} & \ldots & A_{iN} \\ \vdots & & \vdots & \ddots & \vdots \\ A_{N1} & \ldots & A_{Nj} & \ldots & A_{NN} \end{pmatrix}.$$

On a ici $\ell_{ij} = \pi_i \circ \ell \circ \mathrm{inc}_j$ et $\ell = \sum_{1 \leqslant i,j \leqslant N} \mathrm{inc}_i \circ \ell_{ij} \circ \pi_j$ où $\mathrm{inc}_j : S_j \to E$ est l'application d'inclusion de S_j dans E (i.e. $\mathrm{inc}_j(x) = x$ pour $x \in S_j$). Nous laissons ces vérifications évidentes en exercice. □

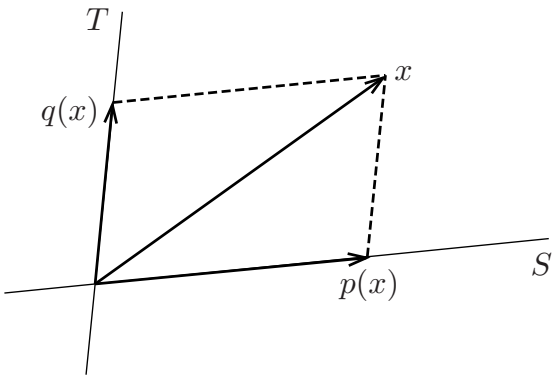
3.6. Caractérisation des projecteurs et des symétries vectorielles

Dans le cas d'une décomposition $E = S \oplus T$ en deux sous-espaces, on a (comme cas particulier de (3.5.9)) des projections

$$p : E \to S \subset E, \quad q : E \to T \subset E \tag{3.6.1}$$

qu'on appelle respectivement projection de E sur S parallèlement à T, et projection de E sur T parallèlement à S. Elles vérifient les relations

$$p \circ p = p, \quad q \circ q = q, \quad p + q = \mathrm{Id}_E\,. \tag{3.6.2}$$



Réciproquement :

3.6.3. Théorème et définition. *Soit E un $\mathbb{K}$-espace vectoriel. On appelle* projecteur *un endomorphisme $p \in \mathrm{End}_{\mathbb{K}}(E)$ tel que $p \circ p = p$ (propriété dite "d'idempotence"). Alors les sous-espaces*

$$S = \mathrm{Inv}\, p = \mathrm{Ker}(\mathrm{Id}_E - p), \quad T = \mathrm{Ker}\, p$$

sont supplémentaires ($E = S \oplus T$), et p est la projection sur S parallèlement à $T = \mathrm{Ker}\, p$. De plus $q = \mathrm{Id}_E - p$ est la projection sur T parallèlement à S, et on a aussi $S = \mathrm{Im}\, p$, $T = \mathrm{Im}\, q$.

Démonstration. Pour tout $x \in E$, on peut écrire

$$x = x_1 + x_2, \quad x_1 = p(x), \quad x_2 = x - p(x),$$

et on a bien $p(x_1) = p \circ p(x) = p(x) = x_1$ (soit encore $(\mathrm{Id}_E - p)(x_1) = 0$), de sorte que $x_1 \in \mathrm{Inv}\, p = \mathrm{Ker}(\mathrm{Id}_E - p)$, et $p(x_2) = p(x) - p \circ p(x) = 0$, de sorte que $x_2 \in \mathrm{Ker}\, p$ et $E = \mathrm{Inv}\, p + \mathrm{Ker}\, p = S + T$. Il reste à montrer que la somme est directe. Si $u \in S \cap T$, alors $u \in S \Rightarrow p(u) = u$ et $u \in T \Rightarrow p(u) = 0$, donc $u = 0$, et on a bien démontré que $E = S \oplus T$. On voit que $x_1 = p(x)$ et $x_2 = q(x)$ sont les projections de x sur S et T suivant la décomposition $E = S \oplus T$. Les égalités $S = \mathrm{Im}\, p$, $T = \mathrm{Im}\, q$ sont immédiates à vérifier. □

3.6.4. Remarque. Dans le cas d'une décomposition $E = S_1 \oplus S_2 \oplus \cdots \oplus S_N$, la projection $\pi_j : E \to S_j \subset E$ définie à la section précédente n'est autre que la projection sur S_j parallèlement à son supplémentaire

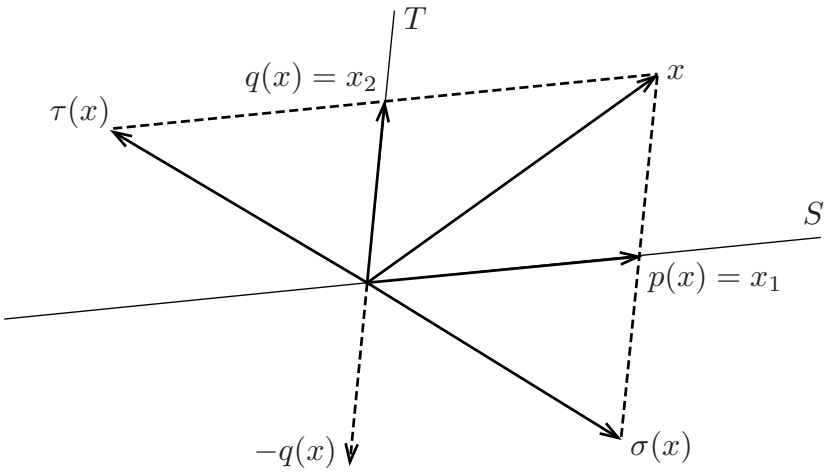
$$T_j = \operatorname{Ker} \pi_j = S_1 \oplus \cdots \oplus S_{j-1} \oplus S_{j+1} \oplus \cdots \oplus S_N. \qquad \square$$

Dans le reste de cette section, on se place dans un corps $\mathbb{K}$ tel que $-1_{\mathbb{K}} \neq 1_{\mathbb{K}}$, c'est-à-dire $2 = 1_{\mathbb{K}} + 1_{\mathbb{K}} \neq 0$ (ce qui est évidemment le cas pour $\mathbb{K} = \mathbb{Q}, \mathbb{R}, \mathbb{C}$). Si $E = S \oplus T$, on appelle symétrie vectorielle par rapport à S parallèlement à T l'endomorphisme $\sigma \in \operatorname{End}_{\mathbb{K}}(E)$ tel que

$$\forall x = x_1 + x_2 \quad \text{avec } x_1 \in S,\ x_2 \in T, \quad \text{on a } \ \sigma(x) = x_1 - x_2. \tag{3.6.5}$$

De même la symétrie vectorielle τ par rapport à T parallèlement à S est définie par

$$\tau(x) = -x_1 + x_2. \tag{3.6.5$'$}$$



On laisse le lecteur vérifier que σ, τ sont linéaires (c'est trivial) et que $\sigma \circ \sigma = \mathrm{Id}_E$, $\tau \circ \tau = \mathrm{Id}_E$; on dit que σ, τ sont des *involutions* ou applications *involutives*. On remarque aussi que $\sigma(x) = p(x) - q(x)$ et $\tau(x) = -\sigma(x)$, d'où

$$\sigma = p - q = p - (\mathrm{Id}_E - p) = 2p - \mathrm{Id}_E = \mathrm{Id}_E - 2q, \tag{3.6.6}$$

$$\tau = -\sigma = \mathrm{Id}_E - 2p = 2q - \mathrm{Id}_E\,. \tag{3.6.6$'$}$$

Réciproquement, les involutions caractérisent les symétries vectorielles :

3.6.7. Théorème. *Soit $\sigma \in \operatorname{End}_{\mathbb{K}}(E)$ une involution, c'est-à-dire une application $\mathbb{K}$-linéaire telle que $\sigma \circ \sigma = \mathrm{Id}_E$. Alors les sous-espaces*

$$S = \{x \in E \,/\, \sigma(x) = x\}, \quad T = \{x \in E \,/\, \sigma(x) = -x\}$$

sont supplémentaires et σ est la symétrie vectorielle par rapport à S parallèlement à T. Ces sous-espaces sont stables par σ.

Démonstration. Comme $2 \neq 0$, l'élément $\frac{1}{2} = 2^{-1} \in \mathbb{K}$ existe et $4 = 2 \times 2 \neq 0$. On pose $p = \frac{1}{2}(\mathrm{Id}_E + \sigma)$. Alors

$$\begin{aligned} p \circ p = p^2 &= \frac{1}{4}(\mathrm{Id}_E + \sigma) \circ (\mathrm{Id}_E + \sigma) \\ &= \frac{1}{4}\big(\mathrm{Id}_E + \sigma \circ \mathrm{Id}_E + \mathrm{Id}_E \circ \sigma + \sigma \circ \sigma\big) = \frac{1}{4}(2\,\mathrm{Id}_E + 2\sigma) = p. \end{aligned}$$

D'après le théorème 3.6.3, p est la projection vectorielle sur

$$S = \{x \in E \,/\, p(x) = x\} = \{x \in E \,/\, \sigma(x) = x\},$$

parallèlement à

$$T = \mathrm{Ker}\, p = \{x \in E \,/\, p(x) = 0\} = \{x \in E \,/\, \sigma(x) = -x\}.$$

Par conséquent, si $q = \mathrm{Id}_E - p$ est la projection complémentaire, on voit que $\sigma = 2p - \mathrm{Id}_E = p - q$ est la symétrie vectorielle par rapport à S parallèlement à T. Le fait que S et T soient stables par σ est évident. □

Chapitre 2
Groupes de permutations

Ce court chapitre a pour but de rappeler les notions de base sur les groupes de permutations et, en particulier, d'établir les propriétés de la signature (la signature d'une permutation est un signe ± 1 décrivant sa parité).

1. Définitions et premières propriétés

1.1. Groupe des permutations d'un ensemble

1.1.1. Définition. *Soit A un ensemble. On note $\mathfrak{S}_A$ l'ensemble des applications bijectives $\sigma : A \to A$ (qu'on appelle aussi permutations de A). Pour la composition des applications $\circ$, on a une structure de groupe $(\mathfrak{S}_A, \circ)$, non commutatif si* $\operatorname{card} A \geqslant 3$.

L'élément neutre du groupe est Id_A et le symétrique d'un élément $\sigma \in \mathfrak{S}_A$ est la bijection inverse σ^{-1} (l'associativité étant toujours vraie pour la composition des applications). On s'intéressera ici surtout au cas où A est un ensemble fini, noté $A = \{a_1, \ldots, a_n\}$.

1.1.2. Notation. *Une permutation $\sigma \in \mathfrak{S}_A$ pourra être définie en donnant la liste des images successives $\sigma(a_i)$ des éléments $a_i \in A$. On notera ainsi*

$$\sigma = \begin{bmatrix} a_1 & a_2 & \ldots & a_n \\ b_1 & b_2 & \ldots & b_n \end{bmatrix}$$

la permutation σ telle que $\sigma(a_i) = b_i$.

1.1.3. Définition. *Le support d'une paermutation $\sigma \in \mathfrak{S}_A$ est par définition la partie*

$$\operatorname{Supp} \sigma = \{x \in A \,/\, \sigma(x) \neq x\}.$$

C'est donc le complémentaire dans A de l'ensemble des éléments invariants, soit

$$\operatorname{Inv} \sigma = \{x \in A \,/\, \sigma(x) = x\}.$$

1.1.4. Définition. *On désigne par $\mathfrak{S}_n$ l'ensemble des permutations de $\{1, 2, ..., n\}$. On a* $\operatorname{card} \mathfrak{S}_n = n!$.

En effet, une telle permutation est obtenue en choisissant $\sigma(1)$ dans $\{1,\dots,n\}$ (n choix possibles), puis $\sigma(2)$ dans $\{1,\dots,n\}\smallsetminus\{\sigma(1)\}$ ($n-1$ choix possibles), puis $\sigma(3)$ dans $\{1,\dots,n\}\smallsetminus\{\sigma(1),\sigma(2)\}$ ($n-2$ choix possibles), etc, ce qui donne

$$\operatorname{card}\mathfrak{S}_n = n\times(n-1)\times(n-2)\times\cdots\times 2\times 1 = n!$$

(il ne reste plus qu'un choix pour le dernier élément $\sigma(n)$, les autres ayant déjà été choisis). □

1.2. Transpositions et cycles

Les exemples fondamentaux de permutations sont les transpositions et les cycles :

1.2.1. Transpositions. Si a,b sont des éléments distincts de A, on note $\tau_{a,b}\in\mathfrak{S}_A$ la permutation définie par

$$\tau_{a,b}(a)=b,\quad \tau_{a,b}(b)=a,\quad \tau_{a,b}(x)=x,\quad \text{si } x\in A\smallsetminus\{a,b\}.$$

La permutation $\tau_{a,b}$ correspond donc à faire l'échange des éléments a,b sans "toucher" aux autres éléments, par suite $\operatorname{Supp}\tau_{a,b}=\{a,b\}$. Il est clair que $\tau_{a,b}$ est une involution, c'est-à-dire que $\tau_{a,b}^2=\mathrm{Id}_A$ (ou encore que c'est un élément d'ordre 2 du groupe $\mathfrak{S}_A$).

1.2.2. Rappel. *Dans un groupe $(G,*)$, un élément x est dit d'ordre fini s'il existe un entier $k\in\mathbb{N}^*$ tel que $x^k=x*x*\cdots*x=1_G$, et on appelle ordre de x, noté* $\operatorname{ordre}(x)$, *le plus petit entier $k\in\mathbb{N}^*$ tel que $x^k=1_G$.*

1.2.3. Cycle de longueur ℓ. Soit $a_1,a_2,\dots,a_\ell$ des éléments 2 à 2 distincts de l'ensemble A. on considère la permutation c définie par

$$c=\begin{bmatrix} a_1 & a_2 & \dots & a_{\ell-1} & a_\ell & b_1 & b_2 & \dots & b_{n-\ell}\\ a_2 & a_3 & \dots & a_\ell & a_1 & b_1 & b_2 & \dots & b_{n-\ell}\end{bmatrix}$$

où $A\smallsetminus\{a_1,\dots,a_\ell\}=\{b_1,\dots,b_{n-\ell}\}$, en d'autres termes c est telle que

$$c: a_1\mapsto a_2\mapsto a_3\mapsto\cdots\mapsto a_{\ell-2}\mapsto a_{\ell-1}\mapsto a_\ell\mapsto a_1$$

et $c(x)=x$ pour $x\notin\{a_1,\dots,a_\ell\}$. Un tel cycle est noté en abrégé

$$c=(a_1\ a_2\ \dots\ a_\ell).$$

Le support du cycle c est donc la partie $\operatorname{Supp}c=\{a_1,\dots,a_\ell\}$, et une transposition $\tau_{a,b}$ n'est pas autre chose qu'un cycle $(a\,b)$ de longueur 2. En général, il est facile de voir que $c^k(a_i)=a_{k+i \bmod \ell}$, c'est-à-dire

$$c^k=\begin{bmatrix} a_1 & a_2 & \dots & a_{\ell-k} & a_{\ell-k+1} & \dots & a_{\ell-1} & a_\ell & b_1 & b_2 & \dots & b_{n-\ell}\\ a_{k+1} & a_{k+2} & \dots & a_\ell & a_1 & \dots & a_{k-1} & a_k & b_1 & b_2 & \dots & b_{n-\ell}\end{bmatrix}$$

pour $k \leqslant \ell - 1$ et $c^\ell = \mathrm{Id}_A$, par conséquent ordre$(c) = \ell$. Il est facile de voir que l'on a pour tout $i = 0, 1, \ldots, \ell - 1$ l'égalité

$$c = (a_1\ a_2\ \ldots\ a_\ell) = (a_{i+1}\ a_{i+2}\ \ldots\ a_\ell\ a_1\ a_2\ \ldots\ a_i),$$

par exemple (1 2 3 4 5) = (4 5 1 2 3), c'est-à-dire que le cycle ne dépend pas de son point de départ, si "l'ordre cyclique" des éléments est préservé. En revanche, le cycle (1 2 3 4 5) n'est pas égal au cycle (1 3 2 4 5).

1.2.4. Exemple. Le groupe $\mathfrak{S}_3$ est constitué des 6 éléments

$$\mathfrak{S}_3 = \{\mathrm{Id}, c, c^2, \tau_{1,2}, \tau_{2,3}, \tau_{1,3}\} \quad \text{où} \quad c = (1\ 2\ 3),\quad c^2 = (1\ 3\ 2),\quad c^3 = \mathrm{Id}\,.$$

On calcule aisément la table de Pythagore du groupe $\mathfrak{S}_3$:

$(u, v) \mapsto u \circ v$

$u \backslash v$	Id	c	c^2	$\tau_{1,2}$	$\tau_{2,3}$	$\tau_{1,3}$
Id	Id	c	c^2	$\tau_{1,2}$	$\tau_{2,3}$	$\tau_{2,3}$
c	c	c^2	Id	$\tau_{1,3}$	$\tau_{1,2}$	$\tau_{2,3}$
c^2	c^2	Id	c	$\tau_{2,3}$	$\tau_{1,3}$	$\tau_{1,2}$
$\tau_{1,2}$	$\tau_{1,2}$	$\tau_{2,3}$	$\tau_{1,3}$	Id	c	c^2
$\tau_{2,3}$	$\tau_{2,3}$	$\tau_{1,3}$	$\tau_{1,2}$	c^2	Id	c
$\tau_{1,3}$	$\tau_{1,3}$	$\tau_{1,2}$	$\tau_{2,3}$	c	c^2	Id

On voit en particulier que le groupe $(\mathfrak{S}_3, \circ)$ est non commutatif, et donc $\mathfrak{S}_n$ est non commutatif pour $n \geqslant 3$ (mais $\mathfrak{S}_1 = \{\mathrm{Id}\}$ et $\mathfrak{S}_2 = \{\mathrm{Id}, \tau_{1,2}\}$ sont commutatifs).

1.3. Décomposition en cycles à supports disjoints

Prenons d'abord l'exemple de la permutation $\sigma \in \mathfrak{S}_8$ telle que

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 6 & 1 & 7 & 8 & 5 & 2 & 3 & 4 \end{bmatrix}.$$

On a $\mathrm{Inv}\,\sigma = \{5\}$ et $\mathrm{Supp}\,\sigma = \{1, 2, 3, 4, 6, 7, 8\}$. Considérons les images des éléments successifs du support :

$$1 \mapsto 6 \mapsto 2 \mapsto 1$$
$$3 \mapsto 7 \mapsto 3$$
$$4 \mapsto 8 \mapsto 4$$

Pour chaque ligne, on prend les images successives et on s'arrête lorsqu'on retombe sur l'élément de départ. On considère ensuite à chaque ligne le premier élément du

support qui n'a pas encore été pris en compte. On voit alors que σ est la composée d'un cycle de longueur 3 et de deux cycles de longueur 2 (transpositions) :

$$\sigma = (1\ 6\ 2) \circ (3\ 7) \circ (4\ 8),$$

avec l'élément 5 qui n'intervient pas (car invariant). L'ordre des composées importe peu, car on a le résultat évident suivant.

1.3.1. Lemme. *Si c et c' sont des cycles dont les supports* $\operatorname{Supp} c$ *et* $\operatorname{Supp} c'$ *sont disjoints* ($\operatorname{Supp} c \cap \operatorname{Supp} c' = \emptyset$), *alors $c' \circ c = c \circ c'$.*

Quel que soit l'ordre de composition, l'image $\sigma(x)$ de la composée σ coïncide en effet avec $c(x)$ si $x \in \operatorname{Supp} c$, avec $c'(x)$ si $x \in \operatorname{Supp} c'$, tandis que $\sigma(x) = x$ si $x \notin \operatorname{Supp} c \cup \operatorname{Supp} c'$. □

En considérant les itérés successifs $\sigma^k(x)$ des éléments du support d'une permutation σ quelconque, on obtient de même le résultat général suivant.

1.3.2. Théorème. *Toute permuation $\sigma \in \mathfrak{S}_A$ d'un ensemble fini A se décompose en un produit commutatif de cycles, c'est-à-dire que*

$$\sigma = c_1 \circ c_2 \circ \cdots \circ c_p$$

avec des cycles c_j dont les supports $\operatorname{Supp} c_j$ *sont 2 à 2 disjoints. Une telle décomposition est unique à l'ordre près des c_j et on a*

$$\operatorname{Supp} \sigma = \operatorname{Supp} c_1 \cup \ldots \cup \operatorname{Supp} c_p.$$

Démonstration. Il faut d'abord voir que si on prend les itérés d'un élément $x_0 \in \operatorname{Supp} \sigma$ quelconque, soit

$$x_0, \quad x_1 = \sigma(x_0), \quad \ldots \quad , \quad x_i = \sigma(x_{i-1}) = \sigma^i(x_0),$$

il y nécessairement un indice $m \in [2, \operatorname{card} A]$ minimal tel que $x_m \in \{x_0, \ldots, x_{m-1}\}$ (sinon on aurait $\operatorname{card}\{x_0, \ldots, x_{i-1}\} \geqslant i$ pour tout i, ce qui contredit la finitude de A). D'autre part, on a nécessairement $x_m = \sigma^m(x_0) = x_0$, sinon on "retomberait" sur $x_m = \sigma^m(x_0) = x_i = \sigma^i(x_0)$ avec $i > 0$, et ceci impliquerait $x_{m-i} = \sigma^{m-i}(x_0) = x_0$, contredisant la minimalité de m. Enfin , si on prend $y_0 \in \operatorname{Supp} \sigma$ en dehors de "l'orbite" $\{x_0, \ldots, x_{m-1}\}$ de x_0, alors tous les itérés $y_i = \sigma^i(y_0)$ sont également en dehors de cette orbite (vérification évidente : $\sigma^i(y_0) = \sigma^j(x_0)$ impliquerait $y_0 = \sigma^{j-i}(x_0)$ ou $y_0 = \sigma^{j+m-i}(x_0)$ suivant que $j \geqslant i$ ou $j < i$). Les orbites qui constituent les supports des cycles sont donc disjointes. □

1.4. Ordre d'une permutation

Soit $\sigma \in \mathfrak{S}_A$ une permutation d'un ensemble fini A, et écrivons

$$\sigma = c_1 \circ c_2 \circ \cdots \circ c_p$$

avec les cycles de longueurs respectives $\ell_1, \ell_2, \ldots, \ell_p$. Comme les cycles commutent, on trouve pour tout $k \in \mathbb{N}^*$

$$\sigma^k = c_1^k \circ c_2^k \circ \cdots \circ c_p^k$$

(on remarquera que dans un groupe non commutatif $(G, \cdot)$, on a en général $(xy)^2 = xyxy$, ce qui ne coïncide avec $x^2y^2 = xxyy$ que si x et y commutent). On a $c_j^k = \mathrm{Id}$ si et seulement si k est multiple de la longeur ℓ_j du cycle c_j. Or, pour $x \in \operatorname{Supp} c_j$, on a $\sigma^k(x) = c_j^k(x)$, donc on voit que $\sigma^k = \mathrm{Id}$ si et seulement si k est simultanément multiple de $\ell_1, \ell_2, \ldots, \ell_p$. Le plus petit entier $k \in \mathbb{N}^*$ tel que $\sigma^k = \mathrm{Id}$ est donc le plus petit commun multiple des ℓ_j. On peut énoncer :

1.4.1. Théorème. *Pour trouver l'ordre d'une permutation σ, on cherche une décomposition en cycles, et alors l'ordre*

$$\operatorname{ordre}(\sigma) = \operatorname{ppcm}(\ell_1, \ell_2, \ldots, \ell_p)$$

est le ppcm des longueurs des cycles $c_1, c_2, \ldots, c_p$ à supports disjoints qui composent σ.

On trouve ainsi par exemple

$$\operatorname{ordre}((1\ 6\ 2) \circ (3\ 7) \circ (4\ 8)) = \operatorname{ppcm}(3, 2, 2) = 6.$$

2. Signature d'une permutation

2.1. Nombre d'inversions et signature

On désigne par P_n l'ensembles des paires $\{i, j\}$ (non ordonnées, $i \neq j$) d'éléments de $\{1, 2, \ldots, n\}$. On a

$$\operatorname{card} P_n = \binom{n}{2} = \frac{n(n-1)}{2}.$$

Si $\sigma \in \mathfrak{S}_n$, alors σ induit une application $\widehat{\sigma} : P_n \to P_n$ définie par

$$\widehat{\sigma}(\{i, j\}) = \{\sigma(i), \sigma(j)\},$$

et il est clair que c'est une bijection de P_n dans P_n, d'inverse $\widehat{\sigma^{-1}}$. On dit que la paire $\{i, j\}$ est inversée par σ (resp. non inversée) si

$$\frac{\sigma(j) - \sigma(i)}{j - 1} < 0, \quad \text{resp.} \quad \frac{\sigma(j) - \sigma(i)}{j - 1} > 0,$$

autrement dit, si $\sigma(i), \sigma(j)$ sont en ordre inverse (ou non) de i, j.

2.1.1. Définition. *Le nombre d'inversions d'une permutation $\sigma \in \mathfrak{S}_n$ est, comme son nom l'indique, le nombre de paires $\{i,j\}$ inversées par σ :*

$$N(\sigma) = \mathrm{card}\Big\{\{i,j\} \in P_n \,/\, \frac{\sigma(j)-\sigma(i)}{j-i} < 0\Big\}.$$

On a donc $N(\sigma) \in \{0, 1, \ldots, \frac{n(n-1)}{2}\}$. La signature $\varepsilon(\sigma)$ de la permutation σ est la valeur ± 1 définie par

$$\varepsilon(\sigma) = (-1)^{N(\sigma)}.$$

2.1.2. Examples.

(a) L'application identique $\sigma = \mathrm{Id}$ n'a pas d'inversions, par conséquent $N(\mathrm{Id}) = 0$, $\varepsilon(\mathrm{Id}) = +1$.

(b) La transposition $\tau_{a,b}$ (avec disons $a < b$) s'écrit

$$\tau_{a,b} = \begin{bmatrix} 1 & 2 & \ldots & a-1 & a & a+1 & \ldots & b-1 & b & b+1 & \ldots n \\ 1 & 2 & \ldots & a-1 & b & a+1 & \ldots & b-1 & a & b+1 & \ldots n \end{bmatrix}$$

donne lieu aux paires inversées $\{a,b\}$ et

$$\{a,i\} \mapsto \{b,i\}, \quad a+1 \leqslant i \leqslant b-1,$$
$$\{i,b\} \mapsto \{i,a\}, \quad a+1 \leqslant i \leqslant b-1,$$

soit $2p+1$ paires inversées avec $p = (b-1)-(a+1)+1 = b-a-1$. On a donc $\varepsilon(\tau_{a,b}) = -1$.

(c) Le cycle $c = (1\ 2\ \cdots\ \ell)$ de longueur ℓ

$$c = \begin{bmatrix} 1 & 2 & \ldots & \ell-1 & \ell & \ell+1 & \ldots & n \\ 2 & 3 & \ldots & \ell & 1 & \ell+1 & \ldots & n \end{bmatrix}$$

donne lieu aux paires inversées $\{i,\ell\} \mapsto \{i+1, 1\}$ pour $1 \leqslant i \leqslant \ell-1$. On obtient par conséquent

$$N(c) = \ell - 1, \quad \varepsilon(c) = (-1)^{\ell-1}.$$

(d) La permutation σ correspondant au renversement de l'ordre

$$c = \begin{bmatrix} 1 & 2 & \ldots & n-1 & n \\ n & n-1 & \ldots & 2 & 1 \end{bmatrix}$$

admet le nombre maximum $N(\sigma) = \frac{n(n-1)}{2}$ d'inversions, et on a par conséquent $\varepsilon(\sigma) = (-1)^{n(n-1)/2}$.

On a la formule importante suivante

2.1.3. Formule de la signature. *Pour tout $\sigma \in \mathfrak{S}_n$, on a*

$$\varepsilon(\sigma) = \prod_{\{i,j\}\in P_n} \frac{\sigma(j)-\sigma(i)}{j-i}.$$

Démonstration. Posons

$$\widetilde{\varepsilon}(\sigma) = \prod_{\{i,j\}\in P_n} \frac{\sigma(j)-\sigma(i)}{j-i} \in \mathbb{Q}^*.$$

Il est clair que le signe de $\widetilde{\varepsilon}(\sigma)$ est $(-1)^{N(\sigma)}$. Mais si on fait le changement de variable bijectif $\{u,v\} = \widehat{\sigma}(\{i,j\}) = \{\sigma(i),\sigma(j)\}$, on voit que le numérateur et le démoninateur de $\widetilde{\varepsilon}(\sigma)$ sont tous les deux égaux en valeur absolue à

$$\prod_{\{u,v\}\in P_n} |v-u| = \prod_{2\leqslant v\leqslant n} \prod_{1\leqslant u\leqslant v-1} (v-u) = \prod_{2\leqslant v\leqslant n} (v-1)! = \prod_{1\leqslant i\leqslant n-1} i! = \prod_{i=1}^{n-1} i^{n-i}.$$

Il en résulte que $|\widetilde{\varepsilon}(\sigma)| = 1$ et donc $\widetilde{\varepsilon}(\sigma) = \varepsilon(\sigma)$. □

2.2. Propriété d'homomorphisme de la signature

On va voir que $\varepsilon : \mathfrak{S}_n \to \{+1,-1\}$ est un homomorphisme du groupe $(\mathfrak{S}_n, \circ)$ dans le groupe multiplicatif $(\{+1,-1\}, \times)$, autrement dit :

2.2.1. Théorème. *Pour toutes permutations $\sigma, \tau \in \mathfrak{S}_n$, on a*

$$\varepsilon(\sigma\circ\tau) = \varepsilon(\sigma)\varepsilon(\tau).$$

Rappelons qu'un homomorphisme $\varphi : G \to H$ entre deux groupes $(G,*)$, $(H,*')$ est une application telle que, pour tous $x, y \in G$, on ait $\varphi(x*y) = \varphi(x) *' \varphi(y)$. Dans ce cas

$$\operatorname{Ker}\varphi = \{x\in G \,/\, \varphi(x) = 1_H\}, \quad \operatorname{Im}\varphi = \{u = \varphi(x) \in H \,/\, x\in G\}$$

sont des sous-groupes de G et H respectivement.

Démonstration. Pour toutes permutations $\sigma, \tau \in \mathfrak{S}_n$, il vient

$$\varepsilon(\sigma\circ\tau) = \prod_{\{i,j\}\in P_n} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{j-i} = \prod_{\{i,j\}\in P_n} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)} \prod_{\{i,j\}\in P_n} \frac{\tau(j)-\tau(i)}{j-i}.$$

Dans le premier produit du membre de droite, faisons le changement de variable bijectif $\{u,v\} = \widehat{\tau}(\{i,j\}) = \{\tau(i),\tau(j)\}$. Ceci donne

$$\prod_{\{i,j\}\in P_n} \frac{\sigma(\tau(j))-\sigma(\tau(i))}{\tau(j)-\tau(i)} = \prod_{\{u,v\}\in P_n} \frac{\sigma(v)-\sigma(u)}{v-u}.$$

Par conséquent

$$\varepsilon(\sigma\circ\tau)=\prod_{\{u,v\}\in P_n}\frac{\sigma(v)-\sigma(u)}{v-u}\prod_{\{i,j\}\in P_n}\frac{\tau(j)-\tau(i)}{j-i}=\varepsilon(\sigma)\varepsilon(\tau).\qquad\square$$

2.2.2. Corollaire. *Soit $A=\{a_1,\ldots,a_n\}$ un ensemble fini. Une permutation $\sigma\in\mathfrak{S}_A$ est définie à l'aide d'une permutation $\alpha\in\mathfrak{S}_n$ par la correspondance bijective*

$$\alpha\in\mathfrak{S}_n\longmapsto\sigma\in\mathfrak{S}_A,\quad \sigma(a_i)=a_{\alpha(i)}.$$

Alors la signature de σ définie par $\varepsilon(\sigma):=\varepsilon(\alpha)$ ne dépend pas de la numérotation des éléments de A, autrement dit, si $A=\{a'_1,\ldots,a'_n\}$ avec une autre numérotation des éléments, et si $\beta\in\mathfrak{S}_n$ est telle que $\sigma(a'_i)=a'_{\beta(i)}$, on a bien $\varepsilon(\alpha)=\varepsilon(\beta)$.

Démonstration. Le changement de numérotation est donné par $a'_i=a_{\gamma(i)}$ avec une certaine permutation $\gamma\in\mathfrak{S}_n$. Posant $j=\gamma(i)$ et $i=\gamma^{-1}(j)$, il vient $a'_{\gamma^{-1}(j)}=a_j$, donc

$$\sigma(a'_i)=\sigma(a_{\gamma(i)})=a_{\alpha(\gamma(i)}=a'_{\gamma^{-1}(\alpha(\gamma(i)))}=a'_{\beta(i)},$$

ce qui montre que les permutations $\alpha,\beta\in\mathfrak{S}_n$ sont liées par $\beta=\gamma^{-1}\circ\alpha\circ\gamma$. Mais on a alors

$$\varepsilon(\beta)=\varepsilon(\gamma)^{-1}\varepsilon(\alpha)\varepsilon(\gamma)=\varepsilon(\alpha).\qquad\square$$

2.2.3. Corollaire. *Pour tout ensemble fini A, il existe un homomorphisme signature $\varepsilon:\mathfrak{S}_A\to\{+1,-1\}$ défini indépendamment de la numérotation des éléments.*

2.3. Calcul de la signature d'une permutation quelconque

2.3.1. Proposition. *Si $c=(a_1\,a_2\ldots a_\ell)$ est un cycle de longueur ℓ dans un ensemble fini A, alors $\varepsilon(c)=(-1)^{\ell-1}$.*

Démonstration. Il suffit de numéroter les éléments en sorte que $a_1,a_2,\ldots,a_\ell$ soient précisément les ℓ premiers éléments de A, et d'observer que le nombre d'inversions de $(1\ 2\ \ldots\ \ell)$ est alors exactement $\ell-1$ (on applique ici le corollaire 2.2.3). $\square$

2.3.2. Corollaire. *Pour une permutation $\sigma\in\mathfrak{S}_A$ décomposée comme*

$$\sigma=c_1\circ c_2\circ\ldots\circ c_p$$

avec des cycles c_j à supports disjoints de longueurs respectives $\ell_1,\ell_2,\ldots,\ell_p$, on a

$$\varepsilon(\sigma)=(-1)^{(\ell_1-1)+(\ell_2-1)+\cdots+(\ell_p-1)}.$$

On observera qu'il est algorithmiquement beaucoup plus efficace de calculer la signature à l'aide d'une décomposition en cycles qu'en examinant les inversions

de toutes les paires $\{i,j\} \in P_n$. En effet, dans le premier cas, on fait un nombre d'opérations d'un ordre de grandeur égal à n, alors que dans le deuxième cas, c'est de l'ordre de $\frac{n(n-1)}{2} \sim \frac{1}{2}n^2$.

2.3.3. Remarque. Pour $\{a_1, a_2, \dots, a_\ell\} \subset \{1, 2, , \dots, n\}$, une façon équivalente de démontrer la proposition 2.3.1 est d'observer que le cycle $c = (a_1\, a_2 \dots a_\ell)$ est le conjugué du cycle $c_\ell = (1\ 2\ \dots\ \ell)$ par la permutation

$$\gamma = \begin{bmatrix} 1 & 2 & \dots & \ell & \ell+1 & \dots & n \\ a_1 & a_2 & \dots & a_\ell & b_1 & \dots & b_{n-\ell} \end{bmatrix}$$

où $\{b_1, \dots, b_{n-\ell}\}$ est le complémentaire de $\{a_1\, a_2 \dots a_\ell\}$ dans $\{1, 2, \dots, n\}$, c'est-à-dire que $c = \gamma \circ c_\ell \circ \gamma^{-1}$ (exercice !)
Plus généralement, on voit facilement que deux permutations $\sigma, \sigma' \in \mathfrak{S}_n$ sont conjuguées, i.e. $\sigma' = \gamma \circ \sigma \circ \gamma^{-1}$ pour un certain élément $\gamma \in \mathfrak{S}_n$, si et seulement si elles ont des décompositions en cycles disjoints

$$\sigma = c_1 \circ c_2 \circ \dots \circ c_p, \quad \sigma' = c'_1 \circ c'_2 \circ \dots \circ c'_p$$

formées du même nombre p de cycles, avec des longueurs identiques $\ell'_j = \ell_j$ (après avoir éventuellement réordonné les composées). Il suffit pour cela de prendre γ qui envoie $\operatorname{Supp} c_j$ sur $\operatorname{Supp} c'_j$ en respectant l'ordre cyclique des éléments dans ces cycles, et qui envoie $\{1, 2, \dots, n\} \smallsetminus \bigcup \operatorname{Supp} c_j$ bijectivement sur $\{1, 2, \dots, n\} \smallsetminus \bigcup \operatorname{Supp} c'_j$.

2.4. Le sous-groupe alterné $\mathcal{A}_n$

2.4.1. Définition. *On pose*

$$\mathcal{A}_n = \ker \varepsilon = \{\sigma \in \mathfrak{S}_n \,/\, \varepsilon(\sigma) = +1\}.$$

C'est un sous-groupe de $\mathfrak{S}_n$.

2.4.2. Proposition. *On a* $\mathcal{A}_1 = \mathfrak{S}_1 = \{\mathrm{Id}\}$, *et pour* $n \geqslant 2$, $\operatorname{card} \mathcal{A}_n = \frac{1}{2}n!$.

Démonstration. Posons

$$\mathfrak{S}_n^+ = \{\sigma \in \mathfrak{S}_n \,/\, \varepsilon(\sigma) = +1\} = \mathcal{A}_n, \quad \mathfrak{S}_n^- = \{\sigma \in \mathfrak{S}_n \,/\, \varepsilon(\sigma) = -1\}.$$

Alors on a la réunion disjointe $\mathfrak{S}_n = \mathfrak{S}_n^+ \cup \mathfrak{S}_n^-$, et pour $n \geqslant 2$, on a une bijection

$$\mathfrak{S}_n^+ \longrightarrow \mathfrak{S}_n^-, \quad \sigma \longmapsto \sigma \circ \tau_{1,2}.$$

Par conséquent $\operatorname{card} \mathfrak{S}_n^+ = \operatorname{card} \mathfrak{S}_n^- = \operatorname{card} \mathcal{A}_n = \frac{1}{2}n!$. □

2.4.3. Complément historique. Pour $n \geqslant 5$, on peut démontrer que $\mathcal{A}_n$ est un groupe simple, c'est-à-dire que $\mathcal{A}_n$ n'a aucun sous-groupe distingué H autre

que $H = \{\mathrm{Id}\}$ et $H = \mathcal{A}_n$ (un sous-groupe distingué H d'un groupe G est un sous-groupe invariant par conjugaison : $\forall \gamma \in G$, $\gamma H \gamma^{-1} = H$) ; d'autre part, $\mathcal{A}_n$ est non commutatif si $n \geqslant 5$. Vers 1830, Évariste Galois a déduit de ce résultat que les racines complexes $z_1, \ldots, z_n$ d'un polynôme général $P \in \mathbb{Q}[X]$ de degré n ne peuvent s'exprimer par radicaux à partir de $\mathbb{Q}$, à savoir comme combinaisons de racines p-ièmes "enchevêtrées" en partant des rationnels – c'était une question ouverte depuis la découverte des formules de résolution des équations de degré 3 et 4 par Tartaglia et Ferrari au 16$^{\mathrm{e}}$ siècle. On vérifie en effet que le corps $\mathbb{K} = \mathbb{Q}[z_1, \ldots, z_n]$ engendré par les racines de P admet un groupe d'automorphismes $\mathrm{Aut}(\mathbb{K})$ de permutation des racines égal à $\mathfrak{S}_n$ si P est général. Or, $\mathfrak{S}_n$ ne peut se "dévisser" à l'aide de groupes abéliens, alors que ce serait le cas pour $\mathrm{Aut}(\mathbb{K})$ si les racines étaient résolubles par radicaux. É. Galois a découvert ces résultats alors qu'il avait à peine 20 ans, et les a consignés fébrilement dans un testament écrit à la veille de son duel. Ils sont restés incompris de la communauté mathématique pendant au moins 20 ans. C'est d'ailleurs à cette occasion qu'il a introduit la notion fondamentale de groupe !

3. Générateurs du groupe des permutations

3.1. Génération par transpositions

3.1.1. Théorème. *Toute permutation $\sigma \in \mathfrak{S}_n$ s'écrit comme un produit d'au plus $\frac{n(n-1)}{2}$ transpositions $\tau_{i,i+1}$ portant sur des éléments consécutifs, $1 \leqslant i \leqslant n-1$, c'est-à-dire*

$$\sigma = \tau_{i_1,i_1+1} \circ \tau_{i_2,i_2+1} \circ \cdots \circ \tau_{i_k,i_k+1}, \quad k \leqslant \frac{n(n-1)}{2}.$$

Démonstration. On raisonne par récurrence sur $N(\sigma)$. Pour $N(\sigma) = 0$, on a $\sigma = \mathrm{Id}$, et le résutat est vrai avec $k = 0$ (produit vide, égal à Id par convention).

Supposons maintenant que $N = N(\sigma) \geqslant 1$ et que le résultat ait déjà été démontré pour les permutations σ' telles que $N(\sigma') = N - 1$. Il existe alors $j \in \{1, 2, \ldots, n-1\}$ tel que $\sigma(j) > \sigma(j+1)$, sinon σ serait strictement croissante (donc $\sigma = \mathrm{Id}$ et $N(\sigma) = 0$ contrairement à notre hypothèse). Posons

$$\begin{aligned}\sigma' &= \sigma \circ \tau_{j,j+1} \\ &= \begin{bmatrix} 1 & 2 & \ldots & j-1 & j & j+1 & j+2 & \ldots & n \\ \sigma(1) & \sigma(2) & \ldots & \sigma(j-1) & \sigma(j+1) & \sigma(j) & \sigma(j+2) & \ldots & \sigma(n) \end{bmatrix}.\end{aligned}$$

Alors l'inversion $\sigma(j+1) < \sigma(j)$ n'est plus une inversion pour σ'. On a donc $N(\sigma') = N(\sigma) - 1 = N - 1$, et par hypothèse de récurrence, il existe une décomposition

$$\sigma' = \tau_{i_1,i_1+1} \circ \tau_{i_2,i_2+1} \circ \cdots \circ \tau_{i_\ell,i_\ell+1},$$

d'où

$$\sigma = \sigma' \circ \tau_{j,j+1} = \tau_{i_1,i_1+1} \circ \tau_{i_2,i_2+1} \circ \cdots \circ \tau_{i_\ell,i_\ell+1} \circ \tau_{j,j+1}.$$

Par récurrence, ce raisonnement fournit une décomposition ayant exactement $k = N(\sigma)$ transpositions $\tau_{i,i+1}$, de sorte que $k \leqslant \frac{N(N-1)}{2}$. □

3.1.2. Corollaire. *En particulier, toute permutation $\sigma \in \mathfrak{S}_n$ est une composée*

$$\sigma = \tau_{a_1,b_1} \circ \tau_{a_2,b_2} \circ \cdots \circ \tau_{a_k,b_k}$$

de transpositions, et la signature $\varepsilon(\sigma) = (-1)^k$ est déterminée par la parité du nombre de transpositions nécessaires (et inversement).

3.2. Génération par une transposition et un cycle

Si $c = (1\ 2\ \ldots\ n)$ est le cycle $1 \mapsto 2 \mapsto \cdots \mapsto (n-1) \mapsto n \mapsto 1$ de longueur n, on a $c^{j-1}(i) = i + j - 1$ modulo n, et on voit facilement que

$$\tau_{j,j+1} = c^{j-1} \circ \tau_{1,2} \circ c^{-(j-1)}$$

puisque $c^{-(j-1)}$ "ramène" $\{j, j+1\}$ sur $\{1, 2\}$, tandis que c^{j-1} "renvoie" $\{2, 1\}$ sur $\{j+1, j\}$. Ceci montre que les transpositions $\tau_{j,j+1}$ sont toutes conjuguées de $\tau_{1,2}$ par des puissances de c. Le théorème 3.1.1 implique alors

3.2.1. Théorème. *Le groupe $\mathfrak{S}_n$ est engendré par le cycle $c = (1\ 2\ \ldots\ n)$ et la permutation $\tau = \tau_{1,2}$, c'est-à-dire que toute permutation σ peut s'écrire comme une composée (non commutative) de τ et de puissances c^i entremêlées :*

$$\sigma = c^{j_0} \circ \tau \circ c^{j_1} \circ \tau \circ \cdots \circ c^{j_{k-1}} \circ \tau \circ c^{j_k}, \qquad 0 \leqslant j_\ell \leqslant n-1.$$

3.2.2. Exercice. On peut démontrer que pour $n \geqslant 3$ le groupe alterné $\mathcal{A}_n$ est engendré par les cycles $(a_1\ a_2\ a_3)$ de longueur 3. Exercice pour le lecteur !

Chapitre 3
Applications multilinéaires et déterminants

Ce chapitre a pour but de développer la théorie des déterminants, un outil pratique et théorique important pour résoudre de nombreux problèmes d'algèbre linéaire, en particulier la réduction des matrices à des formes canoniques.

1. Applications multilinéaires

1.1. Cas bilinéaire, concepts fondamentaux

1.1.1. Définition. *Soient E, F, G des $\mathbb{K}$-espaces vectoriels, et soit*

$$\varphi : E \times F \to G, \quad (x, y) \mapsto \varphi(x, y)$$

une application de $E \times F$ dans G. On dit que φ est une application bilinéaire *si φ est linéaire en chacune des variables. Autrement dit, φ est une application bilinéaire si pour tous $x, x' \in E$, $y, y' \in F$, et tous $\lambda, \mu \in \mathbb{K}$, on a*

(a) $\varphi(x + x', y) = \varphi(x, y) + \varphi(x', y), \quad \varphi(\lambda x, y) = \lambda\, \varphi(x, y)$ (*linéarité en x*),

(b) $\varphi(x, y + y') = \varphi(x, y) + \varphi(x, y'), \quad \varphi(x, \mu y) = \mu\, \varphi(x, y)$ (*linéarité en y*).

En prenant $\lambda = \mu = 0$, on voit qu'on a nécessairement

$$\varphi(0, y) = \varphi(x, 0) = 0 \text{ pour tous } x \in E \text{ et } y \in F.$$

Une autre façon équivalente de formuler les axiomes de la bilinéarité est de vérifier l'unique identité de "distributivité"

$$\varphi(\lambda x + \lambda' x', \mu y + \mu' y') = \lambda\mu\, \varphi(x, y) + \lambda\mu'\, \varphi(x, y') + \lambda'\mu\, \varphi(x', y) + \lambda'\mu'\, \varphi(x', y'),$$

(mais cette forme plus concise n'est pas nécessairement la plus pratique à mettre en œuvre).

1.1.2. Définition. *Une* forme bilinéaire $\varphi : E \times F \to \mathbb{K}$ *est une application bilinéaire à valeurs dans le corps des scalaires, i.e. telle que $G = \mathbb{K}$.*

Très souvent, on sera amené à considérer le cas où $E = F$, c'est-à-dire le cas où les vecteurs x, y sont pris dans le même espace vectoriel E. On introduit alors la terminologie suivante :

1.1.3. Définition. *Une application bilinéaire $\varphi : E \times E \to G$ est dite*

(a) *symétrique si $\varphi(y,x) = \varphi(x,y)$ pour tous $x, y \in E$.*

(b) *antisymétrique si $\varphi(y,x) = -\varphi(x,y)$ pour tous $x, y \in E$.*

(c) *alternée si $\varphi(x,x) = 0$ pour tout $x \in E$.*

Remarquons que l'on a pour tous $x, y \in E$

$$\varphi(x,y) + \varphi(y,x) = \varphi(x+y, x+y) - \varphi(x,x) - \varphi(y,y),$$

donc il résulte de cette identité :

1.1.4. Propriété. *Toute application bilinéaire alternée est antisymétrique.*

Réciproquement, si φ est antisymétrique, en prenant $y = x$ dans 1.1.3 (b), on en déduit que $2\,\varphi(x,x) = 0$, et donc $\varphi(x,x) = 0$ si le corps $\mathbb{K}$ n'est pas de caractéristique 2, c'est-à-dire si $2_{\mathbb{K}} = 1_{\mathbb{K}} + 1_{\mathbb{K}} \neq 0$. (Un exemple de corps de caractéristique 2 est le corps $\mathbb{K} = \{0,1\}$, avec précisément la loi d'addition $1+1=0$; dans un tel corps $\mathbb{K}$, on a $-1_{\mathbb{K}} = 1_{\mathbb{K}}$, et une application bilinéaire φ est antisymétrique si et seulement elle est symétrique !).

1.1.5. Propriété. *Si $\mathbb{K}$ n'est pas de caractéristique 2 (ce qui est le cas des corps usuels $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$), une forme bilinéaire est alternée si et seulement si elle est antisymétrique.*

1.1.6 . Exemples.

(a) L'application

$$\varphi : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}, \quad (x,y) = ((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) \mapsto x \cdot y = \sum_{i=1}^{n} x_i y_i$$

induite par le produit scalaire usuel est une forme bilinéaire symétrique.

(b) Identifions ici $\mathbb{R}^3$ aux vecteurs colonnes de dimension 3. L'application dite de produit vectoriel

$$\varphi : \mathbb{R}^3 \times \mathbb{R}^3 \to \mathbb{R}^3, \quad (x,y) = \left(\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}, \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \right) \mapsto x \wedge y = \begin{pmatrix} x_2 y_3 - x_3 y_2 \\ x_3 y_1 - x_1 y_3 \\ x_1 y_2 - x_2 y_1 \end{pmatrix}$$

est une application bilinéaire alternée (mais ce n'est pas une *forme* bilinéaire, car elle est à valeurs vectorielles et non pas à valeurs scalaires).

(c) Le "crochet de commutation"

$$\varphi : \mathcal{M}_{n\times n}(\mathbb{R}) \times \mathcal{M}_{n\times n}(\mathbb{R}) \to \mathcal{M}_{n\times n}(\mathbb{R}), \quad (M,N) \mapsto [M,N] = MN - NM$$

est une application bilinéaire alternée.

(d) L'application

$$\varphi : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}, \quad ((x_1, x_2), (y_1, y_2)) \mapsto x_1 x_2 + 2x_1 y_2$$

n'est pas bilinéaire. En effet, posons $x = (x_1, x_2)$, $y = (y_1, y_2)$. On a

$$\varphi(\lambda x, y) = (\lambda x_1)(\lambda x_2) + 2(\lambda x_1) y_2 = \lambda^2 x_1 x_2 + 2\lambda x_1 y_2$$

et en général

$$\varphi(\lambda x, y) \neq \lambda\, \varphi(x, y) = \lambda(x_1 x_2 + 2x_1 y_2)$$

(prendre par exemple $x_1 = x_2 = y_1 = y_2 = 1$, $\lambda = 2$).

1.2. Cas multilinéaire général

1.2.1. Définition. *Soient $E_1, \ldots, E_p$ et G des $\mathbb{K}$-espaces vectoriels, et soit*

$$\varphi : E_1 \times \cdots \times E_p \to G, \quad (x_1, \ldots, x_p) \mapsto \varphi(x_1, \ldots, x_p)$$

une application de $E_1 \times \cdots \times E_p$ dans G. On dit que φ est une application multilinéaire (*ou parfois p-multilinéaire*) *si φ est linéaire en chacune des variables. Autrement dit, pour tout indice $j \in \{1, \ldots, p\}$ et pour tous vecteurs $x_1 \in E_1 \ldots, x_{j-1} \in E_{j-1}$, $x_{j+1} \in E_{j+1}, \ldots, x_p \in E_p$ fixés, l'application*

$$E_p \to G, \quad x_j \mapsto \varphi(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots x_p)$$

est linéaire, ou encore :

$$\varphi(x_1, \ldots, x_{j-1}, \lambda x_j + \lambda' x'_j, x_{j+1}, \ldots x_p) = \lambda\, \varphi(x_1, \ldots, x_{j-1}, x_j, x_{j+1}, \ldots x_p) + \lambda'\, \varphi(x_1, \ldots, x_{j-1}, x'_j, x_{j+1}, \ldots x_p)$$

pour tous scalaires $\lambda, \lambda' \in \mathbb{K}$ et tous les vecteurs x_i, x'_i impliqués.

Nous nous intéresserons particulièrement au cas où $E_1 = \ldots = E_p = E$. Dans ce cas on généralise comme suit les notions de symétrie et d'antisymétrie.

1.2.2. Définition. *Une application multilinéaire $\varphi : E^p \to G$ est dite*

(a) *symétrique si pour tous vecteurs $x_1, \ldots, x_p \in E$ et toute permutation $\sigma \in \mathfrak{S}_p$ on a*

$$\varphi(x_{\sigma(1)}, \ldots, x_{\sigma(p)}) = \varphi(x_1, \ldots, x_p).$$

(b) *antisymétrique si pour tous vecteurs $x_1, \ldots, x_p \in E$ et toute permutation $\sigma \in \mathfrak{S}_p$ on a*

$$\varphi(x_{\sigma(1)}, \ldots, x_{\sigma(p)}) = \varepsilon(\sigma)\, \varphi(x_1, \ldots, x_p)$$

où $\varepsilon(\sigma) = \pm 1$ est la signature de σ.

(c) *alternée si $\varphi(x_1, \ldots, x_p) = 0$ chaque fois que l'on a $x_i = x_j$ pour des indices $i \neq j$ quelconques.*

1.2.3. Remarque. Comme les transpositions engendrent $\mathfrak{S}_p$, il est facile de voir que les conditions de symétrie et d'antisymétrie sont respectivement equivalentes à

$$\varphi(x_{\tau(1)},\ldots,x_{\tau(p)}) = \begin{cases} \varphi(x_1,\ldots,x_p) & \text{resp.} \\ -\varphi(x_1,\ldots,x_p) & \end{cases}$$

pour toute transposition $\tau = \tau_{i,j}$ de deux des indices i, j. Il suffirait même de le vérifier pour des transpositions $\tau_{i,i+1}$ d'indices consécutifs, puisque celles-ci engendrent $\mathfrak{S}_p$.

D'autre part, si on fixe tous les vecteurs x_k, $k \neq i, j$, l'application

$$E \times E \to G, \quad (x_i, x_j) \mapsto \varphi(x_1,\ldots,x_i,\ldots,x_j\ldots,x_p)$$

est bien entendu bilinéaire. Par conséquent, on peut encore appliquer ici les propriétés 1.1.4 et 1.1.5 et conclure :

1.2.4. Propriété. *Toute application multilinéaire alternée est antisymétrique, et la réciproque est également vraie si le corps $\mathbb{K}$ n'est pas de caractéristique 2.*

1.3. Écriture d'une forme multilinéaire dans une base

Dans le reste de ce chapitre, on s'intéressera surtout au cas des *formes multilinéaires*

$$\varphi : E^p \to \mathbb{K}.$$

On suppose que E est de dimension finie n et muni d'une base $(e_1,\ldots,e_n)$ quelconque.

1.3.1. Cas bilinéaire ($p = 2$). Si $x = \sum_{i=1}^n x_i e_i$ et $y = \sum_{j=1}^n y_j e_j$, il vient par bilinéarité

$$\begin{aligned}\varphi(x,y) &= \varphi\left(\sum_{i=1}^n x_i e_i, y\right) = \sum_{i=1}^n x_i\,\varphi(e_i,y) \\ &= \sum_{i=1}^n x_i\,\varphi\left(e_i, \sum_{j=1}^n y_j e_j\right) = \sum_{i=1}^n x_i \left(\sum_{j=1}^n y_j\,\varphi(e_i,e_j)\right) \\ &= \sum_{1\leqslant i,j\leqslant n} x_i y_j\,\varphi(e_i,e_j),\end{aligned}$$

c'est-à-dire

$$\varphi(x,y) = \sum_{1\leqslant i,j\leqslant n} c_{ij}\, x_i y_j \tag{1.3.2}$$

en notant $c_{ij} = \varphi(e_i,e_j) \in \mathbb{K}$. Réciproquement, pour des coefficients $c_{ij} \in \mathbb{K}$ quelconques, l'expression (1.3.2) définit bien une forme bilinéaire $\varphi : E^2 \to \mathbb{K}$, et

il est facile de voir que les coefficients c_{ij} sont déterminés de manière unique par la condition $c_{ij} = \varphi(e_i, e_j)$. On peut associer à φ sa matrice de coefficients

$$C = (c_{ij})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_{n\times n}(\mathbb{K}).$$

Les conditions de symétrie et d'antisymétrie se lisent facilement sur les coefficients. Rappelons que la transposée de la matrice C est définie par ${}^tC = C' = (c'_{ij})$ avec $c'_{ij} = c_{ji}$ pour tous $i, j \in \{1, \ldots, n\}$.

1.3.3. Propriétés. *La forme bilinéaire $\varphi : E^2 \to \mathbb{K}$ est*

(a) symétrique *si et seulement si les coefficients vérifient $c_{ji} = c_{ij}$ pour tous i, j, autrement dit, si la matrice C de coefficients est symétrique : ${}^tC = C$;*

(b) antisymétrique *si et seulement si les coefficients vérifient $c_{ji} = -c_{ij}$ pour tous i, j, autrement dit, si la matrice C de coefficients est antisymétrique : ${}^tC = -C$;*

(c) alternée *si et seulement si elle est antisymétrique, i.e. $c_{ji} = -c_{ij}$, et de plus $c_{ii} = 0$ pour tout $i = 1, \ldots, n$.*

1.3.4. Cas p-multilinéaire. Les notations sont plus compliquées. On doit prendre p vecteurs $x_j \in E$ qu'on écrit

$$x_j = \sum_{i=1}^{n} x_{i,j}\, e_i, \quad 1 \leqslant j \leqslant p,$$

avec des matrices colonnes de coordonnées

$$X_j = \begin{pmatrix} x_{1,j} \\ \vdots \\ x_{i,j} \\ \vdots \\ x_{n,j} \end{pmatrix}, \quad i = \text{indice ligne} \in \{1, \ldots, n\},$$

et on peut éventuellement ranger ces coordonnées en une matrice rectangulaire $X = (x_{i,j})_{1 \leqslant i \leqslant n,\, 1 \leqslant j \leqslant p}$. Maintenant, on a

$$\varphi(x_1, \ldots, x_j, \ldots, x_p) = \varphi\left(\sum_{i_1=1}^{n} x_{i_1,1}\, e_{i_1}, \ldots, \sum_{i_j=1}^{n} x_{i_j,j}\, e_{i_j}, \ldots, \sum_{i_p=1}^{n} x_{i_p,p}\, e_{i_p} \right),$$

car on est obligé d'utiliser des indices différents $i_1, \ldots, i_j, \ldots, i_p$ pour chacun des vecteurs $x_1, \ldots, x_j, \ldots, x_p$, ce qui amène à utiliser des sous-indices ! En développant par multilinéarité comme dans le cas $p = 2$, il vient

$$\varphi(x_1, \ldots, x_p) = \sum_{1 \leqslant i_1, \ldots, i_p \leqslant n} x_{i_1,1} \cdots x_{i_p,p}\, \varphi(e_{i_1}, \ldots, e_{i_p}).$$

Comme précédemment, si on introduit les coefficients

$$c_{i_1\ldots i_p} = \varphi(e_{i_1},\ldots,e_{i_p}), \tag{1.3.5}$$

on trouve une expression de la forme

$$\varphi(x_1,\ldots,x_p) = \sum_{1\leqslant i_1,\ldots,i_p\leqslant n} c_{i_1\ldots i_p}\, x_{i_1,1}\cdots x_{i_p,p}, \tag{1.3.6}$$

et réciproquement, toute expression de ce type définit une forme multilinéaire $\varphi : E^p \to \mathbb{K}$, les coefficients $c_{i_1\ldots i_p}$ étant déterminés de manière unique par (1.3.5). Pour obtenir une forme multilinéaire, il est nécessaire que chaque vecteur colonne X_j contribue exactement un terme $x_{i_j,j}$ dans le membre de droite de (1.3.6) (de même qu'il faut exactement une coordonnée x_i et une coordonnée y_j dans l'expression (1.3.2) d'une forme bilinéaire). Dans le cas d'une forme mutilinéaire, la symétrie est l'antisymétrie se lisent comme suit sur les coefficients.

1.3.7. Propriétés. *La forme $\varphi : E^p \to \mathbb{K}$ est*

(a) symétrique *si et seulement si les coefficients vérifient $c_{i_{\sigma(1)}\ldots i_{\sigma(p)}} = c_{i_1\ldots,i_p}$ pour tous $i_1,\ldots,i_p \in \{1,\ldots,n\}$ et toute permutation $\sigma \in \mathfrak{S}_p$.*

(b) antisymétrique *si et seulement si les coefficients vérifient la relation $c_{i_{\sigma(1)}\ldots i_{\sigma(p)}} = \varepsilon(\sigma)\, c_{i_1\ldots,i_p}$ pour tous $i_1,\ldots,i_p \in \{1,\ldots,n\}$ et toute permutation $\sigma \in \mathfrak{S}_p$.*

(c) alternée *si et seulement la condition* (b) *d'antisymétrie est satisfaite, et si de plus $c_{i_1\ldots,i_p} = 0$ chaque fois qu'on a $i_j = i_k$ pour au moins deux indices $j \neq k$, $1 \leqslant j,k \leqslant p$.*

2. Formes n-multilinéaires alternées et déterminants

2.1. Expression des formes n-multilinéaires alternées

On suppose ici que E est un espace vectoriel de dimension finie n, muni d'une base $(e_1,\ldots,e_n)$, et on cherche à déterminer les applications multilinéaires alternées $\varphi : E^n \to \mathbb{K}$ (le cas $p = n$ est plus simple que le cas où $p \neq n$, voir l'exercice ??? pour le cas général – et pour la culture !). Soient $x_1,\ldots,x_n \in E$ des vecteurs représentés par une *matrice carrée* $X \in \mathcal{M}_{n\times n}(\mathbb{K})$ ayant pour colonnes les $X_j = (x_{i,j})_{1\leqslant i\leqslant n}$. On a ici

$$\varphi(x_1,\ldots,x_n) = \sum_{1\leqslant i_1,\ldots,i_n\leqslant n} x_{i_1,1}\cdots x_{i_n,n}\, \varphi(e_{i_1},\ldots,e_{i_n}).$$

Comme φ est alternée, on a $\varphi(e_{i_1},\ldots,e_{i_n}) = 0$ sauf lorsque les indices $i_1,\ldots,i_n$ sont 2 à 2 distincts, ce qui impose en fait qu'il existe une permutation $\sigma \in \mathfrak{S}_n$ telle que $i_1 = \sigma(1),\ldots,\ i_n = \sigma(n)$. Mais comme φ est antisymétrique, on a alors

$$\varphi(e_{\sigma(1)},\ldots,e_{\sigma(n)}) = \varepsilon(\sigma)\,\varphi(e_1,\ldots,e_n).$$

On n'a donc plus à considérer qu'un seul coefficient $c = \varphi(e_1, \dots, e_n) \in \mathbb{K}$, et il vient

$$\varphi(x_1, \dots, x_n) = c \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, x_{\sigma(1),1} \cdots x_{\sigma(n),n}. \tag{2.1.1}$$

La sommation apparaissant dans le membre de droite est par définition le *déterminant* de la matrice $X = (x_{i,j})$.

2.1.2. Définition. *Soit* $X = (x_{i,j})_{1 \leqslant i,j \leqslant n} \in \mathcal{M}_{n\times n}(\mathbb{K})$ *une* matrice carrée, *et pour* $j \in \{1, \dots, n\}$, *soient* $X_j = (x_{i,j})_{1 \leqslant i \leqslant n}$ *les vecteurs colonnes de* X. *On appelle* déterminant *de* X, *noté* $\det(X)$ *ou encore* $\det(X_1, \dots, X_n)$, *l'expression*

$$\det(X) = \det(X_1, \dots, X_n) := \sum_{\sigma \in \mathfrak{S}_n} \varepsilon(\sigma)\, x_{\sigma(1),1} \cdots x_{\sigma(n),n} \in \mathbb{K}.$$

Les calculs que nous avons menés ci-dessus aboutissent à la conclusion suivante.

2.1.3. Théorème. *Si* E *est un espace vectoriel de dimension* n *muni d'une base* $(e_1, \dots, e_n)$, *les formes* n*-multilinéaires alternées* $\varphi : E^n \to \mathbb{K}$ *s'écrivent toutes*

$$\varphi(x_1, \dots, x_n) = c \det(X_1, \dots, X_n)$$

où les X_j *sont les matrices colonnes des vecteurs* x_j *dans la base* $(e_i)_{1 \leqslant i \leqslant n}$ *et où* $c \in \mathbb{K}$ *est le coefficient donné par* $c = \varphi(e_1, \dots, e_n)$.

2.1.4. Notation. Dans la pratique, et particulièrement lorsqu'on a affaire à des matrices numériques, un déterminant se note également en remplaçant les parenthèses par des barres latérales verticales. Si $X = (x_{i,j})_{1 \leqslant i,j \leqslant n}$, on note ainsi

$$\det(X) = \begin{vmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,n} \\ x_{2,1} & x_{2,2} & \dots & x_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n,1} & x_{n,2} & \dots & x_{n,n} \end{vmatrix}.$$

Insistons sur le fait qu'il s'agit nécessairement d'un *tableau carré* ($p = n$), le déterminant d'une matrice non carrée n'a pas de sens !

2.1.5. Calcul en dimensions 1 et 2. Si $n = 1$ on a $\mathfrak{S}_1 = \{\mathrm{Id}\}$ et il est trivial que $\det((x)) = x$ pour $(x) \in \mathcal{M}_{1\times 1}(\mathbb{K})$. Pour $n = 2$, on a $\mathfrak{S}_2 = \{\mathrm{Id}, \tau_{1,2}\}$ et on voit que la somme à 2 termes $\sum_{\sigma \in \mathfrak{S}_2} \varepsilon(\sigma)\, x_{\sigma(1),1} x_{\sigma(2),2}$ donne

$$\begin{vmatrix} x_{1,1} & x_{1,2} \\ x_{2,1} & x_{2,2} \end{vmatrix} = x_{1,1}x_{2,2} - x_{2,1}x_{1,2}\ ,$$

car $\sigma = \mathrm{Id}$ fournit le terme $x_{1,1}x_{2,2}$ (diagonale principale, descendante), et $\sigma = \tau_{2,1}$ le terme $-x_{2,1}x_{1,2}$ (diagonale montante).

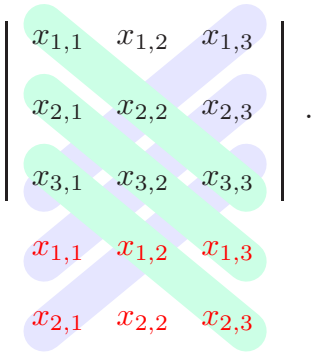
2.1.6. Calcul en dimension 3. On a $3! = 6$ permutations $\sigma \in \mathfrak{S}_3$, et la somme des termes $\varepsilon(\sigma)\, x_{\sigma(1),1} x_{\sigma(2),2} x_{\sigma(3),3}$ donne

$$\begin{vmatrix} x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,1} & x_{3,2} & x_{3,3} \end{vmatrix} = \begin{array}{l} x_{1,1}x_{2,2}x_{3,3} + x_{2,1}x_{3,2}x_{1,3} + x_{3,1}x_{1,2}x_{2,3} \\ -x_{2,1}x_{1,2}x_{3,3} - x_{1,1}x_{3,2}x_{2,3} - x_{3,1}x_{2,2}x_{1,3} \end{array},$$

car les différentes permutations paires et impaires correspondent respectivement aux termes

$$\begin{array}{lll} \sigma = \mathrm{Id} & : & x_{1,1}x_{2,2}x_{3,3} \\ \sigma = c & : & x_{2,1}x_{3,2}x_{1,3} \\ \sigma = c^2 & : & x_{3,1}x_{1,2}x_{2,3} \\ \sigma = \tau_{1,2} & : & -x_{2,1}x_{1,2}x_{3,3} \\ \sigma = \tau_{2,3} & : & -x_{1,1}x_{3,2}x_{2,3} \\ \sigma = \tau_{1,3} & : & -x_{3,1}x_{2,2}x_{1,3}\ . \end{array}$$

Une règle mnémotechnique, valable seulement en dimension 3, est la règle dite de Sarrus, qui peut se traduire comme suit : on récrit les deux premières lignes au dessous du déterminant et on prend les 6 diagonales possibles, avec le signe $+$ pour les diagonales descendantes et le signe $-$ pour les diagonales montantes :



(On peut aussi alternativement récrire les deux premières colonnes à la droite du déterminant, et procéder de même sur 3 lignes et 5 colonnes ; ou encore ajouter la dernière ligne au dessus et la première ligne en dessous ; ou la dernière colonne à gauche et la première colonne à droite, etc.)

En dimension $n \geqslant 4$, le nombre $n!$ de termes croît très vite, et l'usage de la formule de définition devient inefficace. On verra plus loin des méthodes plus rapides (par exemple, usage de combinaisons linéaires de lignes ou de colonnes).

2.1.7. Premières propriétés des déterminants. Dans ce qui suit, on identifie $\mathbb{K}^n$ à l'espace isomorphe $\mathcal{M}_{n\times 1}(\mathbb{K})$ des vecteurs colonnes de taille n, et on considère l'application déterminant

$$\det : \mathbb{K}^n \times \cdots \times \mathbb{K}^n \to \mathbb{K}, \quad (X_1, \ldots, X_n) \longmapsto \det(X_1, \ldots, X_n).$$

(a) det *est n-multilinéaire.*

Démonstration. Cela résulte du fait que la définition 2.1.2 est un cas particulier de la forme générale (1.3.6) des applications multilinéaires (somme de termes comportant exactement un facteur dans chaque vecteur colonne).

(b) det *est une application multilinéaire alternée* (et donc antisymétrique).

Démonstration. Supposons $X_j = X_k$ avec $j \neq k$, disons $j < k$, et considérons la transposition $\tau = \tau_{jk}$. En séparant la sommation suivant les permutations de signature $+1$ et -1, on peut écrire

$$\det(X) = \sum_{\sigma\in\mathfrak{S}_n^+} x_{\sigma(1),1}\cdots x_{\sigma(n),n} - \sum_{\sigma'\in\mathfrak{S}_n^-} x_{\sigma'(1),1}\cdots x_{\sigma'(n),n}.$$

Mais on peut parcourir toutes les permutations négatives $\sigma' \in \mathfrak{S}_n^-$ en prenant $\sigma' = \sigma\circ\tau$, $\sigma\in\mathfrak{S}_n^+$. Or $\sigma'(j) = \sigma(\tau(j)) = \sigma(k)$ et $\sigma'(k) = \sigma(\tau(k)) = \sigma(j)$, donc le terme $x_{\sigma'(1),1}\cdots x_{\sigma'(n),n}$ s'écrit

$$x_{\sigma'(1),1}\cdots x_{\sigma'(j),j}\cdots x_{\sigma'(k),k}\cdots x_{\sigma'(n),n} = x_{\sigma(1),1}\cdots x_{\sigma(k),j}\cdots x_{\sigma(j),k}\cdots x_{\sigma(n),n}.$$

Comme $X_j = X_k$, ce terme coïncide avec $x_{\sigma(1),1}\cdots x_{\sigma(k),k}\cdots x_{\sigma(j),j}\cdots x_{\sigma(n),n}$ qui (par commutativité de la multiplication) est égal à $x_{\sigma(1),1}\cdots x_{\sigma(n),n}$. On a donc bien $\det(X) = 0$. Il résulte de (a) et (b) que le théorème 2.1.3 fournit bien des formes φ n-multilinéaires alternées (à ce stade, on ne savait pas encore que les solutions trouvées au théorème 2.1.3 convenaient, mis à part le cas évident $\varphi = 0$!)

(c) *Pour toute matrice* $X = (x_{i,j})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_{n\times n}(\mathbb{K})$, *on a* $\det({}^tX) = \det(X)$.

Démonstration. Posons $X' = {}^tX = (x'_{i,j})$, avec $x'_{i,j} = x_{j,i}$. On a

$$\det({}^tX) = \det(X') = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma)\, x'_{\sigma(1),1}\cdots x'_{\sigma(n),n} = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma)\, x_{1,\sigma(1)}\cdots x_{n,\sigma(n)}.$$

Désignons par $f_j = x_{j,\sigma(j)}$ les facteurs des termes de la sommation. Comme la multiplication est commutative et que σ^{-1} est une bijection de $\{1,2,\ldots,n\}$, on a

$$x_{1,\sigma(1)}\cdots x_{n,\sigma(n)} = \prod_{j=1}^n x_{j,\sigma(j)} = \prod_{j=1}^n f_j = \prod_{j=1}^n f_{\sigma^{-1}(j)} = \prod_{j=1}^n x_{\sigma^{-1}(j),j}.$$

Ceci donne

$$\det({}^tX) = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma)\, x_{\sigma^{-1}(1),1}\cdots x_{\sigma^{-1}(n),n}.$$

Maintenant, on utilise le fait que $\sigma\mapsto\sigma' = \sigma^{-1}$ est une bijection de $\mathfrak{S}_n$, et aussi le fait que $\varepsilon(\sigma') = \varepsilon(\sigma^{-1}) = (\varepsilon(\sigma))^{-1} = \varepsilon(\sigma)$, pour voir que

$$\det({}^tX) = \sum_{\sigma'\in\mathfrak{S}_n} \varepsilon(\sigma')\, x_{\sigma'(1),1}\cdots x_{\sigma'(n),n} = \det(X). \qquad \square$$

2.2. Méthodes pratiques de calcul des déterminants

2.2.1. Calcul par blocs. Considérons une matrice $M \in \mathcal{M}_{n\times n}(\mathbb{K})$ de la forme

$$M = \begin{pmatrix} M' & R \\ O & M'' \end{pmatrix}$$

où M' et M'' sont des matrices carrées $p\times p$ et $(n-p)\times(n-p)$, et R une matrice rectangulaire $p\times(n-p)$. Si $M=(m_{i,j})$, on a $m_{i,j}=0$ pour $i\in\{p+1,\dots,n\}$ et $j\in\{1,\dots,p\}$. Dans le calcul de

$$\det(M) = \sum_{\sigma\in\mathfrak{S}_n} \varepsilon(\sigma)\, m_{\sigma(1),1}\cdots m_{\sigma(p),p}\, m_{\sigma(p+1),p+1}\cdots m_{\sigma(n),n},$$

on n'obtient donc des termes non nuls que lorsque $\sigma(j)\in\{1,\dots,p\}$ pour $j\in\{1,\dots,p\}$. Ceci implique que σ se décompose en une permutation σ' de $A'=\{1,\dots,p\}$ et une permutation σ'' de $A''=\{p+1,\dots,n\}$. Il est clair que le nombre d'inversions de σ est donné par $N(\sigma)=N(\sigma')+N(\sigma'')$, donc $\varepsilon(\sigma)=\varepsilon(\sigma')\varepsilon(\sigma'')$. On obtient alors

$$\begin{aligned}\det(M) &= \sum_{\sigma'\in\mathfrak{S}_{A'}}\ \sum_{\sigma''\in\mathfrak{S}_{A''}} \varepsilon(\sigma')\varepsilon(\sigma'')\, m_{\sigma'(1),1}\cdots m_{\sigma'(p),p}\, m_{\sigma''(p+1),p+1}\cdots m_{\sigma''(n),n}\\ &= \det(M')\times\det(M'').\end{aligned}$$

Plus généralement, si

$$M = \begin{pmatrix} M_1 & R_{12} & \dots & R_{1s} \\ O & M_2 & \dots & R_{2s} \\ \vdots & & \ddots & \vdots \\ O & O & \dots & M_s \end{pmatrix}$$

est "triangulaire supérieure par blocs" (avec les M_j carrées $p_j\times p_j$ et les R_{ij} rectangulaires de tailles $p_i\times p_j$), on peut appliquer un raisonnement par récurrence sur s en écrivant

$$M = \begin{pmatrix} M_1 & R \\ O & M' \end{pmatrix} \quad\text{où}\quad M' = \begin{pmatrix} M_2 & \dots & R_{2s} \\ \vdots & \ddots & \vdots \\ O & \dots & M_s \end{pmatrix},$$

ce qui implique inductivement

$$\det(M) = \det(M_1)\det(M') = \det(M_1)\det(M_2)\cdots\det(M_s).$$

On obtient ainsi la formule

$$\det\begin{pmatrix} M_1 & R_{12} & \dots & R_{1s} \\ O & M_2 & \dots & R_{2s} \\ \vdots & & \ddots & \vdots \\ O & O & \dots & M_s \end{pmatrix} = \det(M_1)\det(M_2)\cdots\det(M_s). \tag{2.2.2}$$

Pour une matrice M triangulaire inférieure par blocs et de blocs diagonaux M_j, on peut observer que la transposée ${}^tM$ est triangulaire supérieure par blocs, avec les ${}^tM_j$ comme blocs diagonaux. L'invariance du déterminant par passage à la transposée implique alors de manière analogue

$$\det \begin{pmatrix} M_1 & O & \dots & O \\ R_{21} & M_2 & \dots & O \\ \vdots & & \ddots & \vdots \\ R_{s1} & R_{s2} & \dots & M_s \end{pmatrix} = \det(M_1)\det(M_2)\cdots\det(M_s). \tag{2.2.3}$$

2.2.4. Corollaire. *Le déterminant d'une matrice triangulaire (et a fortiori d'une matrice diagonale) est égal au produit des coefficients diagonaux.*

2.2.5. Calcul d'un déterminant par combinaisons linéaires. Pour cela, on commence par faire les observations (a), (b), (c) suivantes.

(a) Si on transpose deux colonnes dans un déterminant $\det(X_1,\dots,X_n)$, le déterminant change de signe, et plus généralement, pour une permutation $\sigma \in \mathfrak{S}_n$, on a $\det(X_{\sigma(1)},\dots,X_{\sigma(n)}) = \varepsilon(\sigma)\det(X_1,\dots,X_n)$, en raison du fait que det est une forme multilinéaire antisymétrique.

(b) Le déterminant $\det(X_1,\dots,X_n)$ reste inchangé si une ajoute à une colonne X_j une combinaison linéaire $\sum_{k\neq j}\lambda_k X_k$ des autres colonnes. En effet, par multilinéarité,

$$\begin{aligned}&\det(X_1,\dots,X_j+\sum_{k\neq j}\lambda_k X_k,\dots,X_n)\\ &\quad=\det(X_1,\dots,X_j,\dots,X_n)+\sum_{k\neq j}\lambda_k\det(X_1,\dots,X_k,\dots X_k,\dots X_n),\end{aligned}$$

et les termes de la sommation sont nuls à cause du fait que det est une application multilinéaire alternée. Si un facteur commun α apparaît dans les coefficients d'une colonne X_j, i.e. $X_j = \alpha X'_j$, on peut en profiter pour "sortir" α :

$$\det(X_1,\dots,\alpha X'_j,\dots,X_n) = \alpha\,\det(X_1,\dots,X'_j,\dots,X_n).$$

(c) Par passage à la transposée, on a des résultats analogues pour les lignes : une permutation σ des lignes induit un facteur $\varepsilon(\sigma)$, le déterminant ne change pas si on ajoute à une ligne L_i une combinaison linéaire $\sum_{k\neq i}\lambda_k L_k$ des autres lignes, et enfin on peut "sortir" un facteur commun α qui apparaît dans une ligne.

(d) L'une des stratégies de calcul les plus efficaces d'un déterminant est la méthode du pivot de Gauss : on effectue des combinaisons linéaires (et éventuellement des permutations de colonnes ou de lignes) pour se ramener à une matrice triangulaire, auquel cas le calcul du déterminant est obtenu par 2.2.2, 2.2.3 ou 2.2.4. Donnons comme exemple le calcul du déterminant $D = \det(X_1,X_2,X_3,X_4)$ suivant :

$$D = \begin{vmatrix} 2 & -1 & 3 & -5 \\ 2 & 2 & -2 & 1 \\ -3 & 4 & 2 & -1 \\ -2 & -1 & 6 & -3 \end{vmatrix}.$$

Il est plus simple d'utiliser des colonnes ayant des coefficients ± 1 pour éviter les calculs fractionnaires. On va par exemple utiliser X_2 dont le premier coefficient est -1. On ajoutera $2X_2$ à X_1, $3X_2$ à X_3 et $-5X_2$ à X_4 pour obtenir

$$D = \begin{vmatrix} 2 & -1 & 3 & -5 \\ 2 & 2 & -2 & 1 \\ -3 & 4 & 2 & -1 \\ -2 & -1 & 6 & -3 \end{vmatrix} = \begin{vmatrix} 0 & -1 & 0 & 0 \\ 6 & 2 & 4 & -9 \\ 5 & 4 & 14 & -21 \\ -4 & -1 & 3 & 2 \end{vmatrix} = - \begin{vmatrix} -1 & 0 & 0 & 0 \\ 2 & 6 & 4 & -9 \\ 4 & 5 & 14 & -21 \\ -1 & -4 & 3 & 2 \end{vmatrix}.$$

(La dernière égalité résulte de la permutation des deux premières colonnes). Il s'agit maintenant d'une matrice triangulaire inférieure par blocs, avec un bloc 1×1 et un bloc 3×3, d'où

$$D = -(-1) \times \begin{vmatrix} 6 & 4 & -9 \\ 5 & 14 & -21 \\ -4 & 3 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -10 & 12 \\ 5 & 14 & -21 \\ -4 & 3 & 2 \end{vmatrix}.$$

(La dernière égalité s'obtenant en retranchant la deuxième ligne à la première, pour faire apparaître un coefficient 1). Pour ce nouveau déterminant $\det(X_1', X_2', X_3')$, on ajoute maintenant $10X_1'$ à X_2' et $-12X_1'$ à X_3' :

$$D = \begin{vmatrix} 1 & 0 & 0 \\ 5 & 64 & -81 \\ -4 & -37 & 50 \end{vmatrix} = \begin{vmatrix} 64 & -81 \\ -37 & 50 \end{vmatrix} = 64 \times 50 - 37 \times 81 = 203.$$

C'est à l'évidence beaucoup plus rapide et moins risqué que d'ajouter les 24 produits de termes constituant D !

2.2.6. Développement d'un déterminant suivant une colonne (ou suivant une ligne). Soit à calculer le déterminant d'une matrice $M = (m_{i,j})_{1 \leqslant i,j \leqslant n}$. Pour effectuer un développement de $\det(M)$ suivant la j-ième colonne, on exprime celle-ci dans la base canonique de $\mathcal{M}_{n \times 1}(\mathbb{K})$ en écrivant

$$e_i = \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}_{i=1,\ldots,n}, \quad X_j = \begin{pmatrix} m_{1,j} \\ \vdots \\ m_{i,j} \\ \vdots \\ m_{n,j} \end{pmatrix} = \sum_{i=1}^{n} m_{i,j} e_i,$$

ce qui donne par multilinéarité

$$\begin{aligned} \det(M) &= \det(X_1, \ldots, X_j, \ldots, X_n) \\ &= \sum_{i=1}^{n} m_{i,j} \det(X_1, \ldots, X_{j-1}, e_i, X_{j+1}, \ldots, X_n) \\ &= \sum_{i=1}^{n} m_{i,j} \begin{vmatrix} m_{1,1} & \ldots & m_{1,j-1} & 0 & m_{1,j+1} & \ldots & m_{1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ m_{i-1,1} & \ldots & m_{i-1,j-1} & 0 & m_{i-1,j+1} & \ldots & m_{i-1,n} \\ m_{i,1} & \ldots & m_{i,j-1} & 1 & m_{i,j+1} & \ldots & m_{i,n} \\ m_{i+1,1} & \ldots & m_{i+1,j-1} & 0 & m_{i+1,j+1} & \ldots & m_{i+1,n} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ m_{n,1} & \ldots & m_{n,j-1} & 0 & m_{n,j+1} & \ldots & m_{n,n} \end{vmatrix}. \end{aligned}$$

Dans le dernier déterminant, on ramène maintenant le coefficient 1 de la position (i,j) à la position $i'=j'=1$, ce qui implique de faire $i-1$ transpositions de lignes consécutives et $j-1$ transpositions de colonnes, et introduit donc un signe $(-1)^{(i-1)+(j-1)}=(-1)^{i+j}$. On aboutit alors à une matrice triangulaire supérieure par blocs $\begin{pmatrix} 1 & R \\ O & M'_{i,j} \end{pmatrix}$, où $M'_{i,j}$ est la matrice $(n-1)\times(n-1)$ obtenue en rayant la ligne i et la colonne j dans la matrice M. Il vient ainsi

$$\det(M)=\sum_{i=1}^{n} m_{i,j}\times(-1)^{i+j}\begin{vmatrix} m_{1,1} & \dots & m_{1,j-1} & m_{1,j+1} & \dots & m_{1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ m_{i-1,1} & \dots & m_{i-1,j-1} & m_{i-1,j+1} & \dots & m_{i-1,n} \\ m_{i+1,1} & \dots & m_{i+1,j-1} & m_{i+1,j+1} & \dots & m_{i+1,n} \\ \vdots & & \vdots & \vdots & & \vdots \\ m_{n,1} & \dots & m_{n,j-1} & m_{n,j+1} & \dots & m_{n,n} \end{vmatrix},$$

(“développement de $\det(M)$ suivant la j-ième colonne”), ce qui permet a priori de ramener le calcul de $\det(M)$ à celui de n déterminants de taille $(n-1)\times(n-1)$.

2.2.7. Définition. *On appelle cofacteur $\widetilde{m}_{i,j}$ de la matrice $M=(m_{i,j})_{1\leqslant i,j\leqslant n}$ associé à la ligne i et la colonne j la quantité*

$$\widetilde{m}_{ij}=(-1)^{i+j}\det(M'_{i,j}),$$

où $M'_{i,j}$ est la matrice $(n-1)\times(n-1)$ obtenue à partir de M en rayant la ligne i et la colonne j.

Les calculs effectués au paragraphe 2.2.6 aboutissent plus généralement au théorème suivant.

2.2.8. Théorème. *Soit $M=(m_{i,j})_{1\leqslant i,j\leqslant n}$ une matrice carrée et $\widetilde{m}_{i,j}$ ses cofacteurs. On a les relations*

(a) $\det(M)=\sum_{i=1}^{n} m_{i,j}\,\widetilde{m}_{i,j}$ *(développement suivant la colonne j)* ;

(b) $\det(M)=\sum_{j=1}^{n} m_{i,j}\,\widetilde{m}_{i,j}$ *(développement suivant la ligne i)* ;

(c) *pour $j\neq j'$ dans $\{1,\dots n\}$,* $\sum_{i=1}^{n} m_{i,j}\,\widetilde{m}_{i,j'}=0$

(d) *pour $i\neq i'$ dans $\{1,\dots n\}$,* $\sum_{j=1}^{n} m_{i,j}\,\widetilde{m}_{i',j}=0$.

Démonstration. La formule (a) a déjà été établie, et (b) s'en déduit en appliquant le résultat à la matrice transposée ${}^tM$. Pour vérifier (c), on considère la matrice $P=(p_{k,\ell})$ obtenue à partir de M en remplaçant la colonne j' par la colonne j

(qui est donc dupliquée). On a par conséquent $\det(P) = 0$. Mais par définition de P on a aussi $p_{i,j'} = m_{i,j}$ et $\widetilde{p}_{i,j'} = \widetilde{m}_{i,j'}$ (la colonne j' étant rayée, les déterminants $(n-1)\times(n-1)$ issus de M et de P sont les mêmes !), donc

$$\det(P) = 0 = \sum_{i=1}^{n} p_{i,j'}\,\widetilde{p}_{i,j'} = \sum_{i=1}^{n} m_{i,j}\,\widetilde{m}_{i,j'}.$$

La formule (d) s'obtient de même en remplaçant la ligne i' par la ligne i dupliquée. □

2.2.9. Exemple de développement par rapport à une ligne ou une colonne. Soit à évaluer

$$D = \begin{vmatrix} 2 & 0 & 1 & -3 \\ -1 & 4 & -7 & 2 \\ 0 & 3 & 5 & 0 \\ -2 & 1 & 0 & 6 \end{vmatrix}.$$

Pour minimiser le nombre de cofacteurs à calculer, on a intérêt à choisir une colonne ou une ligne présentant beaucoup de coefficients nuls. On choisira ici la troisième ligne, ce qui (en tenant compte des signes $(-1)^{i+j}$) donne

$$D = -3\begin{vmatrix} 2 & 1 & -3 \\ -1 & -7 & 2 \\ -2 & 0 & 6 \end{vmatrix} + 5\begin{vmatrix} 2 & 0 & -3 \\ -1 & 4 & 2 \\ -2 & 1 & 6 \end{vmatrix}$$

avec (même méthode, on utilise respectivement la troisième puis la première ligne)

$$\begin{vmatrix} 2 & 1 & -3 \\ -1 & -7 & 2 \\ -2 & 0 & 6 \end{vmatrix} = -2\begin{vmatrix} 1 & -3 \\ -7 & 2 \end{vmatrix} + 6\begin{vmatrix} 2 & 1 \\ -1 & -7 \end{vmatrix} = (-2)(-19) + 6(-13) = -40,$$

$$\begin{vmatrix} 2 & 0 & -3 \\ -1 & 4 & 2 \\ -2 & 1 & 6 \end{vmatrix} = 2\begin{vmatrix} 4 & 2 \\ 1 & 6 \end{vmatrix} - 3\begin{vmatrix} -1 & 4 \\ -2 & 1 \end{vmatrix} = 2\times 22 - 3\times 7 = 23.$$

En définitive, on trouve

$$D = (-3)(-40) + 5\times 23 = 235.$$ □

2.2.10. Définition. *Si $M = (m_{i,j})_{1\leqslant i,j\leqslant n}$ est une matrice $n\times n$ et si les $\widetilde{m}_{i,j}$ en sont les cofacteurs, la matrice*

$$\operatorname{comat}(M) := \widetilde{M} = (\widetilde{m}_{i,j})_{1\leqslant i,j\leqslant n} \in \mathcal{M}_{n\times n}(\mathbb{K})$$

est appelée comatrice *de M.*

2.2.11. Exemples de comatrices. On n'oubliera pas de prêter attention aux signes $(-1)^{i+j}$ dans ce qui suit :

$$\operatorname{comat}\begin{pmatrix} a & c \\ b & d \end{pmatrix} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},$$

$$\operatorname{comat}\begin{pmatrix} a & d & g \\ b & e & h \\ c & f & k \end{pmatrix} = \begin{pmatrix} ek-fh & ch-bk & bf-ce \\ fg-dk & ak-cg & cd-af \\ dh-eg & bg-ah & ae-bd \end{pmatrix}. \qquad \square$$

Avec cette notation, le théorème 2.2.8 peut se retraduire de manière synthétique comme suit.

2.2.12. Théorème. *Soit $M = (m_{i,j})_{1\leqslant i,j\leqslant n}$ une matrice $n \times n$ et*

$$\widehat{M} = (\widehat{m}_{i,j})_{1\leqslant i,j\leqslant n} = {}^t(\operatorname{comat}(M))$$

la transposée de sa comatrice (parfois appelée matrice adjointe et notée ${}^{ad}M$). Alors on a

$$M\widehat{M} = \widehat{M}M = \det(M)\cdot I_n$$

où I_n est la matrice unité de taille n.

Démonstration. Si $\widetilde{M} = (\widetilde{m}_{i,j}) = \operatorname{comat}(M)$, on a $\widehat{m}_{i,j} = \widetilde{m}_{j,i}$, par conséquent

$$\sum_{j=1}^{n} m_{i,j}\,\widehat{m}_{j,k} = \sum_{j=1}^{n} m_{i,j}\,\widetilde{m}_{k,j} = \begin{cases} \det(M) & \text{si } i=k \text{ (d'après 2.2.8 (b))} \\ 0 & \text{si } i\neq k \text{ (d'après 2.2.8 (d))}, \end{cases}$$

$$\sum_{i=1}^{n} \widehat{m}_{j,i}\,m_{i,k} = \sum_{i=1}^{n} \widetilde{m}_{i,j}\,m_{i,k} = \begin{cases} \det(M) & \text{si } j=k \text{ (d'après 2.2.8 (a))} \\ 0 & \text{si } j\neq k \text{ (d'après 2.2.8 (c))}. \end{cases}$$

Ces relations expriment bien les égalités matricielles annoncées. $\square$

2.2.13. Corollaire. *Si $\det(M)\neq 0$, la matrice M est inversible et on a*

$$M^{-1} = \frac{1}{\det(M)}\,\widehat{M} = \frac{1}{\det(M)}\,{}^t(\operatorname{comat}(M)). \qquad \square$$

En dépit de son "élégance théorique", cette formule est en pratique peu efficace pour calculer l'inverse d'une matrice (hormis peut-être le cas $n = 2$) ; très souvent, il vaut mieux utiliser la méthode du pivot pour résoudre le système linéaire $Y = MX$ et trouver la solution sous la forme $X = M'Y$. On a alors $M^{-1} = M'$. $\square$

3. Déterminant des endomorphismes

3.1. Déterminant d'un système de vecteurs dans une base

3.1.1. Définition. *Soit E un $\mathbb{K}$-espace vectoriel de dimension finie, et soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base de E. Si $(x_1, \ldots, x_n)$ est un système de n vecteurs de E et $X_1, \ldots, X_n$ sont les vecteurs colonnes coordonnées de $x_1, \ldots, x_n$ dans la base $\mathcal{B}$, on note*

$$\det_{\mathcal{B}}(x_1, \ldots, x_n) = \det(X_1, \ldots, X_n) = \det(X),$$

où X est la matrice dont les colonnes sont les X_j.

3.1.2. Attention. Dans la notation $\det_{\mathcal{B}}(x_1, \ldots, x_n)$, il convient de ne pas oublier de mentionner la base $\mathcal{B}$ dans laquelle on travaille, car les vecteurs colonnes coordonnées X_j dépendent de la base $\mathcal{B}$, et $\det(X_1, \ldots, X_n)$ aussi. □

Il est évident que $\det_{\mathcal{B}}(\mathcal{B}) = 1$, puisque la base $\mathcal{B}$ est représentée par rapport à elle-même par la matrice unité I_n. Le théorème 2.1.3 peut alors se reformuler comme suit.

3.1.3. Théorème. *Si $\varphi : E^n \to \mathbb{K}$ est une forme n-multilinéaire alternée, alors on peut écrire*

$$\varphi(x_1, \ldots, x_n) = c \det_{\mathcal{B}}(x_1, \ldots, x_n) \quad \text{où} \quad c = \varphi(\mathcal{B}) = \varphi(e_1, \ldots, e_n).$$

Autrement dit, les formes n-multilinéaires alternées sur E sont les multiples de $\det_{\mathcal{B}}$ par un facteur $c \in \mathbb{K}$ quelconque.

Démonstration. On sait par le théorème 2.1.3 que φ s'écrit sous cette forme, et il suffit de prendre $(x_1, \ldots, x_n) = (e_1, \ldots, e_n) = \mathcal{B}$ pour trouver la valeur de c. □

3.1.4. Formule de changement de base pour les déterminants. *Soient $\mathcal{B} = (e_1, \ldots, e_n)$ et $\mathcal{B}' = (e'_1, \ldots, e'_n)$ des bases de E.*

(a) *On a la relation*

$$\det_{\mathcal{B}'}(x_1, \ldots, x_n) = \det_{\mathcal{B}'}(\mathcal{B}) \times \det_{\mathcal{B}}(x_1, \ldots, x_n) \quad \text{où} \quad \det_{\mathcal{B}'}(\mathcal{B}) \times \det_{\mathcal{B}}(\mathcal{B}') = 1.$$

(b) *Si P est la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$, on a*

$$\det_{\mathcal{B}}(\mathcal{B}') = \det(P) \neq 0, \quad \det_{\mathcal{B}'}(\mathcal{B}) = \det(P^{-1}) = (\det(P))^{-1}.$$

Démonstration. (a) est un cas particulier de 3.1.3 avec $\varphi = \det_{\mathcal{B}'}$. L'égalité $\det_{\mathcal{B}'}(\mathcal{B}) \times \det_{\mathcal{B}}(\mathcal{B}') = 1$ s'obtient en prenant $(x_1, \ldots, x_n) = \mathcal{B}' = (e'_1, \ldots, e'_n)$.

(b) La formule $\det_{\mathcal{B}}(\mathcal{B}') = \det(P)$ résulte des définitions, $\det_{\mathcal{B}'}(\mathcal{B}) = \det(P^{-1})$ du fait que P^{-1} est la matrice de passage de $\mathcal{B}'$ à $\mathcal{B}$, et l'égalité avec $(\det(P))^{-1}$ découle de (a). □

3.2. Déterminant d'un endomorphisme

3.2.1. Théorème et définition. *Si $u \in \mathrm{End}_{\mathbb{K}}(E)$ est un endomorphisme d'un espace vectoriel E de dimension finie n, il existe un scalaire noté $\det(u) \in \mathbb{K}$ tel que pour tous vecteurs $x_1, \dots, x_n \in E$ et toute base $\mathcal{B}$ de E on ait*

$$(*) \qquad \det_{\mathcal{B}}(u(x_1), \dots, u(x_n)) = \det(u) \times \det_{\mathcal{B}}(x_1, \dots, x_n).$$

Le scalaire $\det(u)$ est indépendant de la base $\mathcal{B}$ et il est donné par

$$\det(u) = \det_{\mathcal{B}}(u(\mathcal{B})) = \det_{\mathcal{B}}(u(e_1), \dots, u(e_n)) = \det(M) \quad \text{où} \quad M = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(u).$$

Démonstration. On considère ici l'application

$$\varphi(x_1, \dots, x_n) = \det_{\mathcal{B}}(u(x_1), \dots, u(x_n)).$$

Il est immédiat de vérifier que c'est une application n-multilinéaire alternée (la multilinéarité est évidente et on a bien $\varphi(x_1, \dots, x_n) = 0$ si $x_i = x_j$, $i \neq j$). On peut donc appliquer le théorème 3.1.3, ce qui donne $\varphi = c \det_{\mathcal{B}}$ avec un facteur

$$c = \varphi(\mathcal{B}) = \det_{\mathcal{B}}(u(e_1), \dots, u(e_n)) = \det(M).$$

Le scalaire $\det(u) := c$ est indépendant de la base $\mathcal{B}$ choisie, car si on choisit une autre base $\mathcal{B}'$, on a une relation de la forme $\det_{\mathcal{B}'} = \lambda \det_{\mathcal{B}}$ avec $\lambda = \det_{\mathcal{B}'}(\mathcal{B}) \neq 0$, et la constante λ peut se simplifier dans l'égalité $(*)$ définissant $\det(u)$. □

Une conséquence importante est la multiplicativité du déterminant (qui serait difficile à vérifier directement sur la définition du déterminant des matrices !)

3.2.2. Théorème. *Soient $u, v \in \mathrm{End}_{\mathbb{K}}(E)$ des endomorphismes. Alors*

$$\det(u \circ v) = \det(u) \times \det(v).$$

De façon équivalente, si $M, N \in \mathcal{M}_{n\times n}(\mathbb{K})$, on a

$$\det(M \times N) = \det(M) \times \det(N).$$

Démonstration. Fixons une base $\mathcal{B}$, et considérons des vecteurs $x_1, \dots, x_n \in E$ quelconques. Par définition, on a d'une part

$$\det_{\mathcal{B}}((u \circ v)(x_1), \dots, (u \circ v)(x_n)) = \det(u \circ v) \times \det_{\mathcal{B}}(x_1, \dots, x_n),$$

et d'autre part cette quantité est aussi égale à

$$\begin{aligned} \det_{\mathcal{B}}\big(u(v(x_1)), \dots, u(v(x_n))\big) &= \det(u) \times \det_{\mathcal{B}}(v(x_1), \dots, v(x_n)) \\ &= \det(u) \times \det(v) \times \det_{\mathcal{B}}(x_1, \dots, x_n). \end{aligned}$$

La formule $\det(u \circ v) = \det(u) \times \det(v)$ s'en déduit en prenant $(x_1, \dots, x_n) = \mathcal{B}$, et la relation matricielle $\det(M \times N) = \det(M) \times \det(N)$ s'obtient en considérant les endomorphismes u, v de matrices associées $M = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(u)$, $N = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(v)$. □

3.2.3. Théorème. *Soit $u \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme d'un espace vectoriel E de dimension* finie. *Alors les propriétés suivantes sont équivalentes.*

(a) *u est inversible, i.e., il existe $v \in \mathrm{End}_{\mathbb{K}}$ tel que $u \circ v = v \circ u = \mathrm{Id}_E$;*

(a′) *u est inversible à droite, i.e. il existe $v \in \mathrm{End}_{\mathbb{K}}(E)$ tel que $u \circ v = \mathrm{Id}_E$;*

(a″) *u est inversible à gauche, i.e. il existe $v \in \mathrm{End}_{\mathbb{K}}(E)$ tel que $v \circ u = \mathrm{Id}_E$;*

(b) *u est bijectif ;*

(c) *u est injectif ;*

(d) *u est surjectif ;*

(e) $\det(u) \neq 0$.

Dans ce cas, on a $\det(u^{-1}) = (\det(u))^{-1}$.

Démonstration. On commence par observer qu'on a la chaîne d'implications

$$\mathrm{(a)} \Rightarrow \mathrm{(a)}' \Rightarrow \mathrm{(e)} \Rightarrow \mathrm{(a)}, \qquad \mathrm{(a)} \Rightarrow \mathrm{(a)}'' \Rightarrow \mathrm{(e)}.$$

En effet, les 4 premières à partir de la gauche sont évidentes (par exemple, pour $\mathrm{(a)}' \Rightarrow \mathrm{(e)}$, on utilise le fait que $u \circ v = \mathrm{Id}_E$ implique $\det(u) \times \det(v) = 1$, qui implique $\det(u) \neq 0$); pour $\mathrm{(e)} \Rightarrow \mathrm{(a)}$, on utilise le fait déjà démontré que $M \in \mathcal{M}_{n\times n}(\mathbb{K})$, $\det(M) \neq 0$ implique M inversible (corollaire 2.2.13)). Les propriétés (a), (a)′, (a)″, (e) sont donc équivalentes. Maintenant, on a également la chaîne d'implications évidentes

$$\mathrm{(a)} \Rightarrow \mathrm{(b)} \Rightarrow \begin{matrix}\mathrm{(c)}\\ \mathrm{(d)}\end{matrix}, \qquad \mathrm{(c)\ et\ (d)} \Rightarrow \mathrm{(b)} \Rightarrow \mathrm{(a)},$$

en effet $\mathrm{(b)} \Rightarrow \mathrm{(a)}$ provient du fait facile à vérifier que u $\mathbb{K}$-linéaire bijectif implique que $v = u^{-1}$ est aussi $\mathbb{K}$-linéaire. Pour vérifier que (a), (b), (c), (d) sont équivalentes, il reste alors seulement à voir que $\mathrm{(c)} \iff \mathrm{(d)}$, mais ceci résulte du théorème du rang : l'égalité $\dim E = \dim \mathrm{Ker}(u) + \dim \mathrm{Im}(u)$ nous dit que

$$\begin{aligned} u \text{ injectif} &\iff \mathrm{Ker}(u) = \{0\} \iff \dim \mathrm{Ker}(u) = 0 \\ &\iff \dim \mathrm{Im}(u) = \dim E \iff \mathrm{Im}(u) = E \iff u \text{ surjectif.} \end{aligned}$$

□

3.2.4. Remarque. Le théorème 3.2.3 est faux en dimension infinie. Prenons en effet $E = \mathbb{K}[X]$ (espace des polynômes de degré quelconque) et les applications u, v définies par

$$u : P \mapsto Q = u(P), \quad P(X) = \sum_{j=0}^{d} a_j X^j \mapsto Q(X) = XP(X) = \sum_{j=0}^{d} a_j X^{j+1},$$

$$v : Q \mapsto R = v(Q), \quad Q(X) = \sum_{j=0}^{d} b_j X^j \mapsto R(X) = \frac{Q(X) - Q(0)}{X} = \sum_{j=1}^{d} b_j X^{j-1}.$$

Il est facile de voir que $u, v \in \mathrm{End}_{\mathbb{K}}(E)$ et que

$$\begin{aligned} v \circ u &= \mathrm{Id}_E, \quad \text{mais} \\ u \circ v &\neq \mathrm{Id}_E, \quad u \circ v : Q \mapsto Q - Q(0), \end{aligned}$$

donc v est un inverse à gauche mais pas à droite de u. On voit aisément que u est injectif ($\mathrm{Ker}(u) = \{0\}$) mais pas surjectif ($\mathrm{Im}(u)$ ne contient pas les polynômes constants non nuls), tandis que v est surjectif mais pas injectif ($\mathrm{Ker}(v) = \{\text{polynômes constants}\}$). □

3.2.5. Remarque. Si $\mathcal{B} = (e_1, \ldots, e_n)$ est une base, pour qu'un système $\mathcal{A} = (a_1, \ldots, a_n)$ soit une base, il faut et il suffit que $\det_{\mathcal{B}}(a_1, \ldots, a_n) \neq 0$, en effet c'est également la condition nécessaire et suffisante pour que la matrice de passage P de $\mathcal{B}$ à $\mathcal{A}$ soit inversible.

3.2.6. Remarque. On peut (de façon moins élégante) vérifier l'indépendance du déterminant d'un endomorphisme u par rapport à la base $\mathcal{B}$ choisie, en observant que si $M = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(u)$ et $M' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(u)$ et si P est la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$, on a $M' = P^{-1}MP$. Il en résulte alors

$$\det(M') = \det(P^{-1}MP) = (\det(P))^{-1} \det(M) \det(P) = \det(M).$$

4. Application à la résolution des systèmes linéaires

L'objectif de cette section est de résoudre (au moins théoriquement) un système linéaire quelconque

$$(\Sigma) \qquad \begin{cases} a_{11}\, x_1 + \cdots + a_{1p} x_p = b_1 \\ \cdots \\ a_{n1} x_1 + \cdots + a_{np} x_p = b_n \end{cases}$$

à n équations et p inconnues $x_1, \ldots, x_p \in \mathbb{K}$, à coefficients $a_{ij} \in \mathbb{K}$, avec seconds membres les $b_i \in \mathbb{K}$, $1 \leqslant i \leqslant n$. On peut retraduire le système linéaire (Σ) sous la forme matricielle

$$AX = B,$$

avec inconnue $X = (x_i)_{1 \leqslant i \leqslant p} \in \mathcal{M}_{p \times 1}(\mathbb{K})$, coefficients et second membre

$$A = (a_{ij})_{1 \leqslant i \leqslant n, 1 \leqslant j \leqslant p} \in \mathcal{M}_{n \times p}(\mathbb{K}), \quad B = (b_i)_{1 \leqslant i \leqslant n} \in \mathcal{M}_{n \times 1}(\mathbb{K}).$$

4.1. Systèmes de Cramer

Un système de Cramer est par définition un système linéaire carré (c'est-à-dire tel que $p = n$), avec $\det(A) \neq 0$. Alors la matrice A est inversible, par conséquent (Σ) admet une solution unique

$$X = (x_i)_{1 \leqslant i \leqslant n} = A^{-1}B. \tag{4.1.1}$$

Il est parfois utile, au moins théoriquement, de connaître une formule explicite pour cette unique solution. Pour cela, on introduit les colonnes de coefficients a_{ij}, et on écrit le système (Σ) sous la forme équivalente qui suit :

$$A_j = \begin{pmatrix} a_{1j} \\ \vdots \\ a_{ij} \\ \vdots \\ a_{nj} \end{pmatrix}, \quad (\Sigma) \Longleftrightarrow \sum_{j=1}^{n} x_j A_j = B. \tag{4.1.2}$$

Si $X = (x_i)$ est la solution, la multilinéarité du déterminant implique pour tout $i \in \{1, \ldots, n\}$ que

$$\begin{aligned} \det(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots A_n) &= \det(A_1, \ldots, A_{i-1}, \sum_{j=1}^{n} x_j A_j, A_{i+1}, \ldots A_n) \\ &= x_i \det(A_1, \ldots, A_i, \ldots, A_n) \end{aligned}$$

(on utilise aussi le fait que det est alterné, ce qui entraîne que seul le terme $j = i$ intervient). On obtient alors :

4.1.3. Formules de Cramer. *L'unique solution d'un système de Cramer (i.e. $p = n$, $\det A \neq 0$) est donnée par la formule explicite*

$$x_i = \frac{\det(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots A_n)}{\det(A_1, \ldots, A_{i-1}, A_i, A_{i+1}, \ldots A_n)}, \quad 1 \leqslant i \leqslant n,$$

c'est-à-dire encore

$$x_i = \frac{\det(A_1, \ldots, A_{i-1}, B, A_{i+1}, \ldots A_n)}{\det(A)}, \quad 1 \leqslant i \leqslant n.$$

Ces formules élégantes sur le plan théorique, dues au mathématicien genevois Gabriel Cramer (1704–1752), sont malheureusement assez peu efficaces pour les calculs, et donc déconseillées en pratique !

4.2. Déterminants mineurs extraits et rang des applications linéaires

On considère maintenant le cas d'une matrice $A = (a_{ij})_{1 \leqslant i \leqslant n,\, 1 \leqslant j \leqslant p}$ rectangulaire, et on s'intéresse à son rang

$$r = \operatorname{rang}(A) = \dim S, \quad S = \operatorname{vect}(A_1, \ldots, A_p) \subset \mathcal{M}_{n \times 1}(\mathbb{K}) \simeq \mathbb{K}^n,$$

où les $(A_j)_{1 \leqslant j \leqslant p}$ sont les vecteurs colonnes de A. On peut alors extraire exactement r colonnes $(A_{j_1}, \ldots, A_{j_r})$ formant une base de S, et les autres colonnes A_j sont, de manière unique, combinaisons linéaires de $A_{j_1}, \ldots, A_{j_r}$.

4.2.1. Définition. *Soient* $I = \{i_1 < i_2 < \ldots < i_k\} \subset \{1, \ldots, p\}$ *et* $J = \{j_1 < j_2 < \ldots < j_k\} \subset \{1, \ldots, n\}$ *des parties formées chacune de* k *indices* ($k \leqslant \min(n, p)$). *On définit le déterminant mineur* $D_{I,J}$ *associé à la matrice rectangulaire* $A = (a_{ij})$ *et aux parties* I, J *comme étant le déterminant de taille* k

$$D_{I,J} = \det(a_{ij})_{i\in I,\, j\in J} = \det(a_{i_\ell j_m})_{1\leqslant \ell, m\leqslant k}.$$

L'objectif de ce paragraphe est de montrer le résultat suivant.

4.2.2. Théorème. *Si* $A \neq 0$, *le rang* $r = \operatorname{rang}(A)$ *est égal au maximum des entiers* $k \leqslant \min(p, q)$ *tel qu'il existe un déterminant mineur* $D_{I,J} \neq 0$ *de taille* k *extrait de* A (*et bien entendu* $r = 0$ *si* $A = 0$).

Démonstration. Supposons qu'il existe un déterminant mineur $D_{I,J} \neq 0$ de taille k extrait de A. Soit $A' = (a_{ij})_{i\in I,\, j\in J}$ la matrice correspondante. Alors $\det(A') = D_{I,J} \neq 0$ donc A' est inversible et ses colonnes $A'_{j_1}, \ldots, A'_{j_k}$ sont linéairement indépendantes. Mais ceci implique que les colonnes correspondantes $A_{j_1}, \ldots, A_{j_k}$ de A (qui ont éventuellement plus de lignes) sont linéairement indépendantes. Ceci implique $k \leqslant r$, puisque r est le nombre maximum de colonnes de A qui sont linéairement indépendantes.

Dans la direction inverse, choisissons des colonnes $(A_{j_1}, \ldots, A_{j_r})$ de A formant une base du sous-espace S. Désignons par $(e_i)_{1\leqslant i\leqslant n}$ la base canonique de $\mathcal{M}_{n\times 1}(\mathbb{K})$ (formée des matrices colonnes ayant un coefficient 1 à la i-ième ligne et des zéros ailleurs). On sait qu'on peut compléter $(A_{j_1}, \ldots, A_{j_r})$ en une base

$$(A_{j_1}, \ldots, A_{j_r}, e_{i_{r+1}}, \ldots, e_{i_n})$$

de $\mathcal{M}_{n\times 1}(\mathbb{K}) \simeq \mathbb{K}^n$, avec $1 \leqslant i_{r+1} < \cdots < i_n \leqslant n$. Posons $J = \{j_1, \ldots, j_r\}$ et

$$I = \{i_1, \ldots, i_r\} = \{1, \ldots, n\} \smallsetminus \{i_{r+1}, \ldots, i_n\}, \quad 1 \leqslant i_1 < \cdots < i_r \leqslant n.$$

Alors

$$0 \neq \det(A_{j_1}, \ldots, A_{j_r}, e_{i_{r+1}}, \ldots, e_{i_n})) = \pm \det(a_{ij})_{i\in I,\, j\in J}$$

en développant le déterminant suivant les $n-r$ colonnes e_{i_ℓ}, $\ell \geqslant r+1$. Ceci entraîne $D_{I,J} = \det(a_{ij})_{i\in I,\, j\in J} \neq 0$, et il s'agit d'un déterminant de taille $k = r$. □

4.2.3. Conséquences. *Soit* A *une matrice rectangulaire* $n \times p$. *Alors*

(a) $\operatorname{rang}({}^tA) = \operatorname{rang}(A)$.

(b) *Le rang* $r = \operatorname{rang}(A)$ *est le nombre maximum de colonnes* C_j *de* A *linéairement indépendantes, aussi bien que le nombre maximum de lignes* L_i *de* A *linéairement indépendantes.*

(c) *Si* $A \neq 0$ *et si* $D_{I,J} \neq 0$ *est un déterminant mineur extrait de taille* $r = \operatorname{rang}(A)$, $I = \{i_1, \ldots, i_r\}$ *et* $J = \{j_1, \ldots, j_r\}$, *alors les colonnes* C_j *de* A, $j \in \complement J$, *sont* (*de manière unique*) *combinaisons linéaires des colonnes* $C_{j_1}, \ldots, C_{j_r}$, *et les lignes* L_i *de* A, $i \in \complement I$, *sont* (*de manière unique*) *combinaisons linéaires des lignes* $L_{i_1}, \ldots, L_{i_r}$.

Démonstration. (a) est une conséquence directe du théorème 4.2.2 et du fait que $\det(M) = \det({}^tM)$ pour toute matrice carrée. Les assertions concernant les colonnes résultent des définitions, et on déduit les énoncés analogues pour les lignes en considérant ${}^tA$. □

4.2.4. Calcul du noyau. La donnée de la matrice A est équivalente à celle de l'application linéaire

$$u : \mathbb{K}^p \to \mathbb{K}^n \quad \text{écrite matriciellement} \quad X \mapsto Y = AX.$$

On a $r = \operatorname{rang}(A) = \dim \operatorname{Im}(u) = \dim \operatorname{vect}(A_1, \ldots, A_p)$, et donc

$$\dim \operatorname{Ker}(u) = p - \dim \operatorname{Im}(u) = p - r.$$

La recherche du noyau consiste à résoudre le système $AX = 0$. On notera $\operatorname{Ker}(A)$ ce noyau. Choisissons $D_{I,J} = \det(A') \neq 0$ un déterminant mineur extrait de A, de taille r. Quitte à permuter les lignes et les colonnes (et donc quitte à renuméroter les coordonnées x_j dans un autre ordre), on va pouvoir écrire

$$A = \begin{pmatrix} A' & A'' \\ M' & M'' \end{pmatrix} \quad \text{avec} \quad \begin{cases} A' & r \times r \\ A'' & r \times (p-r) \\ M' & (n-r) \times r \\ M' & (n-r) \times (p-r) \end{cases}, \quad X = \begin{pmatrix} X' \\ X'' \end{pmatrix}$$

avec X', X'' des colonnes de tailles respectives r et $p - r$, et donc

$$AX = \begin{pmatrix} A'X' + A''X'' \\ M'X' + M''X'' \end{pmatrix}.$$

On sait que les lignes de $(M' \quad M'')$ sont combinaisons linéaires des lignes de $(A' \quad A'')$, par conséquent le système $AX = 0$ est *équivalent* au système partiel $A'X' + A''X'' = 0$, i.e. $A'X' = -A''X''$. Mais ici la matrice A' est inversible, on voit alors qu'on peut choisir $X'' \in \mathcal{M}_{(p-r)\times 1}(\mathbb{K})$ arbitraire et qu'il suffit de prendre $X' = -(A')^{-1}A''X''$ (il le faut aussi !). Le noyau $\{X \,/\, AX = 0\}$ est donc l'ensemble des vecteurs colonnes

$$X = \begin{pmatrix} X' \\ X'' \end{pmatrix} = \begin{pmatrix} -(A')^{-1}A''X'' \\ X'' \end{pmatrix} = \begin{pmatrix} -(A')^{-1}A'' \\ I_{p-r} \end{pmatrix} X'', \quad X'' \in \mathcal{M}_{(p-r)\times 1}(\mathbb{K}).$$

On voit sur cette formule que le noyau est bien de dimension $(p - r)$. L'usage de la formule $(A')^{-1}A'' = \frac{1}{\det(A')}{}^t(\operatorname{comat}(A'))A''$ revient (en théorie) à appliquer les formules de Cramer pour résoudre $A'X' = -A''X''$. En pratique, il vaut mieux utiliser la méthode du pivot.

4.3. Compatibilité des systèmes linéaires

Pour résoudre un système linéaire avec second membre

$$(\Sigma) : \quad AX = B$$

il suffit de connaître une solution particulière X_{part} (s'il y en a) : en soustrayant $AX_{\text{part}} = B$ et en posant $\Xi = X - X_{\text{part}}$, on se ramène en effet à résoudre

$$A(X - X_{\text{part}}) = A\,\Xi = 0.$$

Ceci montre alors que les solutions de (Σ) sont les matrices colonnes

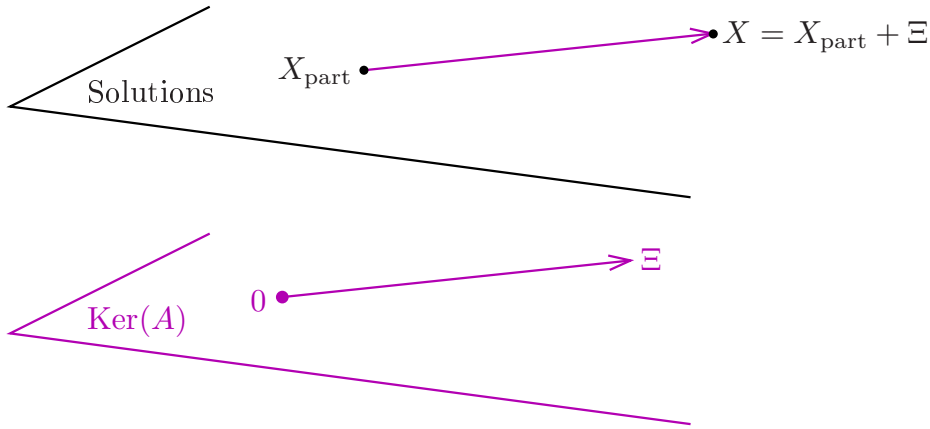
$$X = X_{\text{part}} + \Xi, \quad \Xi \in \operatorname{Ker}(A), \tag{4.3.1}$$

en on sait que $\dim \operatorname{Ker}(A) = p - r$ (avec les notations de 4.2.4). On peut énoncer

4.3.2. Théorème. *L'ensemble des solutions d'un système linéaire quelconque $(\Sigma) : AX = B$ où A est une matrice rectangulaire $n \times p$ de rang r est ou bien vide, ou bien c'est un sous-espace affine de dimension $(p - r)$ de $\mathcal{M}_{p\times 1}(\mathbb{K})$ de la forme*

$$X_{\text{part}} + \operatorname{Ker}(A),$$

où X_{part} est une solution particulière de (Σ). [Un sous-espace affine d'un espace vectoriel est par définition un translaté d'un sous-espace vectoriel, un tel sous-espace affine ne contient en général pas 0 et n'est donc pas un sous-espace vectoriel].



Il nous reste à caractériser les conditions assurant que l'ensemble des solutions d'un système linéaire donné (Σ) est non vide ; on dit alors que (Σ) est *compatible*, sinon, lorsque l'ensemble des solutions est vide, on dit que le système est *incompatible*.

4.3.3. Exemple. Soit à résoudre sur $\mathbb{R}$ le sytème linéaire

$$(\Sigma) : \quad \begin{cases} 3x - y + z = b_1 \\ x + 2y - 5z = b_2 \\ 4x + y - 4z = b_3. \end{cases}$$

La matrice $A = \begin{pmatrix} 3 & -1 & 1 \\ 1 & 2 & -5 \\ 4 & 1 & -4 \end{pmatrix}$ est de rang 2, en effet la troisième ligne est somme des deux premières, tandis que l'on a par exemple $\begin{vmatrix} 3 & -1 \\ 1 & 2 \end{vmatrix} \neq 0$. La condition

nécessaire et suffisante de compatibilité du système est donc que $b_3 = b_1 + b_2$, et dans ce cas la troisième équation est superflue puisque combinaison linéaire des deux premières. Si $b_3 = b_1 + b_2$, on a donc

$$(\Sigma) \iff \begin{cases} 3x - y + z = b_1 \\ x + 2y - 5z = b_2 \end{cases} \iff \begin{cases} 3x - y = b_1 - 3z \\ x + 2y = b_2 + 5z. \end{cases}$$

Lorsque $z \in \mathbb{R}$ est fixé (arbitrairement), il s'agit d'un système de Cramer par rapport à (x, y), et on trouve

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} \frac{2b_1+b_2-z}{7} \\ \frac{-b_1+3b_2+18z}{7} \\ z \end{pmatrix} = \begin{pmatrix} \frac{2b_1+b_2}{7} \\ \frac{-b_1+3b_2}{7} \\ 0 \end{pmatrix} + \frac{z}{7} \begin{pmatrix} -1 \\ 18 \\ 7 \end{pmatrix}$$

où $P = \begin{pmatrix} \frac{2b_1+b_2}{7} \\ \frac{-b_1+3b_2}{7} \\ 0 \end{pmatrix}$ est une solution particulière et $V = \begin{pmatrix} -1 \\ 18 \\ 7 \end{pmatrix}$ une base du noyau $\mathrm{Ker}(A)$. On voit que l'ensemble des solutions est la droite affine $\Delta = P + \mathbb{R}V$ passant par P et de vecteur directeur V. En général, ce n'est pas une droite vectorielle, Δ ne contient 0 que si $b_1 = b_2 = 0$ (et $b_3 = 0$ pour la compatibilité). □

De manière générale, pour une matrice $A \in \mathcal{M}_{n\times p}(\mathbb{K})$, on choisit un déterminant mineur $D_{I,J} = \det(A') \neq 0$ de taille r extrait de A, et on reprend la discussion du 4.2.4. Après permutation éventuelle des lignes et des colonnes, on est ramené à résoudre un système de la forme

$$\begin{pmatrix} A'X' + A''X'' \\ M'X' + M''X'' \end{pmatrix} = \begin{pmatrix} B' \\ B'' \end{pmatrix}, \quad B' \in \mathcal{M}_{r\times 1}(\mathbb{K}), \quad B'' \in \mathcal{M}_{(n-r)\times 1}(\mathbb{K}).$$

On sait ici que chaque ligne $M_k = (M'_k \;\; M''_k)$ de $M = (M' \;\; M'')$, $k \in \complement I$, est combinaison linéaire des lignes $A_i = (A'_i \;\; A''_i)$, $i \in I$, de A, i.e. $M_k = \sum_{i=1} \lambda_i A_i$, ou encore

$$M'_k = \sum_{i=1}^{r} \lambda_i A'_i, \quad M''_k = \sum_{i=1}^{r} \lambda_i A''_i,$$

pour des coefficients $\lambda_1 \dots \lambda_r \in \mathbb{K}$ (dépendant évidemment de l'indice ligne k). Pour que le système (Σ) soit compatible, il faut et il suffit que dans chacune des équations correspondantes

$$\sum_{j=1}^{p} m_{kj} x_j = b''_k, \tag{4.3.4}$$

le coefficient b''_k soit précisément égal à la même combinaison linéaire $b''_k = \sum_{i=1}^{r} \lambda_i b'_i$ des constantes $b'_1, \dots, b'_r$. En effet, dans ce cas, les équations (4.3.4) sont conséquences des r premières, et (Σ) se ramène à un système de Cramer automatiquement compatible $A'X' = B' - A''X''$ (ses solutions sont données par $X' = (A')^{-1}(B'' - A''X'')$ avec X'' quelconque).

4.3.5. Proposition. *On suppose choisi un déterminant mineur*

$$D_{I,J} = \det(A') \neq 0$$

extrait de A, de taille $r = \operatorname{rang}(A)$. La condition nécessaire et suffisante de compatibilité de $(\Sigma) : AX = B$ est qu'on ait annulation des $(n-r)$ déterminants $(r+1)\times(r+1)$

$$\det \begin{pmatrix} A' & B' \\ M'_k & b''_k \end{pmatrix} = 0, \quad k \in \complement I$$

où M'_k est la k-ième ligne de M' (à savoir les portions de lignes de A qui correspondent aux lignes complémentaires $\complement I$ et aux colonnes J). Dans ce cas, (Σ) est équivalent au système de Cramer

$$A'X' = B' - A''X''$$

où $X'' \in \mathcal{M}_{(n-r)\times 1}(\mathbb{K})$ est quelconque, et sinon l'ensemble des solutions est vide.

Démonstration. En effet, comme A' est de rang r, l'annulation des déterminants ci-dessus équivaut au fait que la dernière ligne soit combinaison linéaire des r premières, avec des coefficients λ_i forcément uniques exprimant M'_k comme combinaison linéaire $M'_k = \sum_{i=1}^{r} \lambda_i A'_i$ (les lignes $A'_1, \ldots, A'_r$ étant indépendantes). On a alors aussi $b''_k = \sum_{i=1}^{r} \lambda_i b'_i$, et M''_k est encore la même combinaison linéaire $M''_k = \sum_{i=1}^{r} \lambda_i A''_i$, d'où la compatibilité de (Σ). □

4.3.6. Remarque. Dans le cas où le système est compatible, on obtient une solution particulière en prenant tout simplement $X'' = 0$, ce qui donne

$$X_{\text{part}} = \begin{pmatrix} (A')^{-1}B' \\ O \end{pmatrix}, \quad O = \text{matrice nulle} \in \mathcal{M}_{(p-r)\times 1}(\mathbb{K}),$$

et on a la solution générale

$$X = \begin{pmatrix} (A')^{-1}B' \\ O \end{pmatrix} + \begin{pmatrix} -(A')^{-1}A'' \\ I_{p-r} \end{pmatrix} X'', \quad X'' \in \mathcal{M}_{(p-r)\times 1}(\mathbb{K}) \ \text{ quelconque.}$$

Chapitre 4
Arithmétique entière et polynomiale

Dans ce chapitre, nous énonçons et démontrons les propriétés arithmétiques élémentaires partagées par l'anneau des entiers et l'anneau des polynômes à une indéterminée $\mathbb{K}[X]$ (où $\mathbb{K}$ est un corps commutatif). Afin de donner une présentation unifiée, il est commode de mettre en œuvre un certain nombre de notions générales concernant la divisibilité et la factorisation dans des anneaux commutatifs unitaires quelconques, en particulier la notion d'idéal d'un anneau, introduite en 1921 par Emmy Noether (éminente mathématicienne allemande, 1882–1935).

1. Généralités sur les anneaux et la divisibilité

1.1. Définitions principales

1.1.1. Définition. *Un anneau est un triplet $(A, +, \times)$ formé d'un ensemble A et de deux lois de composition interne*

$$\begin{aligned} + &: A \times A \to A, \quad (x, y) \mapsto x + y, \\ \times &: A \times A \to A, \quad (x, y) \mapsto x \times y \end{aligned}$$

ayant les propriétés suivantes :

(a) *$(A, +)$ est une groupe commutatif, c'est-à-dire que la loi $+$ appelée addition est associative, commutative, dotée d'un élément neutre 0_A, et que tout élément x possède un symétrique x', en sorte que $x + x' = 0_A$ (on le note $x' = -x$ et on l'appelle opposé de x) ;*

(b) *la multiplication $\times$ est associative et distributive par rapport à $+$, et possède un élément neutre noté 1_A.*

1.1.2. Remarque. On omettra assez souvent l'indice A dans l'écriture de 0_A et 1_A. Nous avons ici supposé l'existence d'un élément neutre 1_A pour la multiplication, mais certains auteurs n'incluent pas cet axiome et parlent alors d'*anneau unitaire*. Il est à noter qu'on ne suppose pas nécessairement $1_A \neq 0_A$, par exemple $A = \{0\}$ est bien un anneau et 0 y est élément neutre à la fois pour l'addition et la multiplication. En fait, comme les axiomes impliquent $0_A \times x = 0_A$ et $1_A \times x = x$, on voit que l'on a $1_A = 0_A$ si et seulement si l'anneau est réduit à $\{0_A\}$, donc ce cas est unique et très "dégénéré". Un anneau A est appelé un *corps* si $1_A \neq 0_A$ et si tout élément $x \neq 0_A$ possède un inverse x' pour la multplication, c'est-à-dire tel que $x \times x' = x' \times x = 1_A$. Un corps $\mathbb{K}$ possède donc au moins deux éléments $0_{\mathbb{K}}$ et $1_{\mathbb{K}}$. Il peut fort bien se réduire à ces deux éléments, c'est le cas

du corps noté $\mathbb{F}_2 = \{0,1\}$ (ou encore $\mathbb{Z}/2\mathbb{Z}$, voir plus loin), dans lequel on prend $1+1=0$.

1.1.3. Définition. *Un morphisme d'anneaux $\varphi : A \to B$ entre deux anneaux A et B (non nécessairement commutatifs) est une application possédant les trois propriétés suivantes :*

(a) $\forall x, y \in A,\ \varphi(x+y) = \varphi(x) + \varphi(y)$;

(b) $\forall x, y \in A,\ \varphi(x \times y) = \varphi(x) \times \varphi(y)$;

(c) $\varphi(1_A) = 1_B$.

Il résulte de (a) que φ est en particulier un morphisme de groupes additifs de $(A,+)$ dans $(B,+)$, ce qui entraîne que l'on a aussi $\varphi(0_A) = 0_B$ [prendre $x = y = 0_A$]. L'axiome (c) sert à éviter l'application inintéressante $\varphi = 0$ et d'autres pathologies.

Dans la suite de ce chapitre, mis à part l'exemple 1.1.4 (e) ci-dessous, on s'intéressera exclusivement à des anneaux $(A,+,\times)$ commutatifs et à des corps $(\mathbb{K},+,\times)$ commutatifs, c'est-à-dire tels que la multiplication $\times$ soit *commutative*. Pour simplifier les notations, on omettra souvent l'écriture des lois, et on parlera d'un anneau A et d'un corps $\mathbb{K}$ (qui seront implicitement supposés commutatifs, sauf mention explicite du contraire).

1.1.4. Exemples d'anneaux. Ce chapitre sera presque exclusivement consacré aux deux exemples fondamentaux (a) et (b) ci-dessous et à leurs propriétés arithmétiques.

(a) $(\mathbb{Z},+,\times)$, l'anneau des entiers relatifs.

(b) $(\mathbb{K}[X],+,\times)$, l'anneau des polynômes à coefficients dans un corps $\mathbb{K}$. Un polynôme est une expression formelle $P = \sum_{i=0}^{d} a_i X^i$, $a_i \in \mathbb{K}$, qui peut être codée comme une somme $P = \sum_{i\in\mathbb{N}} a_i X^i$ possédant seulement un nombre fini de coefficients a_i non nuls, ou encore, si on veut, comme une suite infinie de coefficients $(a_0, a_1, \ldots, a_n, \ldots)$ presque tous nuls. Étant donné

$$P = \sum_{i\in\mathbb{N}} a_i X^i, \quad Q = \sum_{i\in\mathbb{N}} b_i X^i,$$

l'addition et la multiplication sont définies par

$$P + Q = \sum_{i\in\mathbb{N}} (a_i + b_i) X^i, \quad P \times Q = \sum_{k\in\mathbb{N}} \left(\sum_{i+j=k} a_i b_j \right) X^k.$$

Au polynôme $P = \sum_{i=0}^{d} a_i X^i$, $a_i \in \mathbb{K}$, on peut associer la fonction polynomiale

$$f_P : \mathbb{K} \to \mathbb{K}, \quad x \mapsto f_P(x) = \sum_{i=0}^{d} a_i x^i,$$

mais il convient de distinguer soigneusement P de f_P : P n'est pas une fonction, mais un objet "formel", élément d'un espace vectoriel de dimension

infinie sur $\mathbb{K}$, dont $(X^i)_{i\in\mathbb{N}}$ est une base. Par exemple, si on prend le corps à 2 éléments $\mathbb{K} = \mathbb{F}_2 = \{0, 1\}$, tous les polynômes $P_i = X^i$, $i \geqslant 1$, fournissent la même fonction polynomiale $f_i : \mathbb{K} \to \mathbb{K}$, $x \mapsto f_i(x) = x^i$ telle que $f_i(0) = 0$ et $f_i(1) = 1$. Par ailleurs, le polynôme non nul $P = X^2 - X = X(X - 1)$ a pour fonction polynôme associée $f_P = 0$ sur $\mathbb{F}_2$!

(c) L'anneau $A = \mathbb{Z}/6\mathbb{Z}$ (la notation sera expliquée plus loin), formé des entiers calculés modulo 6. On prend $A = \{0, 1, 2, 3, 4, 5\}$ et on calcule le résultat des opérations $+$ et $\times$ en considérant que 6 est nul (de sorte que par exemple $3 + 5 = 8 = 6 + 2 = 2$, $3 \times 5 = 15 = 6 + 6 + 3 = 3$). Ceci donne les tables de Pythagore suivantes pour les lois $+$ et $\times$ de l'anneau :

$+$	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

$\times$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	4	4	2
5	0	5	4	3	2	1

Dans cet anneau A "étrange", on a donc des éléments non nuls dont le produit est nul, par exemple $2 \times 3 = 0$.

(d) L'ensemble $\mathcal{F}_{\mathbb{R}}$ des fonctions de $\mathbb{R}$ dans $\mathbb{R}$, muni de l'addition et de la multiplication des fonctions est un anneau commutatif qui a aussi la propriété "étrange" ci-dessus : le produit des fonctions non nulles $f(x) = x + |x|$ et $g(x) = x - |x|$ est la fonction $fg = 0$ (puisque $f(x)g(x) = x^2 - |x|^2 = 0$ sur $\mathbb{R}$).

(e) L'ensemble des matrices carrées $\mathcal{M}_{n\times n}(\mathbb{K})$, muni de l'addition et de la multiplication des matrices, est un anneau non commutatif si $n \geqslant 2$ (si $n = 1$, il s'identifie au corps commutatif $\mathbb{K}$). Le chapitre suivant donnera quelques lumières sur ses propriétés.

1.1.5. Exemples de morphismes d'anneaux.

(a) L'application $\varphi : \mathbb{Z} \to \mathbb{F}_2 = \{0, 1\}$ telle que $\varphi(x) = 0$ si x est pair et $\varphi(x) = 1$ si x est impair est un morphisme d'anneaux.

(b) Si $w \in \mathbb{K}$ est un élément fixé, l'application $\varphi_w : \mathbb{K}[X] \to \mathbb{K}$ qui à un polynôme P associe sa valeur $\varphi_w(P) = P(w)$ au point w est un morphisme d'anneaux.

(c) L'ensemble $\mathbb{D}$ des nombres décimaux est un anneau, et l'inclusion $\varphi : \mathbb{Z} \hookrightarrow \mathbb{D}$ est un morphisme d'anneaux.

(d) L'application $\varphi_w : \mathcal{F}_{\mathbb{R}} \to \mathbb{R}$ qui associe à une fonction $f \in \mathcal{F}_{\mathbb{R}}$ sa valeur $\varphi_w(f) = f(w)$ en un point $w \in \mathbb{R}$ est un morphisme d'anneaux.

(e) Pour tout anneau A, on a un morphisme dit *canonique*

$$\varphi_{\mathrm{can}} : \mathbb{Z} \to A, \quad n \mapsto n_A$$

où l'on définit $n_A := 1_A + \cdots + 1_A$ (n fois) si $n \in \mathbb{N}^*$, $(-n)_A := -(n_A)$, l'élément 0_A étant ici encore l'élément neutre de A pour l'addition. L'additivité de φ_{can} est évidente, et la multiplicativité résulte de la distributivité de $\times$ par rapport à $+$ dans A. On prendra garde au fait que l'on peut avoir $n_A = 0$ même si $n \neq 0$, comme c'est le cas pour $n = 6$ dans l'exemple 1.1.4 (e) ; cependant, il est fréquent que l'on omette l'indice A et que l'on désigne par le même symbole n un élément de $\mathbb{Z}$ et son image dans A, sauf si on veut absolument éviter les confusions.

On donne un nom au phénomène apparu dans les exemples 1.1.4 (c) et (d).

1.1.6. Définition. *Soit A un anneau (commutatif, on ne le répétera plus !).*

(a) *On appelle diviseur de 0 dans A tout élément $x \neq 0$ pour lequel il existe $y \neq 0$ tel que $xy = 0$.*

(b) *L'anneau A est dit* intègre *s'il ne possède pas de diviseurs de 0, c'est-à-dire si*

$$\forall x, y \in A, \quad x \neq 0 \ \text{ et } \ y \neq 0 \implies xy \neq 0.$$

ou encore, par contraposition, si

$$\forall x, y \in A, \quad xy = 0 \implies x = 0 \ \text{ ou } \ y = 0.$$

1.1.7. Exemples d'anneaux intègres et non intègres.

(a) Comme il est bien connu, $\mathbb{Z}$ est un anneau intègre.

(b) Tout corps $\mathbb{K}$ est un anneau intègre : en effet, si $x, y \in \mathbb{K}$ sont non nuls, il possèdent des inverses x', y', et donc $x'xyy' = 1$, ce qui implique $xy \neq 0$.

(c) L'anneau $\mathbb{K}[X]$ est également intègre : en effet si on a deux polynômes non nuls $P = \sum_{j=0}^{d} a_j X^j$, $Q = \sum_{j=0}^{\delta} b_j X^j$ de termes dominants non nuls $a_d X^d$ et $b_\delta X^\delta$, alors PQ est de terme dominant non nul $a_d b_\delta X^{d+\delta}$. En particulier on voit que

$$\deg(PQ) = d + \delta = \deg(P) + \deg(Q).$$

Afin que cette égalité soit encore vraie lorsque $P = 0$ ou $Q = 0$, on convient de définir le degré du polynôme nul comme étant $-\infty$, avec la règle $-\infty + \delta = -\infty$ pour tout $\delta \in \mathbb{N} \cup \{-\infty\}$.

(d) Si A est un anneau commutatif, on peut définir l'anneau de polynômes $A[X]$ à coefficients dans A de la même manière que pour un corps, et le raisonnement du (c) implique que l'anneau $A[X]$ est intègre dès que A est lui-même intègre.

(e) Les anneaux $\mathbb{Z}/6\mathbb{Z}$ et $\mathcal{F}_{\mathbb{R}}$ définis ci-dessus ne sont pas intègres. L'anneau non commutatif $\mathcal{M}_{n\times n}(\mathbb{K})$ non plus, si $n \geqslant 2$: par exemple le carré de la matrice $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ est nul.

1.2. Notions liées à la divisibilité

1.2.1. Définitions et notations. *Soit $(A,+,\times)$ un anneau intègre non trivial.*

(a) *On notera $A^* = A\setminus\{0\}$ l'ensemble des éléments non nuls de A. L'hypothèse A intègre signifie que la multiplication est une loi de composition interne sur A^*.*

(b) *Si $x,y\in A^*$, on dit que x* divise *y, et on écrit $x\mid y$, s'il existe $\lambda\in A^*$ tel que $y=\lambda x$. Dans ce cas, on dit aussi que y est un* multiple *de x et que x est un* diviseur *de y. Par extension, on convient aussi de dire que 0 est un multiple de x (on prend $\lambda=0$).*

Les propriétés suivantes sont à peu près évidentes.

1.2.2. Propriétés.

(a) *(transitivité) si $x,y,z\in A^*$, $x\mid y$ et $y\mid z$ entraîne $x\mid z$.*

(b) *pour $x\in A^*$ et $y,z,\alpha,\beta\in A$, si $x\mid y$ et $x\mid z$ alors $x\mid(\alpha y+\beta z)$, à moins que $\alpha y+\beta z=0$.*

Démonstration. (a) Si $y=\lambda x$ et $z=\mu y$, alors $z=\mu(\lambda x)=(\mu\lambda)x$ dans A^*.

(b) Si $y=\lambda x$ et $z=\mu x$, alors $\alpha y+\beta z=(\alpha\lambda+\beta\mu)x$. □

1.2.3. Éléments inversibles. *Soit A est un anneau non trivial (i.e. $1_A\neq 0_A$), non nécessairement commutatif. Un élément $u\in A$ est dit inversible s'il existe $u'\in A$ tel que*

$$uu'=u'u=1_A.$$

On note $A^\times$ l'ensemble des éléments inversibles. C'est un sous-ensemble de A^*.

1.2.4. Proposition. *$(A^\times,\times)$ est un groupe.*

Démonstration. Il est facile de voir que le produit de deux éléments inversibles u,v est inversible et que $(uv)^{-1}=v^{-1}u^{-1}$, donc la multiplication est une loi de composition interne. Par ailleurs 1_A est inversible et $1_A\in A^\times$. Comme la multiplication est associative et que l'inverse u^{-1} d'un inversible est lui aussi inversible, la proposition s'ensuit. □

1.2.5. Exemples.

(a) Dans $\mathbb{Z}$, l'ensemble des éléments inversibles est $\mathbb{Z}^\times=\{1,-1\}$.

(b) Dans $\mathbb{K}[X]$ si on a $PQ=1$ alors P et Q sont non nuls et $\deg(P)+\deg(Q)=0$, ce qui implique que P et Q sont des polynômes constants non nuls. Par conséquent

$$\mathbb{K}[X]^\times=\{\text{polynômes constants } a_0\neq 0\}\simeq\mathbb{K}^*.$$

(c) Considérons

$$A=\{x=a+b\sqrt{2}\,/\,a,b\in\mathbb{Z}\}$$

Comme

$$(a+b\sqrt{2})(a'+b'\sqrt{2}) = (aa'+2bb') + (ab'+a'b)\sqrt{2},$$

on voit aisément que $(A,+,\times)$ est un anneau. Dans cet anneau

$$(1+\sqrt{2})(-1+\sqrt{2}) = 1$$

donc $\pm(1+\sqrt{2})$ et $\pm(-1+\sqrt{2})$ sont inversibles. Par la propriété de groupe de $A^\times$, tous les éléments $u = \pm(1+\sqrt{2})^k$, $k \in \mathbb{Z}$ sont inversibles. On peut montrer qu'il n'y en a pas d'autres.

1.2.6. Éléments irréductibles. *Soit A un anneau intègre non trivial. Un élément $p \in A^*$ est dit irréductible si p n'est pas inversible et si on ne peut pas décomposer p sous la forme*

$$p = xy \quad \textit{avec } x, y \in A^* \textit{ non inversibles,}$$

ou, de façon équivalente, si p est non inversible et si

$$\forall x, y \in A^*, \quad p = xy \implies x \textit{ ou } y \textit{ inversible.}$$

On notera qu'on peut toujours décomposer un élément $x \in A^*$ quelconque sous la forme $x = u \times u^{-1}x$ pour tout élément inversible u, ce qui justifie le fait d'exiger une décomposition en éléments non inversibles.

1.2.7. Exemples.

(a) Dans $\mathbb{Z}$, les éléments irréductibles sont exactement les nombres premiers et leurs opposés :

$$\mathcal{P} = \{2, 3, 5, 7, 11, \dots\},$$
$$-\mathcal{P} = \{-2, -3, -5, -7, -11, \dots\}.$$

(b) Dans $\mathbb{K}[X]$, les éléments irréductibles sont nécessairement des polynômes de degré $\geqslant 1$, car les constantes (non nulles) sont inversibles. Il est clair que tout polynôme $P = aX + b$ de degré 1 ($a \neq 0$) est irréductible, puisque si $P = QR$ avec Q, R non inversibles, alors $\deg(Q) \geqslant 1$ et $\deg(R) \geqslant 1$, donc $\deg(P) \geqslant 2$. Notons que $aX + b = a(X + b/a)$ admet la racine $x = -b/a \in \mathbb{K}$.

(c) Il peut se produire que des polynômes de degré $\geqslant 2$ soient irréductibles. Par exemple, dans $\mathbb{R}[X]$, $P = X^2 + 1$ est irréductible, car s'il avait une décomposition $P = QR$, les facteurs seraient nécessairement de degré 1, disons $Q = aX + b$, $R = a'X + b$, $a, a' \in \mathbb{R}^*$, $b, b' \in \mathbb{R}$, ce qui donnerait des racines réelles pour P.

(d) Le même raisonnement montre que $P = X^3 - 2$ est irréductible dans $\mathbb{Q}[X]$. Sinon $P = QR$ avec un facteur de degré 1, disons $Q = aX + b$ avec $a \in \mathbb{Q}^*$, $b \in \mathbb{Q}$, ce qui donnerait une racine rationnelle $x = -b/a$. Mais $P = X^3 - 2$ admet seulement la racine réelle $x = \sqrt[3]{2}$ qui n'est pas rationnelle, cf. 2.4.5.

1.3. Division euclidienne

Donnons d'abord la définition dans un anneau commutatif intègre quelconque.

1.3.1. Définition. *Soit A un anneau commutatif intègre (non trivial). On dit que A est un anneau euclidien, ou que A possède une division euclidienne, s'il existe une fonction $v : A^* \to \mathbb{N}$ (appelée stathme euclidien ou simplement stathme) telle que*

$$\forall a \in A,\ \forall b \in A^*,\ \exists q, r \in A \text{ tels que } a = bq + r \text{ avec } r = 0 \text{ ou } r \neq 0,\ v(r) < v(b).$$

Ici a est appelé dividende, b diviseur, q quotient et r reste de la division euclidienne.

Il est à noter qu'on n'exige pas en général l'unicité du couple (q, r).

1.3.2. Cas des anneaux $\mathbb{Z}$ et $\mathbb{K}[X]$.

(a) $\mathbb{Z}$ possède une division euclidienne* associée au stathme $v(x) = |x|$. Dans ce cas on a une division euclidienne sous la forme plus précise

$$a = bq + r \quad \text{avec} \quad 0 \leqslant r < |b|,$$

et sous cette condition, le couple (q, r) est *unique*. Redémontrons cette propriété. Si $b < 0$, quitte à changer b en $-b$ et q en $-q$, on se ramène au cas $b > 0$. Pour $a \in \mathbb{N}$, on montre alors la propriété par récurrence sur a : elle est vraie si $0 \leqslant a < b$ (prendre $q = 0$ et $r = a$), et si $a \geqslant b$, on peut appliquer l'hypothèse de récurrence à $a' = a - b < a$ pour obtenir $a' = bq' + r$, $0 \leqslant r < b$, ce qui donne

$$a = a' + b = b(q' + 1) + r = bq + r \quad \text{avec} \quad q = q' + 1.$$

Lorsque $a < 0$, on peut appliquer la division à $a'' = a + |a|b \geqslant |a|(b - 1) \geqslant 0$, et la division $a'' = bq'' + r$ donne par soustraction $a = bq + r$ avec $q = q'' - |a|$. Ici, il y a unicité de (q, r), car si $a = bq + r = bq' + r'$ avec $0 \leqslant r, r' < b$, on en déduit que $r' - r = b(q - q')$ est multiple de b, et comme $-b < r' - r < b$, la seule possibilité est $r' - r = 0$, d'où aussi $q - q' = 0$.

(b) $\mathbb{K}[X]$ possède une division euclidienne associée au stathme $v(P) = \deg(P)$: si $B \in \mathbb{K}[X]^*$, tout polynôme $A \in \mathbb{K}[X]$ peut s'écrire de *manière unique*

$$A = BQ + R \quad \text{avec} \quad \deg(R) < \deg(B)$$

(ce qui inclut la possibilité que $R = 0$ avec notre convention que $\deg(0) = -\infty$). La démonstration est similaire à celle faite dans $\mathbb{Z}$: on raisonne par récurrence sur $d = \deg(A)$, le résultat étant évident si $d = \deg(A) < \delta = \deg(B)$ (on prend

* En arithmétique pure, il y a bien d'autres exemples que $\mathbb{Z}$ admettant une division euclidienne, par exemple l'anneau des "entiers de Gauss" $A = \mathbb{Z}[i] = \{x + yi \,/\, x, y \in \mathbb{Z}\}$. Si $a \in A$ et $b \in A^*$, on vérifie aisément que le point $q \in A$ le plus proche du quotient exact $\frac{a}{b} \in \mathbb{C}$ est tel que $r = a - bq$ satisfait $N(r) < N(b)$ où $N(z) = N(x + yi) = x^2 + y^2 \in \mathbb{N}$. On prend ici $v = N$ comme stathme.

alors $Q = 0$, $R = A$). Supposons maintenant $d \geqslant \delta$ et le résultat démontré pour $d' < d$, et écrivons

$$A = a_0 + a_1 X + \cdots + a_d X^d, \quad B = b_0 + b_1 X + \cdots + b_\delta X^\delta, \quad a_d \neq 0,\ b_\delta \neq 0.$$

Si on prend la différence des termes dominants dans

$$A' := A - ((a_d/b_\delta) X^{d-\delta}) B \in \mathbb{K}[X]$$

on trouve $a_d X^d - ((a_d/b_\delta) X^{d-\delta})(b_\delta X^\delta) = 0$, donc $d' = \deg(A') < d$. Par hypothèse de récurrence, on peut écrire $A' = BQ' + R$, ce qui donne alors

$$A = BQ + R \quad \text{avec} \quad Q = (a_d/b_\delta) X^{d-\delta} + Q'.$$

Notons que cette démonstration est précisément l'explicitation théorique de l'algorithme de division des polynômes. Ici encore on a unicité, puisque

$$A = BQ + R = BQ' + R' \quad \text{avec} \quad \deg(R), \deg(R') < \deg(B)$$

implique $R' - R = B(Q - Q')$ avec $\deg(R' - R) < \deg(B)$, d'où $R' - R = 0$ et $Q - Q' = 0$.

1.3.3. Exemple de division dans $\mathbb{K}[X]$. Voici un exemple de division polynomiale ; on procède comme pour la division entière, en appliquant la procédure de récurrence décrite plus haut.

$$\begin{array}{l|l}
3X^5 - 2X^4 + 4X^3 - 5X^2 + 2X + 4 & 2X^2 - 3X + 4 \\
3X^5 - \frac{9}{2}X^4 + 6X^3 & \hline \\
\hline
\frac{5}{2}X^4 - 2X^3 - 5X^2 & \frac{3}{2}X^3 + \frac{5}{4}X^2 + \frac{7}{8}X - \frac{59}{16} \\
\frac{5}{2}X^4 - \frac{15}{4}X^3 + 5X^2 & \\
\hline
\frac{7}{4}X^3 - 10X^2 + 2X & \\
\frac{7}{4}X^3 - \frac{21}{8}X^2 + \frac{7}{2}X & \\
\hline
-\frac{59}{8}X^2 - \frac{3}{2}X + 4 & \\
-\frac{59}{8}X^2 + \frac{177}{16}X - \frac{59}{4} & \\
\hline
-\frac{201}{16}X + \frac{75}{4} &
\end{array}$$

Le degré du dernier reste $\deg(R) = 1$ est inférieur au degré $\deg(B) = 2$ du diviseur $B = 2X^2 - 3X + 4$, donc l'opération est terminée. On en déduit l'égalité

$$\begin{aligned}
&3X^5 - 2X^4 + 4X^3 - 5X^2 + 2X + 4 \\
&\qquad = (2X^2 - 3X + 4)\big(\tfrac{3}{2}X^3 + \tfrac{5}{4}X^2 + \tfrac{7}{8}X - \tfrac{59}{16}\big) + \big(-\tfrac{201}{16}X + \tfrac{75}{4}\big).
\end{aligned}$$

1.3.4. Remarque. Comme on le voit avec l'exemple ci-dessus, la division euclidienne n'est pas possible en général dans un anneau de polynômes $\mathbb{A}[X]$ dont les coefficients sont pris dans un anneau $\mathbb{A}$, car on est amené à utiliser des fractions. En revanche, si le polynôme B est *unitaire* de degré δ, c'est-à-dire de coefficient dominant $b_\delta = 1$,

$$B = X^\delta + b_{\delta-1}X^{\delta-1} + \cdots + b_1X + b_0, \quad b_j \in \mathbb{A},$$

alors la division euclidienne est possible, car dans ce cas l'étape de récurrence décrite au 1.3.2 (b) ne nécessite plus de fractions. En particulier, si on fait dans $\mathbb{Z}[X]$ une division par un polynôme unitaire, le quotient et le reste sont bien dans $\mathbb{Z}[X]$.

1.4. Racines des polynômes et factorisation

Soit $P \in A[X]$ un polynôme à coefficients dans un anneau intègre A. Si l'on effectue la division par $(X-w)$ avec $w \in A$, ce qui est possible puisque $(X-w)$ est unitaire, le reste $R \in A[X]$ doit vérifier $\deg(R) < \deg(X-a) = 1$, par conséquent ce reste est une constante : $P(X) = (X-w)Q(X)+c$. Mais si on substitue $X = w$, il vient $P(w) = c$. On en déduit aisément :

1.4.1. Théorème. *Pour tout polynôme $P \in A[X]$ non constant et tout élément $w \in A$, il existe un polynôme $Q \in A[X]$ de degré $\deg(Q) = \deg(P) - 1$ tel que*

$$P(X) = (X-w)Q(X) + P(w).$$

En particulier, si $P(w) = 0$, alors P est divisible par $(X-w)$.
[Nota. *Si P est constant, ceci est vrai aussi avec $Q = 0$*].

1.4.2. Corollaire. *Soit $P \in A[X]$ un polynôme admettant des racines 2 à 2 distinctes $w_1, \ldots, w_k$. Alors P est divisible par $(X-w_1)\ldots(X-w_k)$, c'est-à-dire qu'il existe $Q \in A[X]$ tel que*

$$P(X) = (X-w_1)\ldots(X-w_k)Q(X)$$

Démonstration. On raisonne par récurrence sur k, le résultat ayant déjà été démontré pour $k = 1$. Supposons le résultat déjà démontré pour k, et supposons que P admette une autre racine w_{k+1}. Alors

$$0 = P(w_{k+1}) = (w_{k+1}-w_1)\ldots(w_{k+1}-w_k)Q(w_{k+1}).$$

Comme l'anneau A est intègre et que $w_{k+1}-w_j \neq 0$ pour $1 \leqslant j \leqslant k$, on en conclut que $Q(w_{k+1}) = 0$, mais alors $Q(X)$ est divisible par $(X-w_{k+1})$ et par conséquent $P(X)$ est divisible par $(X-w_1)\ldots(X-w_k)(X-w_{k+1})$, ce qui démontre bien la propriété à l'ordre $k+1$. □

Les propriétés qui précèdent sont très utiles également dans les anneaux de polynômes à plusieurs variables. On définit ainsi un polynôme $P \in A[X,Y]$ comme une expression formelle

$$P(X,Y) = \sum_{1\leqslant i\leqslant d,\, 1\leqslant j\leqslant d'} a_{i,j} X^i Y^j, \quad a_{i,j} \in A.$$

Le degré total $\deg(P)$ (resp. les degrés partiels $\deg_X(P)$, $\deg_Y(P)$) désigne le maximum des entiers $i+j$ (resp. i, resp. j) tels qu'il existe un coefficient $a_{i,j} \neq 0$. Si l'on écrit P sous forme factorisée

$$P(X,Y) = \sum_{1\leqslant j\leqslant d'} \left(\sum_{1\leqslant i\leqslant d} a_{i,j} X^i \right) Y^j,$$

on en conclut que $A[X,Y] = A[X][Y]$, c'est-à-dire que $A[X,Y]$ n'est autre que l'anneau des polyômes en l'indéterminée Y dont les coefficients sont des éléments de l'anneau $A[X]$. En appliquant le théorème 1.4.1 à l'anneau $A'[Y]$ avec $A' = A[X]$, on obtient alors la conséquence suivante.

1.4.3. Corollaire. *Soit $P(X,Y) \in A[X,Y]$. Si P est tel que $P(X,X) = 0$, alors $P(X,Y)$ est divisible par $Y - X$ dans $A[X,Y]$.*

Le même raisonnement donne un résultat analogue pour un nombre quelconque d'indéterminées.

1.4.4. Corollaire. *Soit $P \in A[X_1, X_2, \ldots, X_n]$. Si $P(X_1, X_2, \ldots, X_n)$ devient nul lorsqu'on fait la substitution $X_j := X_i$, alors $P(X_1, X_2, \ldots, X_n)$ est divisible par $X_j - X_i$ dans $A[X_1, X_2, \ldots, X_n]$.*

1.4.5. Déterminant de Vandermonde. Nous allons illustrer ce qui précède en calculant la valeur du déterminant dit de Vandermonde

$$\Delta = \begin{vmatrix} 1 & 1 & \ldots & 1 & 1 \\ X_1 & X_2 & \ldots & X_{n-1} & X_n \\ \vdots & \vdots & & \vdots & \vdots \\ X_1^{i-1} & X_2^{i-1} & \ldots X_j^{i-1} \ldots & X_{n-1}^{i-1} & X_n^{i-1} \\ \vdots & \vdots & & \vdots & \vdots \\ X_1^{n-2} & X_2^{n-2} & \ldots & X_{n-1}^{n-2} & X_n^{n-2} \\ X_1^{n-1} & X_2^{n-1} & \ldots & X_{n-1}^{n-1} & X_n^{n-1} \end{vmatrix},$$

dont le coefficient (i,j) est X_j^{i-1}, et qui définit un polynôme $\Delta \in \mathbb{Z}[X_1, X_2, ..., X_n]$. Comme Δ est une somme de produits de facteurs dont un est pris dans chaque ligne, chaque monôme, tel que le monôme diagonal $X_1^0 X_2^1 \ldots X_n^{n-1}$, est de degré $0 + 1 + \cdots + (n-1) = n(n-1)/2$. Par conséquent Δ est un polynôme homogène de degré $n(n-1)/2$. Or Δ s'annule si on substitue $X_j := X_i$,

$j > i$, et le raisonnement du corollaire 1.4.2 implique que Δ est divisible par le produit $P = \prod_{1\leqslant i<j\leqslant n}(X_j - X_i)$ (on peut appliquer la récurrence faite plus haut, car la substition $X_j := X_i$ annule seulement le facteur $X_j - X_i$). Mais P est également un polynôme homogène de degré $n(n-1)/2$, de sorte que le quotient Δ/P est une constante. Comme par ailleurs P contient le monôme $\prod_{1\leqslant i<j\leqslant n} X_j = \prod_{1\leqslant j\leqslant n} X_j^{j-1}$, le quotient Δ/P est égal à 1. Ceci donne la formule classique

$$\Delta = \prod_{1\leqslant i<j\leqslant n} (X_j - X_i).$$

1.5. Dérivation des polynômes

Soit $P \in A[X]$ un polynôme de degré d sur un anneau intègre A, écrit

$$P(X) = \sum_{j=0}^{d} a_j X^j, \quad a_j \in A.$$

On commence par définir le polynôme dérivé $P'(X)$ de manière purement algébrique – sans avoir à prendre de véritable limite comme dans $\mathbb{R}$ ou $\mathbb{C}$. Soit Y une autre indéterminée. On forme le "taux d'accroissement"

$$\begin{aligned}\tau_P(X,Y) = \frac{P(Y)-P(X)}{Y-X} &= \sum_{j=0}^{d} a_j \frac{Y^j - X^j}{Y-X} \\ &= \sum_{j=0}^{d} a_j (X^{j-1} + X^{j-2}Y + \cdots + XY^{j-2} + Y^{j-1}).\end{aligned}$$

On voit alors que $\tau_P(X,Y) \in A[X,Y]$ est un polynôme de degré total $d-1$.

1.5.1. Définition. *On appelle polynôme dérivé de P le polynôme $P' \in A[X]$ de degré $d-1$ tel que*

$$P'(X) = \tau_P(X,X) = \sum_{j=0}^{d} j a_j X^{j-1}.$$

Il est facile de voir que $\tau_{P+Q}(X,Y) = \tau_P(X,Y) + \tau_Q(X,Y)$ et que

$$\begin{aligned}\tau_{PQ}(X,Y) = \frac{PQ(Y)-PQ(X)}{Y-X} &= \frac{P(Y)-P(X)}{Y-X}Q(Y) + P(X)\frac{Q(Y)-Q(X)}{Y-X} \\ &= \tau_P(X,Y)Q(Y) + P(X)\tau_Q(X,Y),\end{aligned}$$

ce qui, après substitution $Y := X$, fournit les formules usuelles de dérivation :

1.5.2. Proposition. *Pour tous $P, Q \in A[X]$ on a*

$$(P+Q)' = P' + Q', \quad (PQ)' = P'Q + PQ'.$$

1.5.3. Formule de Leibniz. *Pour tous polynômes $P, Q \in A[X]$, la dérivée k-ième du produit PQ est donnée par*

$$(PQ)^{(k)} = \sum_{j=0}^{k} \binom{k}{j} P^{(j)} Q^{(k-j)}$$

en notant $P^{(0)} = P$.

Démonstration. On procède par récurrence sur k, en utilisant les propriétés du triangle de Pascal pour passer de k à $k+1$, à savoir que

$$\binom{k+1}{j} = \binom{k}{j} + \binom{k}{j-1}.$$

Nous laissons le détail de la vérification au lecteur. □

1.5.4. Remarque. Sur certains corps tels que $\mathbb{K} = \mathbb{F}_2 = \{0, 1\}$, la dérivation des polynômes peut présenter des propriétés "bizarres". Ainsi, au polynôme $P = X^2 + X \in \mathbb{F}_2[X]$ est associée la fonction polynôme nulle $f_P = 0 : \mathbb{F}_2 \to \mathbb{F}_2$. Cependant, le polynôme dérivé $P' = 2X + 1 = 1 \in \mathbb{F}_2[X]$ a pour fonction polynôme associée la fonction constante $f_{P'} = 1$.

1.6. Multiplicité des racines d'un polynôme

Soit A un anneau intègre et $P \in A[X]$ un polynôme de degré $d \geqslant 1$ admettant $w \in A$ comme racine. On sait qu'on peut écrire $P(X) = (X - w)P_1(X)$ avec $\deg(P_1) = d - 1$, et le polynôme P_1 peut (ou non) admettre de nouveau w comme racine. Si c'est le cas, on a $P_1(X) = (X - w)P_2(X)$ avec $\deg(P_2) = d - 2$, donc $P(X) = (X - w)^2 P_2(X)$. On peut répéter le raisonnement avec des polynômes P_i de degrés $d - i$ de plus en plus petits jusqu'à ce que $P_i(w) \neq 0$, ce qui se produit nécessairement à une certaine étape, au plus tard lorsque P_i devient de degré 0 (donc constant et non nul).

1.6.1. Définition. *Soit A un anneau intègre et $P \in A[X]$ un polynôme de degré d admettant $w \in A$ comme racine. On dit que w est une racine de multiplicité $m \geqslant 1$ si on peut écrire*

$$P(X) = (X - w)^m Q(X) \quad \textit{avec} \quad \deg(Q) = d - m, \quad Q(w) \neq 0.$$

Si w n'est pas racine de P, c'est-à-dire si $P(w) \neq 0$, l'égalité ci-dessous est encore valide si $m = 0$ et $Q = P$; on convient donc de dire que la multiplicité de w dans P est égale à 0. On a toujours

$$m \leqslant d = \deg(P).$$

Nous allons voir que la multiplicité d'une racine peut se déterminer en utilisant les dérivées successives du polynôme P. La dérivée j-ième de $(X - w)^m$ est

$m(m-1)\cdots(m-j+1)(X-w)^{m-j}$ pour $j \leqslant m$ (et est identiquement nulle pour $j > m$). Au point $X = w$, la seule valeur non nulle est celle de la dérivée m-ième qui vaut

$$\left(\frac{d}{dX}\right)^m (X-w)^m = m!\ .$$

La formule de Leibniz appliquée à $P(X) = (X-w)^m Q(X)$ donne

$$P^{(k)}(X) = \sum_{j=0}^{k} \binom{k}{j} m(m-1)\cdots(m-j+1)(X-w)^{m-j} Q^{(k-j)}(X).$$

Pour $P^{(m)}(w)$, on trouve en particulier que le seul terme non nul du membre de droite correspond à $j = m = k$. On obtient par conséquent :

1.6.2. Formule. *Si $P(X) = (X-w)^m Q(X)$, alors*

$$\begin{aligned} P(w) = P'(w) = \ldots = P^{(m-1)}(w) = 0,\\ P^{(m)}(w) = m!\, Q(w). \end{aligned}$$

1.6.3. Corollaire. *Soit $\mathbb{K}$ un corps "de caractéristique 0", c'est-à-dire tel que $n_{\mathbb{K}} \neq 0$ pour tout $n \in \mathbb{N}^*$, par exemple $\mathbb{K} = \mathbb{Q}, \mathbb{R}$ ou $\mathbb{C}$. Si $P(X) = (X-w)^m Q(X)$ dans $\mathbb{K}[X]$ avec $Q(w) \neq 0$, on a*

$$P(w) = P'(w) = \ldots = P^{(m-1)}(w) = 0, \quad P^{(m)}(w) = m!\, Q(w) \neq 0,$$

autrement dit la multiplicité de w est le plus petit entier $j \in \mathbb{N}$ tel que $P^{(j)}(w) \neq 0$.

1.6.4. Exemples. (a) Dans $\mathbb{Q}[X]$, on considère

$$P(X) = X^4 + X^3 - 30\,X^2 + 76\,X - 56.$$

On s'aperçoit que $P(2) = 16+8-120+152-56 = 0$. Pour trouver la multiplicité, on peut calculer les dérivées successives

$$\begin{aligned} &P'(X) = 4\,X^3 + 3\,X^2 - 60\,X + 76, && P'(2) = 0,\\ &P''(X) = 12\,X^2 + 6\,X - 60, && P''(2) = 0,\\ &P'''(X) = 24\,X + 6, && P'''(2) = 54 \neq 0, \end{aligned}$$

donc $x = 2$ est une racine triple, c'est-à-dire de multiplicité 3. Le quotient $P(X)/(X-2)^3$ est unitaire de degré 1 et on voit que l'on a la factorisation

$$P(X) = X^4 + X^3 - 30\,X^2 + 76\,X - 56 = (X-2)^3(X+7),$$

car le coefficient constant du quotient est $(-56)/(-2)^3 = 7$.

(b) Dans $\mathbb{C}[X]$, on considère

$$P(X) = X^5 - 3\,X^4 + 2\,X^3 - 6\,X^2 + X - 3.$$

Nous avons $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$, d'où $P(i) = i - 3 - 2i + 6 + i - 3 = 0$. On calcule alors

$$\begin{aligned} &P'(X) = 5\,X^4 - 12\,X^3 + 6\,X^2 - 12\,X + 1, \quad && P'(i) = 0, \\ &P''(X) = 20\,X^3 - 36\,X^2 + 12\,X - 12, && P''(i) = 24 - 8i \neq 0. \end{aligned}$$

Par conséquent $x = i$ est racine double de P. Mais comme P est à coefficients réels, on voit en prenant les conjugués que $P(-i) = P'(-i) = 0$ et $P''(-i) = 24 + 8i \neq 0$, de sorte que $-i$ est aussi racine double. On en conclut que $P(X)$ est divisible par $(X-i)^2(X+i)^2 = (X^2+1)^2$. Le quotient $P(X)/(X^2+1)^2$ est unitaire de degré 1 et de coefficient constant -3, d'où la factorisation

$$P(X) = X^5 - 3\,X^4 + 2\,X^3 - 6\,X^2 + X - 3 = (X-i)^2(X+i)^2(X-3).$$

1.7. Théorème de d'Alembert-Gauss

Le théorème de d'Alembert-Gauss, connu aussi sous le nom de *théorème fondamental de l'algèbre*, stipule que tout polynôme $P \in \mathbb{C}[X]$ non constant admet au moins une racine. Les nombres complexes ont été introduits par le mathématicien Bombelli entre 1560 et 1572, mais ce n'est pas avant d'Alembert, en 1746, qu'on se pose le problème d'une preuve rigoureuse du théorème fondamental de l'algèbre. En 1815, Gauss parvient finalement à une preuve complète, utilisant une récurrence algébrique très subtile sur le degré du polynôme ; elle repose d'une part sur la possibilité de résoudre les équations du second degré complexes, ce qui se ramène à des calculs de racines carrées complexes, et d'autre part sur le fait que tout polynôme réel $P \in \mathbb{R}[X]$ de degré impair admet une racine réelle. Pour cette dernière affirmation, on utilise le théorème des valeurs intermédiaires en observant que pour $x \in \mathbb{R}$, $P(x)$ varie entre $-\infty$ et $+\infty$, ou entre $+\infty$ et $-\infty$, en fonction du signe du coefficient dominant. Nous donnerons ici une preuve plus simple et plus directe s'appuyant sur des idées d'Argand (1806) et de Cauchy (1821), qui repose seulement sur le fait qu'une fonction réelle continue sur un segment y atteint son infimum en un point ; les détails de cette preuve restent tout de même assez subtils.

Commençons par des préliminaires généraux. Soit $P = \sum_{k=1}^{d} a_k X^k \in \mathbb{C}[X]$ un polynôme de degré $d \geqslant 1$, avec $a_d \neq 0$. Soit $r \geqslant 0$ fixé. Pour $\theta \in [0, 2\pi]$, la fonction $\theta \mapsto |P(re^{i\theta})|$ est continue, et atteint donc son infimum

$$m(r) = \inf_{\theta \in [0,2\pi]} |P(re^{i\theta})| = \min_{|z|=r} |P(z)|$$

en un point du cercle $|z| = r$. On a

$$\begin{aligned} |P(z)| &= \left|a_d z^d + a_{d-1} z^{d-1} + \cdots + a_1 z + a_0\right| \\ &\geqslant |a_d||z|^d - (|a_{d-1}||z|^{d-1} + \cdots + |a_1||z| + |a_0|), \end{aligned}$$

ce qui entraîne le lemme suivant.

1.7.1. Lemme. *On a la minoration*

$$m(r) \geqslant |a_d| r^d - (|a_{d-1}| r^{d-1} + \cdots + |a_1| r + |a_0|) \quad \textit{avec} \quad |a_d| > 0,$$

par conséquent $\lim_{r \to +\infty} m(r) = +\infty$. □

1.7.2. Lemme. *La fonction $r \mapsto m(r)$ est continue sur $[0, +\infty[$.*

Démonstration. Pour des complexes $z, w \in \mathbb{C}$ vérifiant $|z| \leqslant R$ et $|w| \leqslant R$, on a

$$w^k - z^k = (w - z)(z^{k-1} + z^{k-2} w + \cdots + z w^{k-2} + w^{k-1})$$

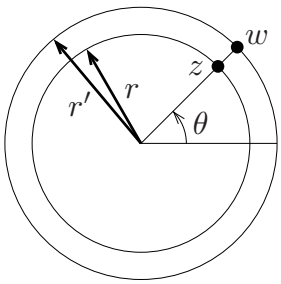
et chacun des k termes $z^j w^{k-1-j}$ est majoré par R^{k-1}. Il s'ensuit par conséquent que $|w^k - z^k| \leqslant k R^{k-1} |w - z|$ et

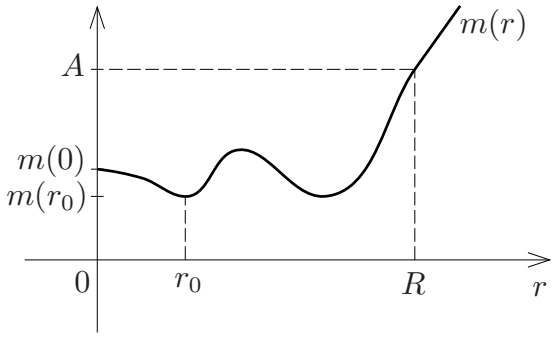
$$|P(w) - P(z)| = \left| \sum_{k=0}^{d} a_k (w^k - z^k) \right| \leqslant |w - z| \sum_{k=1}^{d} k |a_k| R^{k-1}.$$

Maintenant si on prend l'infimum de $|P(z)|$ sur les cercles de rayons $r, r' \in [0, R]$ et si on applique l'inégalité précédente à $z = re^{i\theta}$ et $w = r' e^{i\theta}$ en observant que $|w - z| = |r' - r|$, on voit en passant à l'infimum que

$$|m(r') - m(r)| \leqslant |r' - r| \sum_{k=1}^{d} k |a_k| R^{k-1}.$$

Ceci entraîne que $r \mapsto m(r)$ est continue sur $[0, R]$ pour tout $R > 0$, d'où la conclusion. □





Choisissons $A > m(0)$. Comme $\lim_{r \to +\infty} m(r) = +\infty$, il existe R suffisamment grand pour que $m(r) \geqslant A > m(0)$ pour $r \in [R, +\infty]$. Ceci implique que la fonction continue $r \mapsto m(r)$ atteint son infimum sur $[0, R]$, et on sait que cet infimum est atteint en un certain point $r_0 \in [0, R]$ (pas nécessairement unique).

1.7.3. Corollaire. *Il existe un point $z_0 = r_0 e^{i\theta_0}$ du cercle $|z| = r_0$ en lequel*

$$|P(z_0)| = m(r_0) = \inf_{r \in [0, +\infty[} m(r) = \inf_{z \in \mathbb{C}} |P(z)|.$$ □

Nous avons besoin d'un dernier lemme, énoncé par Argand en 1806 (mais publié seulement en 1814), qui est le point crucial de la démonstration.

1.7.4. Lemme. *Supposons $P \in \mathbb{C}[X]$ non constant. Soit $z_1 \in \mathbb{C}$ un point tel que $P(z_1) \neq 0$. Alors z_1 ne peut pas être un minimum local pour $z \mapsto |P(z)|$, autrement dit, pour tout $\delta > 0$, il existe $w \in \mathbb{C}$ tel que $|w| < \delta$ et $|P(z_1 + w)| < |P(z_1)|$.*

Notons que ce lemme est violemment faux sur $\mathbb{R}$: le polynôme $P(X) = X^2 + 1$ est non constant, positif, mais passe par un minimum non nul $P(0) = 1$ en $x = 0$.

Démonstration. Posons $Q(X) = P(z_1 + X)/P(z_1)$ de sorte que $Q(0) = 1$, et pour tout $w \in \mathbb{C}$ écrivons

$$Q(w) = 1 + b_k w^k + \cdots + b_d w^d$$

où $b_k \neq 0$ est le coefficient non nul d'indice $k \geqslant 1$ minimal (il existe puisque les polynômes P et Q sont non constants). On va s'arranger pour avoir un terme $b_k w^k$ qui soit réel négatif (et petit). Écrivons $b_k = \beta e^{i\varphi}$ avec $\beta = |b_k| > 0$ et prenons $w = r e^{i(\pi-\varphi)/k}$ avec $r \in]0, \delta[$. On a alors $|w| = r$ et

$$b_k w^k = \beta e^{i\varphi}\, r^k e^{i(\pi-\varphi)} = \beta r^k e^{i\pi} = -\beta r^k,$$

donc

$$Q(w) = 1 - \beta r^k + b_{k+1} w^{k+1} + \cdots + b_d w^d.$$

Choisissons r assez petit pour que $1 - \beta r^k > 0$. En prenant le module, ceci implique

$$\begin{aligned} \frac{|P(z_1 + w)|}{|P(z_1)|} = |Q(w)| &\leqslant 1 - \beta r^k + |b_{k+1}| r^{k+1} + \cdots + |b_d| r^d \\ &= 1 - r^k(\beta - |b_{k+1}| r - \cdots - |b_d| r^{d-k}) < 1 \end{aligned}$$

pour $|w| = r$ assez petit tel que $|b_{k+1}| r + \cdots + |b_d| r^{d-k} < \beta/2$ (et $\beta r^k < 1$). Le lemme est démontré. □

1.7.5. Théorème de d'Alembert-Gauss. *Tout polynôme $P \in \mathbb{C}[X]$ non constant admet au moins une racine $z_1 \in \mathbb{C}$.*

Démonstration. On considère le point $z_1 = z_0$ donné par le corollaire 1.7.3, en lequel $|P(z_1)| = \inf_{z \in \mathbb{C}} |P(z)|$. On veut démontrer que $P(z_1) = 0$. Or, si $P(z_1) \neq 0$, on peut appliquer le lemme 1.7.4 pour trouver un point $z_1' = z_1 + w$ tel que $|P(z_1')| < |P(z_1)|$. C'est une contradiction. Par conséquent $P(z_1) = 0$. □

1.7.6. Théorème. *Soit $P(X) = \sum_{j=0}^{d} a_j X^j \in \mathbb{C}[X]$ de degré d, i.e. $a_d \neq 0$. Alors on peut le factoriser en facteurs de degré 1 sous la forme*

$$P(X) = a_d \prod_{j=1}^{s} (X - z_j)^{m_j}, \quad \sum_{j=1}^{s} m_j = d = \deg P,$$

où $z_1, \ldots, z_s \in \mathbb{C}$ sont les racines complexes 2 à 2 distinctes et les $m_j \in \mathbb{N}^*$ leurs multiplicités.

Démonstration. On raisonne par récurrence sur d. Si $d = 0$, il n'y a rien à démontrer, on a $P(X) = a_0$ et un produit vide égal à 1, avec $s = 0$. Si $d = 1$, on a $P(X) = a_1X + a_0 = a_1(X - z_1)$ où $z_1 = -a_0/a_1$. Supposons $d \geqslant 2$, et le résultat déjà démontré pour $d-1$. D'après le théorème de d'Alembert-Gauss, $P(X)$ possède une racine complexe z_1, et on peut alors factoriser $P(X)$ sous la forme

$$P(X) = (X - z_1)Q(X).$$

Or $Q(X)$ est un polynôme de degré $d-1$ de terme dominant a_dX^{d-1}, on peut alors le factoriser entièrement sous la forme indiquée (avec somme des multiplicités égale à $d-1$). Le résultat s'ensuit pour $P(X)$. □

2. Idéaux et éléments d'arithmétique

2.1. Idéaux, idéaux principaux, anneaux principaux

On introduit d'abord de manière générale la notion d'idéal d'un anneau A : si les éléments de l'anneau sont vus par analogie comme des "scalaires", la définition est voisine de celle de sous-espace vectoriel d'un espace vectoriel.

2.1.1. Définition. *Soit A un anneau. Un sous-ensemble non vide I de A est appelé* idéal *de A si*

(i) $\forall x, y \in I, \quad x + y \in I$ (*stabilité pour +*) ;

(ii) $\forall \lambda \in A, \ \forall x \in I, \quad \lambda x \in A$ (*stabilité par la multiplication scalaire de A*).

Il est équivalent de demander que I soit stable par "combinaisons linéaires", autrement dit, la conjection des axiomes (i) *et* (ii) *est équivalente à l'axiome ci-dessous :*

(iii) $\forall \lambda, \mu \in A, \ \forall x, y \in I, \quad \lambda x + \mu y \in A$ (*stabilité par combinaisons linéaires*).

L'hypothèse que l'idéal I soit non vide, combinée à l'axiome (ii) avec $\lambda = 0$ implique que I contient nécessairement l'élément 0 de A.

2.1.2. Exemples. (a) Si $g \in A$, l'ensemble

$$\langle g \rangle = \{ \lambda g \,/\, \lambda \in A \}$$

des multiples de g, aussi noté gA, est un idéal de A ; par exemple, dans $\mathbb{Z}$,

$$\langle 7 \rangle = 7\mathbb{Z} = \{ \ldots, -21, -14, -7, 0, 7, 14, 21, \ldots \}$$

est un idéal. En particulier $\langle 1 \rangle = A$ est un idéal (appelé "idéal unité" de A), de même que $\langle 0 \rangle = \{0\}$ ("idéal zéro").

(b) Plus généralement, si $g_1, \ldots g_N \in A$, l'ensemble des combinaisons linéaires

$$\langle g_1, \ldots, g_N \rangle = \left\{ \sum_{i=1}^{N} \lambda_i g_i \,/\, \lambda_i \in A \right\}$$

est évidemment stable par combinaisons linéaires, donc c'est un idéal de A. Encore plus généralement, on peut considérer une famille finie ou infinie $(g_i)_{i\in S}$ et poser

$$\langle g_i\rangle_{i\in S} = \Big\{ \sum_{\text{finies}} \lambda_i g_i \,/\, i\in S,\ \lambda_i\in A\Big\}.$$

On l'appelle idéal engendré par la famille $(g_i)_{i\in S}$. On dit aussi que $(g_i)_{i\in S}$ est un *système de générateurs* d'un idéal I donné si on a précisément $I=\langle g_i\rangle_{i\in S}$.

(c) Dans l'anneau $\mathbb{K}[X,Y]$ des polynômes à 2 indéterminées, l'ensemble $I_{(0,0)}$ des polynômes P tels que $P(0,0)=0$ est un idéal (puisque cet ensemble est à l'évidence stable par combinaisons linéaires, si $P_1,P_2\in I_{(0,0)}$, alors $P=Q_1P_1+Q_2P_2$ vérifie bien $P(0,0)=0$, donc $P\in I_{(0,0)}$). L'idéal $I_{(0,0)}$ consiste en les polynômes $P=\sum_{i,j} c_{i,j}X^iY^j$ ayant un coefficient constant $c_{0,0}=0$, on peut alors écrire

$$P=\Big(\sum_{i\neq 0,j} c_{i,j}X^{i-1}Y^j\Big)X+\Big(\sum_{i=0,j\neq 0} c_{0,j}Y^{j-1}\Big)Y$$

et on en conclut que

$$I_{(0,0)}=\langle X,Y\rangle.$$

(d) On se place dans l'anneau $\mathbb{K}[X]$ des polynômes sur l'un des corps $\mathbb{K}=\mathbb{Q}$, $\mathbb{R}$ ou $\mathbb{C}$. Soient $w_1,\dots,w_s\in\mathbb{K}$ des points 2 à 2 distincts et $I_{w_1,\dots,w_s}$ l'ensemble des polynômes P tels que $P(w_1)=\dots=P(w_s)=0$. Alors $I_{w_1,\dots,w_s}$ est un idéal de $\mathbb{K}[X]$, et en fait les résultats de la section précédente montrent que

$$I_{w_1,\dots,w_s}=\langle (X-w_1)\dots(X-w_s)\rangle$$

consiste précisément en l'ensemble des polynômes

$$P=Q(X-w_1)\dots(X-w_s)$$

qui sont multiples de $(X-w_1)\dots(X-w_s)$. Plus généralement, si on se donne des multiplicités $m_1,\dots,m_s\in\mathbb{N}$ et qu'on considère l'ensemble $I_{w_1(m_1),\dots,w_s(m_j)}$ des polynômes P ayant une multiplicité m_i en w_i, c'est-à-dire tels que $P^{(j)}(w_i)=0$ pour tout i et tout $j=0,1,\dots,m_i-1$, alors

$$I_{w_1(m_1),\dots,w_s(m_s)}=\langle (X-w_1)^{m_1}\dots(X-w_s)^{m_s}\rangle$$

est l'idéal engendré par $(X-w_1)^{m_1}\dots(X-w_s)^{m_s}$.

2.1.3. Définition. *Soit A un anneau.*

(a) *Un idéal I de A est dit* principal *s'il admet un seul générateur, autrement dit, si on peut trouver $g\in I$ tel que $I=\langle g\rangle$ coïncide avec l'ensemble des multiples de g.*

(b) *L'anneau A lui-même est dit* principal *si A est intègre et si tout idéal I de A est principal.*

2.1.4. Exemples.

(a) Les idéaux $I_{w_1,\dots,w_s}$ et $I_{w_1(m_1),\dots,w_s(m_s)}$ de $\mathbb{K}[X]$ sont des idéaux principaux.

(b) L'idéal $I_{(0,0)} = \langle X, Y\rangle$ de $\mathbb{K}[X,Y]$ n'est pas principal. En effet si on avait $I_{(0,0)} = \langle G\rangle$ avec $G \in \mathbb{K}[X,Y]$, il existerait des polynômes $P, Q \in \mathbb{K}[X,Y]$ tels que $X = PG$ et $Y = QG$, donc $G \neq 0$, $G \mid X$ et $G \mid Y$. Mais la condition $G \mid X$ implique $\deg_Y(G) \leqslant \deg_Y(X) = 0$ et la condition $G \mid Y$ implique $\deg_X(G) \leqslant \deg_X(Y) = 0$. Par conséquent G serait une constante non nulle $c \in \mathbb{K}^*$ et on aurait $G(0,0) = c \neq 0$, ce qui est contradictoire puisque $G \in I_{(0,0)}$. Il en résulte que l'anneau $\mathbb{K}[X,Y]$ *n'est pas un anneau principal.*

Nous démontrons maintenant un théorème très important.

2.1.5. Théorème. *Tout anneau euclidien A est principal. En particulier $\mathbb{Z}$ est principal, et les anneaux de polynômes $\mathbb{K}[X]$ sur les corps sont principaux.**

Démonstration. Par hypothèse, il existe un stathme $v : A^* \to \mathbb{N}$ et un algorithme de division euclidienne :

$$\forall a \in A,\ \forall b \in A^*,\ \ \exists q, r \in A,\ \text{tels que } a = bq + r \ \text{ avec } r = 0 \text{ ou } r \neq 0,\ v(r) < v(b).$$

Soit I un idéal de A. Si $I = \{0\}$, alors $I = \langle 0\rangle$ est principal. Si $I \neq \{0\}$ on choisit un élément $g \in I \setminus \{0\}$ tel que le stathme $v(g) \in \mathbb{N}$ prenne la valeur minimale $\min_{x \in I\setminus\{0\}} v(x)$ (ce qui est toujours possible puisqu'on est dans les entiers naturels). Soit $x \in I$ quelconque. On effectue la division euclidienne de x par g :

$$\exists q, r \in A,\ \ x = gq + r \ \text{ avec } r = 0 \text{ ou } r \neq 0,\ v(r) < v(g).$$

Mais alors $r = x - gq = 1 \times x + (-q) \times g \in I$ car $x, g \in I$. D'après le choix de g on ne peut pas avoir $r \in I \setminus \{0\}$, $v(r) < v(g)$, c'est donc que $r = 0$ et que $x = qg$ est multiple de g. Par conséquent $I \subset \langle g\rangle$. Mais comme $g \in I$, on a aussi $\langle g\rangle = \{\lambda g \,/\, \lambda \in A\} \subset I$ et donc $I = \langle g\rangle$. □

2.1.6. Lien avec la divisibilité. Dans un anneau A intègre, les notions de divisibilité peuvent se relier de manière simple à l'inclusion et l'égalité des idéaux principaux.

(a) Si $x, y \in A^*$, alors $\langle x\rangle \subset \langle y\rangle$ si et seulement si $y \mid x$.

En effet, si $y \mid x$, soit $x = \alpha y$, tout multiple $\lambda x = \lambda\alpha y$ est aussi multiple de y, donc $\langle x\rangle \subset \langle y\rangle$. Réciproquement, si $\langle x\rangle \subset \langle y\rangle$, on a $x \in \langle y\rangle$, donc x est multiple de y, i.e. $y \mid x$.

(b) Si $x, y \in A^*$, alors $\langle x\rangle = \langle y\rangle$ si et seulement s'il existe $u \in A^\times$ tel que $y = ux$

(u étant donc inversible). On dit alors que x, y sont multiplicativement équivalents.

* En revanche, il n'est pas vrai que "A principal $\Longrightarrow$ $A[X]$ principal". Par exemple $A = \mathbb{K}[Y]$ est principal, mais $A[X] = \mathbb{K}[Y][X] = \mathbb{K}[X,Y]$ n'est pas principal.

En effet, d'après (a), si $\langle x\rangle = \langle y\rangle$, il existe $\alpha, \beta \in A^*$ tels que $x = \alpha y$ et $y = \beta x$, donc $x = \alpha\beta x$, et comme A est intègre, on en déduit que $\alpha\beta = 1$. On a donc bien $y = ux$ (et $x = u^{-1}y$) avec $u = \beta$ et $u^{-1} = \alpha$. La réciproque est claire.

(c) Si $x \in A^*$, alors $\langle x\rangle = \langle 1\rangle = A$ si et seulement si $x \in A^\times$, i.e. x inversible.

C'est un cas particulier du (b).

Une conséquence de ce qui précède est que dans toutes les questions liées à la divisibilité, on considère comme équivalents des éléments qui ne diffèrent que par un élément inversible. Par exemple, dans $\mathbb{Z}$, on considère 7 et -7 comme équivalents, et lorsqu'on a affaire à des éléments irréductibles p (nombres premiers), on pourra toujours choisir plutôt $p > 0$, à équivalence près. De même, l'ensemble des inversibles de $\mathbb{K}[X]$ consiste en les constantes $\alpha \in \mathbb{K}^*$, et si on a affaire à un polynôme irréductible $P \in \mathbb{K}[X]$, on pourra toujours diviser par son coefficient dominant de façon à se ramener à un polynôme unitaire.

2.2. Opérations sur les idéaux

Soit A un anneau et $I_1, I_2, \ldots, I_k$ des idéaux de A. On leur associe alors des nouveaux idéaux comme suit.

2.2.1. Intersection. $I = I_1 \cap I_2 \cap \ldots \cap I_k$ *est un idéal de* A.

Démonstration. Notons d'abord que $0 \in I$, donc I n'est pas vide. Soient $\lambda, \mu \in A$ et $x, y \in I$. Alors pour tout $\ell = 1, 2, \ldots, k$ on a $x, y \in I_\ell$ donc $\lambda x + \mu y \in I_\ell$. Par conséquent $\lambda x + \mu y \in I$, et I est bien un idéal.

2.2.2. Exemple. Dans $\mathbb{Z}$, l'intersection $\langle 4\rangle \cap \langle 10\rangle$ représente les entiers qui sont à la fois multiples de 4 et de 10, on voit qu'il s'agit donc des multiples de 20. Par conséquent

$$\langle 4\rangle \cap \langle 10\rangle = \langle 20\rangle.$$

2.2.3. Somme. *On définit*

$$I_1 + I_2 + \cdots + I_k = \big\{x_1 + x_2 + \cdots + x_k \,/\, x_\ell \in I_\ell\big\}.$$

Alors $I_1 + I_2 + \cdots + I_k$ *est un idéal de* A.

Démonstration. Comme les I_ℓ sont stables par combinaisons linéaires, la stabilité de la somme par combinaisons linéaires est également évidente. □

2.2.4. Exemple. Dans $\mathbb{Z}$, la somme $\langle 4\rangle + \langle 10\rangle$ est constitué d'entiers de la forme $4n + 10p$ qui sont tous pairs, donc $\langle 4\rangle + \langle 10\rangle \subset \langle 2\rangle$. Mais d'autre part $2 = (-2) \times 4 + 10 \in \langle 4\rangle + \langle 10\rangle$ donc $\langle 2\rangle \subset \langle 4\rangle + \langle 10\rangle$. Ceci implique

$$\langle 4\rangle + \langle 10\rangle = \langle 2\rangle.$$

2.2.5. Produit. (*La définition est plus subtile* !). *Pour des idéaux* I, J, *on pose*

$$IJ = \Big\{\sum_\ell x_\ell\, y_\ell \,/\, x_\ell \in I,\ y_\ell \in J\Big\}.$$

Il est facile de voir que IJ est bien un idéal (on a $0 \in IJ$, et IJ est stable par combinaisons linéaires). Plus généralement, on pose

$$I_1 I_2 \cdots I_k = \Big\{ \sum_\ell x_{1,\ell}\, x_{2,\ell} \cdots x_{k,\ell} \,/\, x_{1,\ell} \in I_1, x_{2,\ell} \in I_2, \ldots, x_{k,\ell} \in I_k \Big\},$$

les sommes étant toujours prises finies.

2.2.6. Exemple. Il n'est pas difficile de voir que dans un anneau A quelconque, un produit d'idéaux principaux est donné par la formule

$$\langle g \rangle \langle h \rangle = \langle gh \rangle,$$

par exemple $\langle 4 \rangle \langle 10 \rangle = \langle 40 \rangle$ dans $\mathbb{Z}$. Plus généralement, pour des idéaux non nécessairement principaux, on a (exercice !)

$$\langle g_i \rangle_{i \in S}\, \langle h_j \rangle_{j \in T} = \langle g_i h_j \rangle_{(i,j) \in S \times T}. \qquad \square$$

2.2.7. Réunion. En général, ce n'est pas un idéal ! Par exemple, dans $\mathbb{Z}$

$$\langle 4 \rangle \cup \langle 10 \rangle$$

contient 4 et 10, mais ne contient pas 14 qui n'est ni multiple de 4 ni multiple de 10, donc il n'y a pas stabilité pour l'addition. $\square$

2.2.8. Réunion croissante. Pour la réunion, il y a tout de même un cas intéressant, celui d'une suite croissante infinie d'idéaux

$$I_0 \subset I_1 \subset \cdots \subset I_k \subset \cdots, \quad k \in \mathbb{N}.$$

Alors la réunion $I = \bigcup_{k \in \mathbb{N}} I_k$ est bien un idéal. En effet, si on prend une combinaison linéaire $\lambda x + \mu y$ d'éléments $x, y \in I$ avec $\lambda, \mu \in A$, alors x appartient à un certain I_k et y à un certain I_ℓ, mais alors $x, y \in I_m$ avec $m = \max(k, \ell)$. Donc $\lambda x + \mu y \in I_m \subset I$. $\square$

2.2.9. Exemple*. Illustrons par un exemple une situation où on a une telle réunion croissante. Soit A l'anneau des fonctions $f : \mathbb{R}_+ \to \mathbb{R}$ qui peuvent s'écrire comme des somme finies

$$\mathbb{R}_+ \ni x \longmapsto f(x) = \sum\nolimits_\ell a_\ell\, x^{b_\ell}, \quad a_\ell \in \mathbb{R},\ b_\ell \in \mathbb{Q}_+.$$

Les exposants b_ℓ sont donc ici des rationnels positifs ou nuls, et non des entiers comme dans le cas des polynômes. Il est facile de voir que $(A, +, \times)$ est un anneau intègre (exercice !). Prenons pour I l'idéal des fonctions f telles que $f(0) = 0$: ce sont les fonctions dont le coefficient a_0 du terme constant $a_0\, x^0$ est nul, puisque tous les autres termes $a_\ell\, x^{b_\ell}$ avec $b_\ell > 0$ s'annulent en 0. D'autre part, soit $(\varepsilon_k)_{k \in \mathbb{N}}$ une suite strictement décroissante de rationnels positifs convergeant vers 0, par exemple $\varepsilon_k = 2^{-k}$. Pour $f \in I$, on a un plus petit exposant $b_m > 0$ qui intervient dans

l'écriture de f avec un coefficient $a_m \neq 0$, et alors f est divisible par la fonction $x \mapsto x^{\varepsilon_k}$ dès que $k \in \mathbb{N}^*$ est pris assez grand pour que $\varepsilon_k \leqslant b_m$, puisqu'alors tous les quotients $a_\ell\, x^{b_\ell}/x^{\varepsilon_k} = a_\ell\, x^{b_\ell-\varepsilon_k}$ sont d'exposants $b_\ell - \varepsilon_k \geqslant b_m - \varepsilon_k \geqslant 0$ dans $\mathbb{Q}_+$. Ceci montre que I est la réunion des idéaux principaux

$$I_k = \langle x \mapsto x^{\varepsilon_k} \rangle.$$

Cette réunion d'idéaux est strictement croissante, car $x^{\varepsilon_{k+1}}$ divise x^{ε_k}, sans que le quotient $x^{\varepsilon_k}/x^{\varepsilon_{k+1}} = x^{\varepsilon_k-\varepsilon_{k+1}}$ définisse une fonction inversible dans A. Il résulte du lemme ci-dessous que I n'est pas un idéal principal, et donc que A n'est pas un anneau principal.

2.2.10. Lemme (Emmy Noether). *Si A possède une suite infinie strictement croissante d'idéaux*

$$I_0 \subsetneq I_1 \subsetneq \cdots \subsetneq I_k \subsetneq I_{k+1} \subsetneq \cdots\ ,$$

alors l'idéal $I = \bigcup_{k\in\mathbb{N}}$ n'est pas principal, et donc l'anneau A n'est pas principal. Par conséquent, un anneau principal ne peut pas posséder une telle suite infinie strictement croissante d'idéaux.

Démonstration. Supposons $I = \langle g \rangle$. Alors $g \in I$ appartiendrait à un certain idéal I_k, et on aurait donc $I = \langle g \rangle \subset I_k$, par suite $I_{k+1} \subset I \subset I_k$, contradiction. □

2.3. PPCM, PGCD et algorithme d'Euclide

On suppose dans toute cette section que A est un *anneau principal*. Étant donné des éléments $x_1, \ldots, x_s \in A^*$, l'idéal intersection $\langle x_1 \rangle \cap \cdots \cap \langle x_s \rangle$ consiste en l'ensemble des multiples communs aux x_j. Cet idéal n'est pas réduit à $\{0\}$, puisqu'il contient $x_1 x_2 \cdots x_s$. Comme l'anneau A est principal, il existe un élément $m \in A^*$ tel que

$$\langle x_1 \rangle \cap \cdots \cap \langle x_s \rangle = \langle m \rangle,$$

et les multiples communs à $x_1, \ldots, x_s$ sont donc précisément les multiples de m.

2.3.1. Définition du ppcm. *Dans un anneau principal A, on appelle* plus petit commun multiple *de $x_1, \ldots, x_s \in A^*$, noté $m = \mathrm{ppcm}(x_1, \ldots, x_s)$ tout élément $m \in A^*$ tel que*

$$\langle x_1 \rangle \cap \cdots \cap \langle x_s \rangle = \langle m \rangle.$$

On observera que m n'est défini de manière unique qu'à un facteur inversible près $u \in A^\times$: l'élément $\widetilde{m} = um$ conviendrait aussi puisque $\langle \widetilde{m} \rangle = \langle m \rangle$. En fait, seul l'idéal $\langle m \rangle$ est défini de manière unique. Pour pallier cette absence d'unicité, on pourra convenir dans $\mathbb{Z}$ de toujours choisir $m > 0$, et dans $\mathbb{K}[X]$ de prendre un polynôme $M \in \mathbb{K}[X]$ qui soit unitaire, mais c'est un choix purement "esthétique". Dans le cas de deux éléments $x, y \in A^*$, on peut écrire

$$\langle x \rangle \cap \langle y \rangle = \langle \mathrm{ppcm}(x, y) \rangle. \tag{2.3.2}$$

L'opération ppcm correspond simplement à l'intersection des idéaux; comme l'intersection des idéaux est commutative et associative, il en est de même pour le ppcm :

2.3.3. Propriété. *Le* ppcm *est commutatif et associatif, c'est-à-dire*

$$\forall x, y \in A^*, \quad \operatorname{ppcm}(x,y) = \operatorname{ppcm}(y,x),$$
$$\forall x, y, z \in A^*, \quad \operatorname{ppcm}(x,y,z) = \operatorname{ppcm}(x, \operatorname{ppcm}(y,z)) = \operatorname{ppcm}(\operatorname{ppcm}(x,y), z)$$

(*les égalités ayant lieu à des éléments inversibles près*).

Pour calculer un ppcm d'un nombre quelconque d'éléments, on peut donc procéder dans un ordre quelconque et se ramener à des ppcm de 2 éléments seulement. On verra plus tard plusieurs méthodes de calcul pratique du ppcm (via le pgcd).

2.3.4. Cas du pgcd. Étant donné $x_1, \dots, x_s \in A^*$, l'idéal somme $\langle x_1 \rangle + \cdots + \langle x_s \rangle$ est la même chose que l'idéal engendré par les x_j :

$$\langle x_1 \rangle + \cdots + \langle x_s \rangle = \big\{ \lambda_1 x_1 + \cdots + \lambda_s x_s \,/\, \lambda_j \in A \big\} = \langle x_1, \dots, x_s \rangle.$$

Cet idéal n'est évidemment pas réduit à $\{0\}$, et comme l'anneau A est principal, il existe un élément $d \in A^*$ tel que

$$\langle x_1 \rangle + \cdots + \langle x_s \rangle = \langle d \rangle.$$

En particulier $x_j \in \langle d \rangle$, ce qui implique que d est un diviseur commun à $x_1, \dots, x_s$. Mais réciproquement, si $d' \in A^*$ est un diviseur commun à $x_1, \dots, x_s$, c'est-à-dire $x_j = k_j d'$, alors tout élément de l'idéal somme $\sum_{j=1}^p \lambda_j x_j = \big(\sum_{j=1}^p \lambda_j k_j\big) d'$ est multiple de d', et en particulier d doit être un multiple de d'. L'élément $d \in A^*$ est donc le "plus grand commun diviseur" des x_j, ou "plus grand" doit se comprendre au sens de la relation de divisibilité : $d' \mid d$.

2.3.5. Définition du pgcd. *Dans un anneau principal A, on appelle* plus grand commun diviseur *de $x_1, \dots, x_s \in A^*$, noté $d = \operatorname{pgcd}(x_1, \dots, x_s)$ tout élément $d \in A^*$, défini à un facteur inversible près $u \in A^\times$, tel que*

$$\langle x_1 \rangle + \cdots + \langle x_s \rangle = \langle d \rangle.$$

Dans le cas de deux éléments $x, y \in A^*$, on peut écrire

$$\langle x \rangle + \langle y \rangle = \langle \operatorname{pgcd}(x,y) \rangle, \tag{2.3.6}$$

et comme l'addition des idéaux est commutative et associative, on en déduit aussitôt :

2.3.7. Propriété. *Le pgcd est commutatif et associatif, c'est-à-dire*

$$\forall x, y \in A^*, \quad \operatorname{pgcd}(x,y) = \operatorname{pgcd}(y,x),$$
$$\forall x, y, z \in A^*, \quad \operatorname{pgcd}(x,y,z) = \operatorname{pgcd}(x, \operatorname{pgcd}(y,z)) = \operatorname{pgcd}(\operatorname{pgcd}(x,y), z)$$

(*les égalités ayant lieu à des éléments inversibles près*).

Une conséquence immédiate de la définition du pgcd est l'identité dite de Bézout ou de Bachet-Bézout (du nom des mathématiciens Français Claude-Gaspard Bachet de Méziriac, 1581–1638, et Étienne Bézout, 1730–1783) :

2.3.8. Identité de Bézout. *Si* $d = \mathrm{pgcd}(x_1, \ldots, x_s)$ *avec* $x_1, \ldots, x_s \in A^*$, *il existe des éléments* $\lambda_1, \ldots, \lambda_s \in A$ *tels que*

$$\lambda_1 x_1 + \cdots + \lambda_s x_s = d.$$

2.3.9. Théorème et définition. *On dit que les éléments* $x_1, \ldots, x_s \in A^*$ *sont* premiers entre eux dans leur ensemble *si*

$$\mathrm{pgcd}(x_1, \ldots, x_s) = 1.$$

Pour cela, il faut et il suffit que

$$\exists\, \lambda_1, \ldots, \lambda_s \in A \quad \textit{tels que} \quad \lambda_1 x_1 + \cdots + \lambda_s x_s = 1.$$

Démonstration. En effet, l'existence d'éléments λ_j comme ci-dessus est bien équivalente au fait que l'idéal engendré $\langle x_1 \rangle + \cdots + \langle x_s \rangle$ coïncide avec l'idéal unité $\langle 1 \rangle = A$. □

2.3.10. Remarque. Si $\mathrm{pgcd}(x_1, \ldots, x_s) = d$ alors on peut "simplifier" le diviseur commun d en écrivant $x_1 = dx'_1, \ldots, x_s = dx'_s$ et l'identité de Bézout 2.3.5 pour $x_1, \ldots, x_s$ implique $\lambda_1 x'_1 + \cdots + \lambda_s x'_s = 1$ après simplification, donc

$$\mathrm{pgcd}(x'_1, \ldots, x'_s) = 1.$$

2.3.11. Attention. Dans $\mathbb{Z}$, on a $\mathrm{pgcd}(6, 10, 15) = 1$ (puisque par exemple $6 + 10 - 15 = 1$), mais $\mathrm{pgcd}(6, 10) = 2$, $\mathrm{pgcd}(6, 15) = 3$, $\mathrm{pgcd}(10, 15) = 5$, il n'y a donc pas équivalence entre le fait que $x_1, \ldots, x_s$ soient premiers entre eux dans leur ensemble, ou premiers entre eux deux à deux (ce qui est une hypothèse bien plus forte d'après 2.3.7).

2.3.12. Distributivité de la multiplication par rapport au ppcm et pgcd. *Pour tous* $a \in A^*$ *et* $x_1, \ldots, x_s \in A^*$, *on a*

(a) $\mathrm{ppcm}(ax_1, \ldots, ax_s) = a\ \mathrm{ppcm}(x_1, \ldots, x_s)$,

(b) $\mathrm{pgcd}(ax_1, \ldots, ax_s) = a\ \mathrm{pgcd}(x_1, \ldots, x_s)$.

Démonstration. (a) Il s'agit de voir que $\langle ax_1 \rangle \cap \ldots \cap \langle ax_s \rangle = \langle a \rangle \big(\langle x_1 \rangle \cap \ldots \cap \langle x_s \rangle \big)$. Or $\langle ax_1 \rangle \cap \ldots \cap \langle ax_s \rangle$ consiste en les éléments $y \in A$ tels qu'il existe $\lambda_1, \ldots, \lambda_s \in A$ vérifiant

$$y = \lambda_1 a x_1 = \ldots = \lambda_s a x_s.$$

Mais comme A est intègre, on peut simplifier les égalités par a, ce qui donne

$$y = az \quad \text{avec} \quad z = \lambda_1 x_1 = \ldots = \lambda_s x_s \in \langle x_1\rangle \cap \ldots \cap \langle x_s\rangle.$$

Ceci implique $\langle ax_1\rangle \cap \ldots \cap \langle ax_s\rangle \subset \langle a\rangle\big(\langle x_1\rangle \cap \ldots \cap \langle x_s\rangle\big)$. L'inclusion inverse est évidente.

(b) L'affirmation découle de la distributivité de l'addition par rapport à la multiplication, qui implique aussitôt

$$a(\lambda_1 x_1 + \cdots + \lambda_s x_s) = \lambda_1(ax_1) + \ldots + \lambda_s(ax_s)$$

(en tenant compte aussi de l'associativité et de la commutativité de $\times$) et donc

$$\langle a\rangle\big(\langle x_1\rangle + \cdots + \langle x_s\rangle\big) = \langle ax_1\rangle + \cdots + \langle ax_s\rangle. \qquad \square$$

2.3.13. Algorithme d'Euclide. On suppose ici que A est un *anneau euclidien*, muni d'un stathme $v : A^* \to \mathbb{N}$. On cherche à calculer le pgcd de deux éléments $a, b \in A^*$. Soit $v_1 = \min(v(a), v(b))$ le minimum de la valeur du stathme pour a et b. On peut toujours ordonner les éléments en sorte que $v(b) \leqslant v(a)$, de façon que $v_1 = v(b)$ [on choisit pour b le "plus petit" des deux éléments]. On effectue alors la division euclidienne de a par b. Ceci donne des éléments $q, r \in A$ tels que

$$a = bq + r \quad \text{avec } r = 0 \text{ ou } r \neq 0,\ v(r) < v(b).$$

Or les combinaisons linéaires de a, b peuvent s'écrire

$$\lambda a + \mu b = \lambda(bq + r) + \mu b = (\lambda q + \mu)b + \lambda r,$$

et inversement les combinaisons linéaires de b, r peuvent s'écrire

$$\lambda' b + \mu' r = \lambda' b + \mu'(a - bq) = \mu' a + (\lambda' - \mu' q)b.$$

Ceci implique $\langle a\rangle + \langle b\rangle = \langle b\rangle + \langle r\rangle$, c'est-à-dire

$$\operatorname{pgcd}(a, b) = \operatorname{pgcd}(b, r).$$

Si $r = 0$, on a $b \mid a$ et $\operatorname{pgcd}(a, b) = b$, tandis que si $r \in A^*$, on peut procéder par récurrence en posant $r_{-1} = a$, $r_0 = b$, $q_0 = q$ et $r_1 = r$, ce qui donne $r_{-1} = q_0 r_0 + r_1$ avec $v(r_1) = v(r) < v(b) = v(r_0)$. L'égalité $\operatorname{pgcd}(a, b) = \operatorname{pgcd}(b, r)$ implique

$$\operatorname{pgcd}(r_{-1}, r_0) = \operatorname{pgcd}(r_0, r_1), \quad \text{et on a } \ v(r_1) < v(r_0).$$

On peut alors procéder inductivement. Tant que les restes r_i obtenus sont non nuls, on produit ainsi des couples successifs $(r_{i-1}, r_i)_{i\geqslant 0}$ en partant de (a, b) et en effectuant des divisions euclidiennes

$$r_{i-1} = q_i r_i + r_{i+1}, \quad \text{avec } r_{i+1} = 0 \text{ ou } r_{i+1} \neq 0,\ v(r_{i+1}) < v(r_i).$$

On obtient ainsi une suite strictement décroissante de valeurs $v(r_i)$ avec

$$r_{i+1} = 0 \quad \text{ou} \quad v(r_{i+1}) < v(r_i) < \cdots < v(r_1) < v(r_0).$$

Comme il s'agit d'une suite strictement décroissante d'entiers naturels, le procédé s'arrête nécessairement, ce qui implique qu'on finit par obtenir $r_{i+1} = 0$ à une certaine étape, par conséquent $r_i \mid r_{i-1}$ et

$$\operatorname{pgcd}(a,b) = \operatorname{pgcd}(r_{-1}, r_0) = \operatorname{pgcd}(r_0, r_1) = \ldots = \operatorname{pgcd}(r_{i-1}, r_i) = r_i.$$

On retiendra la règle suivante :

2.3.14. Règle. *Dans l'algorithme d'Euclide,* $\operatorname{pgcd}(a,b)$ *est égal au dernier reste* r_i *non nul calculé (ou à* $r_0 = b$*, si* $b \mid a$*).* □

2.3.15. Retour sur l'identité de Bézout. Un autre mérite de l'algorithme d'Euclide est de permettre le calcul effectif d'éléments $\lambda, \mu \in A$ tels que $\lambda a + \mu b = d$. En effet, d'après la règle ci-dessus, on a

$$\begin{aligned} d = r_i &= r_{i-2} - q_{i-1} r_{i-1} \\ &= r_{i-2} - q_{i-1}(r_{i-3} - q_{i-2} r_{i-2}) \\ &= (q_{i-1} q_{i-2} + 1) r_{i-2} - q_{i-1} r_{i-3} \\ &= (q_{i-1} q_{i-2} + 1)(r_{i-4} - q_{i-3} r_{i-3}) - q_{i-1} r_{i-3} \ \ldots, \end{aligned}$$

et on peut ainsi remonter jusqu'à $r_{-1} = a$, $r_0 = b$.

2.3.16. Exemples. (a) Dans $\mathbb{Z}$, on demande de calculer $d = \operatorname{pgcd}(1662, 1356)$ et de déterminer des entiers $\lambda, \mu \in \mathbb{Z}$ tels que $\lambda\, 1662 + \mu\, 1356 = d$. On effectue pour cela des divisions successives en prenant à chaque fois les restes obtenus comme nouveaux diviseurs :

$$\begin{aligned} 1662 &= 1 \times 1356 + 306, \\ 1356 &= 4 \times 306 + 132, \\ 306 &= 2 \times 132 + 42, \\ 132 &= 3 \times 42 + 6, \\ 42 &= 7 \times 6 + 0. \end{aligned}$$

Le pgcd est le dernier reste non nul, donc $\operatorname{pgcd}(1662, 1356) = 6$, et en effet on a bien $1662 = 6 \times 277$, $1356 = 6 \times 226$ avec $\operatorname{pgcd}(277, 226) = 1$. Pour trouver les coefficients λ et μ, on "remonte" dans les divisions en écrivant

$$\begin{aligned} 6 &= 132 - 3 \times 42 = 132 - 3 \times (306 - 2 \times 132) \\ &= -3 \times 306 + 7 \times 132 = -3 \times 306 + 7 \times (1356 - 4 \times 306) \\ &= 7 \times 1356 - 31 \times 306 = 7 \times 1356 - 31 \times (1662 - 1 \times 1356) \\ &= -31 \times 1662 + 38 \times 1356. \end{aligned}$$

Une solution possible est donc $(\lambda,\mu) = (-31, 38)$. Cette solution n'est pas unique : il est facile de voir que pour tout entier $k \in \mathbb{Z}$, le couple $(\lambda,\mu) = (-31 - 226\,k, 38 + 277\,k)$ vérifie encore

$$\begin{aligned}\lambda\,1662 + \mu\,1356 &= (-31 - 226\,k) \times 1662 + (38 + 277\,k) \times 1356 \\ &= 6 + k(-226 \times 6 \times 277 + 277 \times 6 \times 226) = 6.\end{aligned}$$

(b) Considérons maintenant un exemple de calcul de $\operatorname{pgcd}(A, B)$ pour des polynômes A, B de l'anneau $\mathbb{K}[X]$, à savoir le pgcd de

$$A = X^a - 1, \quad B = X^b - 1, \quad a, b \in \mathbb{N}^*.$$

On supposera par exemple $a \geqslant b$, et on utilise l'algorithme d'Euclide pour calculer $d = \operatorname{pgcd}(a, b)$. Pour cela, on commence par effectuer une division $a = bq + r$, $0 \leqslant r < b$. On a l'identité

$$X^{bq} - 1 = (X^b)^q - 1 = (X^b - 1)(X^{b(q-1)} + X^{b(q-2)} + \cdots + X^b + 1),$$

ce qui donne

$$X^a - 1 = X^{bq+r} - 1 = (X^{bq} - 1)X^r + (X^r - 1),$$

par conséquent

$$X^a - 1 = (X^b - 1)(X^{b(q-1)} + X^{b(q-2)} + \cdots + X^b + 1)X^r + (X^r - 1),$$

soit $A = BQ + R$, $\deg(R) < \deg(B)$, avec

$$Q = (X^{b(q-1)} + X^{b(q-2)} + \cdots + X^b + 1)X^r, \quad R = X^r - 1.$$

D'après l'algorithme d'Euclide appliqué dans l'anneau $\mathbb{K}[X]$, on en conclut que

$$\operatorname{pgcd}(X^a - 1, X^b - 1) = \operatorname{pgcd}(A, B) = \operatorname{pgcd}(B, R) = \operatorname{pgcd}(X^b - 1, X^r - 1).$$

Si (r_{i-1}, r_i) est la suite produite par les divisions successives jusqu'à l'obtention d'un reste $r_{i+1} = 0$, on obtient $R_{i+1} = X^{r_{i+1}} - 1 = 0$, donc $\operatorname{pgcd}(a, b) = r_i$ et $\operatorname{pgcd}(X^a - 1, X^b - 1) = X^{r_i} - 1$, d'où la formule amusante

$$\operatorname{pgcd}(X^a - 1, X^b - 1) = X^{\operatorname{pgcd}(a,b)} - 1.$$

2.4. Décomposition en facteurs irréductibles

2.4.1. Un exemple d'anneau non factoriel. Pour donner un exemple de situation où les choses "se passent mal", considérons l'anneau noté $\mathbb{Z}[i\sqrt{5}]$ tel que

$$\mathbb{Z}[i\sqrt{5}] = \left\{a + bi\sqrt{5} \,/\, a, b \in \mathbb{Z}\right\} \subset \mathbb{C}.$$

Il s'agit bien d'un anneau, puisque la multiplication est une loi interne (comme l'est aussi trivialement l'addition) :

$$(a+bi\sqrt{5})(a'+b'i\sqrt{5}) = (aa'+5bb') + (ab'+ba')i\sqrt{5}.$$

Pour $z = a+bi\sqrt{5} \in \mathbb{Z}[i\sqrt{5}]$, on a $\overline{z} = a-bi\sqrt{5} \in \mathbb{Z}[i\sqrt{5}]$, et on définit

$$N(z) = z\overline{z} = a^2+5b^2 \in \mathbb{N}.$$

Si $z, z' \in \mathbb{Z}[i\sqrt{5}]$, il vient $N(zz') = N(z)N(z')$. Lorsque $u \in \mathbb{Z}[i\sqrt{5}]^\times$ est inversible d'inverse u', la relation $uu' = 1$ implique $N(u)N(u') = 1$ et la seule possibilité est que $N(u) = 1$. Réciproquement, si $N(u) = 1$, il vient $u\overline{u} = 1$, et u est inversible d'inverse $\overline{u}$. On voit alors que les éléments inversibles de l'anneau sont les $u = a+bi\sqrt{5}$ tels que $a^2+5b^2 = 1$. La seule possibilité dans les entiers est $a = \pm 1$, $b = 0$, donc $\mathbb{Z}[i\sqrt{5}]^\times = \{1,-1\}$. Maintenant, l'élément 6 de l'anneau admet les décompositions

$$6 = 2\times 3 = (1+i\sqrt{5})\times(1-i\sqrt{5}),$$

et nous affirmons que les quatre éléments $2, 3, 1+i\sqrt{5}, 1-i\sqrt{5}$ sont irréductibles. En effet

$$N(2) = 4, \quad N(3) = 9, \quad N(1+i\sqrt{5}) = N(1-i\sqrt{5}) = 6,$$

mais il n'existe pas d'éléments $z = a+bi\sqrt{5}$, $a, b \in \mathbb{Z}$ tels que $N(z) = a^2+5b^2 = 2$ ou $N(z) = a^2+5b^2 = 3$, donc les seules "décompositions" possibles de $2, 3$, $1+i\sqrt{5}, 1-i\sqrt{5}$ comportent nécessairement l'un des facteurs inversibles ± 1. L'anneau $\mathbb{Z}[i\sqrt{5}]$ a la propriété surprenante que l'élément 6 possède deux décompositions en facteurs irréductibles totalement différentes !* □

Lorsque l'anneau considéré est principal, la décomposition en facteurs irréductibles a les propriétés attendues. On commence pour cela par démontrer quelques résultats préliminaires, énoncés par Gauss dans ses Disquisitiones arithmeticae (1801), mais probablement déjà connus antérieurement par des mathématiciens comme Fermat (1607–1665).

2.4.2. Lemme (Gauss). *On considère un anneau principal A et des éléments $a, b_1, \ldots, b_s, b, c, p \in A^*$.*

(a) *On suppose que $a \mid bc$ et* $\mathrm{pgcd}(a,b) = 1$. *Alors $a \mid c$.*

(b) *On suppose que* $\mathrm{pgcd}(a,b_1) = 1, \ldots, \mathrm{pgcd}(a,b_s) = 1$. *Alors* $\mathrm{pgcd}(a, b_1 \ldots b_s) = 1$.

* Au niveau des idéaux, les choses se passent en revanche beaucoup mieux. Si on introduit les idéaux non principaux

$$I = \langle 2, 1+i\sqrt{5}\rangle, \quad J = \langle 2, 1-i\sqrt{5}\rangle, \quad K = \langle 3, 1+i\sqrt{5}\rangle, \quad L = \langle 3, 1-i\sqrt{5}\rangle,$$

le lecteur vérifiera facilement (exercice !) que

$$\langle 2\rangle = IJ, \quad \langle 3\rangle = KL, \quad \langle 1+i\sqrt{5}\rangle = IK, \quad \langle 1-i\sqrt{5}\rangle = JL, \quad \langle 6\rangle = IJKL.$$

C'est de là historiquement que vient la terminologie de "nombres idéaux", permettant de rétablir un substitut adéquat à la décomposition bancale en facteurs irréductibles, de la même manière qu'on avait inventé les "imaginaires" pour suppléer à l'inexistence de $\sqrt{-1}$ dans $\mathbb{R}$.

(c) *Si p est irréductible et $p \nmid a$, alors* $\operatorname{pgcd}(p,a) = 1$.

(d) *Si p est irréductible et $p \mid b_1 \dots b_s$, alors il existe j tel que $p \mid b_j$.*

Démonstration. (a) Il est clair que $a \mid ac$ et par hypothèse $a \mid bc$, donc

$$a \mid \operatorname{pgcd}(ac,bc), \quad \text{et d'autre part} \quad \operatorname{pgcd}(ac,bc) = c\operatorname{pgcd}(a,b) = c.$$

(b) Par récurrence sur s, il suffit de le voir pour $s = 2$, disons pour $b_1 = b$ et $b_2 = c$. Or si $\operatorname{pgcd}(a,b) = 1$ et $\operatorname{pgcd}(a,c) = 1$, il existe $\lambda, \mu, \lambda', \mu' \in A$ tels que

$$\lambda a + \mu b = 1, \quad \lambda' a + \mu' c = 1,$$

par conséquent

$$1 = (\lambda a + \mu b)(\lambda' a + \mu' c) = (\lambda\lambda' a + \mu\lambda' b + \lambda\mu' c)a + (\mu\mu')bc = 1,$$

ce qui montre que $\operatorname{pgcd}(a,bc) = 1$.

(c) Si p est irréductible, les seules "décompositions" possibles de p sont de la forme $p = u(u'p)$ avec $u, u' \in A^\times$, $uu' = 1$, donc les seuls diviseurs de p sont les inversibles u et les éléments de la forme $u'p$. Mais par hypothèse $p \nmid a$ équivaut à dire que $u'p \nmid a$, donc seuls restent les éléments inversibles u comme diviseurs communs possibles à p et a. Ceci montre bien que $\operatorname{pgcd}(p,a) = 1$.

(d) Si p ne divisait aucun des b_j, on aurait $\operatorname{pgcd}(p,b_j) = 1$ d'après (c), et donc $\operatorname{pgcd}(p, b_1 \dots b_s) = 1$ d'après (b), contradiction. On peut aussi déduire (d) de (a) et (c) par récurrence sur s (la propriété étant triviale si $s = 1$) : si $p \nmid b_1$ alors $\operatorname{pgcd}(p,b_1) = 1$ et comme $p \mid b_1(b_2 \dots b_s)$, la propriété (a) implique $p \mid b_2 \dots b_s$, d'où la conclusion par hypothèse de récurrence. □

2.4.3. Définition. *Soit A un anneau intègre. Une partie $\mathcal{P} \subset A^*$ sera appelé sous-ensemble "représentatif" des éléments irréductibles de A si $\mathcal{P}$ contient exactement un élément p dans chaque classe d'équivalence $\{up\}$ d'éléments multiplicativement équivalents, de sorte que $\mathcal{P} \ni p \mapsto \langle p \rangle$ soit une bijection de $\mathcal{P}$ sur l'ensemble des idéaux principaux de A.*

Par exemple, dans $\mathbb{Z}$, on choisira pour $\mathcal{P}$ l'ensemble des nombres premiers $p > 0$, et dans $\mathbb{K}[X]$ on choisira l'ensemble des polynômes irréductibles qui sont unitaires.

2.4.4. Définition. *Soit A un anneau intègre et $\mathcal{P} \subset A^*$ un ensemble "représentatif" des éléments irréductibles. On dit que A est* factoriel *si tout élément $x \in A^*$ peut se décomposer sous la forme d'un produit*

$$x = u\, p_1^{m_1} p_2^{m_2} \cdots p_s^{m_s}, \quad u \in A^\times, \quad p_j \in \mathcal{P}, \quad m_j \in \mathbb{N}^*$$

(les $p_j \in \mathcal{P}$ étant deux à deux distincts et s étant éventuellement nul si $x = u$ est inversible), et si de plus la décomposition est unique *à l'ordre près des facteurs $p_j^{m_j}$.*

2.4.5. Théorème. *Tout anneau principal A est factoriel.*

En résumé, pour un anneau intègre A quelconque, on a la chaîne d'implications

$$\text{A euclidien} \;\Rightarrow\; \text{A principal} \;\Rightarrow\; \text{A factoriel},$$

et on peut montrer par des exemples (que nous n'étudierons pas ici) que les implications réciproques ne sont pas vraies.

Démonstration du théorème 2.4.5. Démontrons d'abord *l'existence* de la décomposition 2.4.4. Supposons par l'absurde qu'on ait un élément $x_0 \in A^*$ non décomposable en facteurs irréductibles. Alors x_0 n'est ni inversible ni irréductible, sinon on aurait $x_0 = u$, resp. $x_0 = up$ avec $p \in \mathcal{P}$ et $u \in A^\times$. Par conséquent x_0 peut s'écrire comme un produit $x_0 = x_1 x_1'$ d'éléments $x_1, x_1' \in A^*$ non inversibles. Nécessairement l'un au moins des éléments x_1, x_1' est non décomposable en facteurs irréductibles (sinon $x_0 = x_1 x_1'$ le serait !). Quitte à échanger x_1, x_1', on peut supposer que x_1 est non décomposable. Par récurrence, on construit ainsi une suite $x_{k-1} = x_k x_k'$, $k \in \mathbb{N}^*$, avec x_k non décomposable et x_k' non inversible. On obtient alors une suite infinie strictement croissante d'idéaux

$$\langle x_0\rangle \subsetneq \langle x_1\rangle \subsetneq \cdots \subsetneq \langle x_{k-1}\rangle \subsetneq \langle x_k\rangle \subsetneq \cdots\ ,$$

ce qui contredit le lemme de Noether 2.2.10. Cette contradiction montre que tout élément $x \in A^*$ est bien décomposable en facteurs irréductibles.

En ce qui concerne *l'unicité*, on raisonne par récurrence sur la "multiplicité totale" $m = \sum_{1\leqslant j\leqslant s} m_j$ de l'une des décompositions $x = u\, p_1^{m_1} p_2^{m_2} \cdots p_s^{m_s}$. Si $m = 0$, i.e. $s = 0$, on voit que $x = u$ est inversible, et dans ce cas x ne peut être divisible par aucun facteur irréductible q (sinon $x = u = \lambda q$, $\lambda \in A^*$, et donc $1 = (u^{-1}\lambda)q$, ce qui est contradictoire puisque q n'est pas inversible) ; lorsque $m = 0$, $x = u$ est donc la seule décomposition possible. En général, supposons qu'on ait un élément x possédant deux décompositions

$$x = u\, p_1^{m_1} p_2^{m_2} \cdots p_s^{m_s} = v\, q_1^{r_1} q_2^{r_2} \cdots q_t^{r_t}, \quad p_j, q_j \in \mathcal{P}, \quad m_j, r_j \in \mathbb{N}^*, \quad u, v \in A^\times$$

et supposons l'unicité déjà démontrée pour $m-1$ avec $m = \sum m_j \geqslant 1$. Alors en particulier p_1 divise le produit $v\, q_1^{r_1} q_2^{r_2} \cdots q_t^{r_t}$. D'après le lemme de Gauss, p_1 doit diviser l'un des facteurs. Mais p_1 qui est irréductible ne peut diviser l'élément inversible v. Donc p_1 divise l'un des q_j, et l'irréductibilité de q_j implique alors $q_j = w p_1$ avec $w \in A^\times$ inversible. Puisque $p_1, q_j \in \mathcal{P}$ et que $\mathcal{P}$ est représentatif, on en déduit $p_1 = q_j$. Quitte à permuter les facteurs q_j, on peut supposer $p_1 = q_1$. Après simplification, on trouve

$$u\, p_1^{m_1-1} p_2^{m_2} \cdots p_s^{m_s} = v\, q_1^{r_1-1} q_2^{r_2} \cdots q_t^{r_t}.$$

Comme l'unicité est supposée vraie pour $m-1$ par hypothèse de récurrence, on conclut après permutation des facteurs que $t = s$, $q_j = p_j$ et $r_j = m_j$. □

2.4.6. Remarque. Dans l'exemple de l'anneau $A = \mathbb{Z}[i\sqrt{5}]$, on peut voir que la décomposition en facteurs irréductibles existe bien, même si elle n'est pas

unique : ceci résulte du fait que si $x \in A^*$ s'écrit comme un produit $\prod y_j$ alors $N(x) = \prod N(y_j) \in \mathbb{N}^*$ et lorsqu'on a atteint des facteurs ayant des valeurs $N(y_j)$ minimales, les y_j sont nécessairement irréductibles. En revanche, dans le cas de l'anneau A introduit en 2.2.9, la fonction $f(x) = x^b$ vérifie par exemple $f(x) = (x^{b/2})^2 = \ldots = (x^{b/2^n})^{2^n}$ et on ne peut jamais atteindre d'éléments irréductibles, donc la décomposition en facteurs irréductibles n'existe pas !

2.4.7. Application. (a) Soit $\mathcal{P} \subset \mathbb{N}^*$ l'ensemble des nombres premiers usuels. Alors tout nombre rationnel $x \in \mathbb{Q}^*$ peut s'écrire de manière unique

$$(*) \qquad x = \pm \prod_{p \in \mathcal{P}} p^{m_p}, \quad m_p \in \mathbb{Z}$$

avec une suite d'entiers relatifs $(m_2, m_3, m_5, m_7, \ldots)$ presque tous nuls. En effet si on écrit $x = \frac{a}{b}$ avec $a \in \mathbb{Z}^*$ et $b \in \mathbb{N}^*$, l'existence de la décomposition $(*)$ provient de la décomposition de a et b en facteurs premiers. Pour l'unicité, on utilise le fait qu'une égalité $\prod_{p\in\mathcal{P}} p^{\alpha_p} = \prod_{p\in\mathcal{P}} p^{\beta_p}$ avec les α_p, β_p presque tous nuls se ramène à $\prod_{p\in\mathcal{P}} p^{\gamma_p} = 1$ avec $\gamma_p = \beta_p - \alpha_p \in \mathbb{Z}$, puis à

$$\prod_{p\in\mathcal{P}'} p^{\gamma_p} = \prod_{p\in\mathcal{P}''} p^{-\gamma_p}, \quad \mathcal{P}' = \{p \in \mathcal{P} \,/\, \gamma_p > 0\}, \quad \mathcal{P}'' = \{p \in \mathcal{P} \,/\, \gamma_p < 0\}.$$

Comme $\mathcal{P}'$, $\mathcal{P}''$ sont disjoints, l'égalité ci-dessus n'est possible que si $\mathcal{P}' = \mathcal{P}'' = \emptyset$, ce qui implique $\gamma_p = 0$ pour tout $p \in \mathcal{P}$. La notation standard est

$$x = \varepsilon \prod_{p \in \mathcal{P}} p^{v_p(x)}$$

où $\varepsilon = \pm 1$ et où $m_p = v_p(x) \in \mathbb{Z}$ s'appelle la *valuation p-adique* de x. Une autre façon d'interpréter le résultat ci-dessus est de dire qu'on a un isomorphisme de groupes

$$(\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}^{(\mathcal{P})}, +) \longrightarrow (\mathbb{Q}^*, \times), \quad (a, m_2, m_3, m_5, \ldots) \longmapsto (-1)^a \prod_{p\in\mathcal{P}} p^{m_p}$$

où $a = 0$ ou 1 modulo 2 et $\mathbb{Z}^{(\mathcal{P})}$ désigne l'ensemble des familles d'entiers relatifs $(m_p)_{p\in\mathcal{P}}$ presque tous nuls.

(b) Étant donné $x \in \mathbb{Q}_+^*$ et un entier $q \geqslant 2$, on se demande quelle est la condition nécessaire et suffisante pour que $\sqrt[n]{x} \in \mathbb{Q}$. Pour que ce soit le cas, il faut et il suffit que $x = y^n$ avec $y = \sqrt[n]{x} \in \mathbb{Q}_+^*$. Cette égalité se traduit sous la forme

$$\prod_{p\in\mathcal{P}} p^{v_p(x)} = \left(\prod_{p\in\mathcal{P}} p^{v_p(y)} \right)^n = \prod_{p\in\mathcal{P}} p^{n\, v_p(y)},$$

il faut donc que $v_p(x) = n\, v_p(y)$ soit multiple de n pour tout $p \in \mathcal{P}$; on peut alors prendre y tel que $v_p(y) = \frac{1}{n} v_p(x) \in \mathbb{Z}$. La condition nécessaire et suffisante

est donc que toutes les valuations p-adiques $v_p(x)$ soient multiples de n. Ainsi $\sqrt[3]{729/500} \notin \mathbb{Q}$, car $729/500 = 2^{-2}3^6 5^{-3}$ et l'exposant $-2 = v_2(729/500)$ n'est pas multiple de $n = 3$.

2.4.8. Généralisation. Soit A un anneau factoriel quelconque et $\mathcal{P}$ un ensemble représentatif d'éléments irréductibles. Tout élément $x \in A^*$ s'exprime sous forme d'un unique produit

$$x = \varepsilon(x) \prod_{p\in\mathcal{P}} p^{v_p(x)}, \quad v_p(x) \in \mathbb{N},$$

avec $\varepsilon(x) \in A^\times$ inversible. Il est alors facile de vérifier la proposition suivante.

2.4.9. Proposition. *Soient A un anneau factoriel A, et $x, y \in A^*$. Alors $x \mid y$ si et seulement si pour tout $p \in \mathcal{P}$ on a $v_p(x) \leqslant v_p(y)$ (i.e. si l'exposant de p dans la factorisation en irréductibles de x est inférieur à l'exposant correspondant de la factorisation de y).*

De là on tire aussi que pour $x, y \in A^*$ donnés, les éléments z qui sont multiples à la fois de x et y (resp. qui divisent à la fois x et y) sont ceux tels que $v_p(z) \geqslant \max(v_p(x), v_p(y))$ (resp. tels que $v_p(z) \leqslant \min(v_p(x), v_p(y))$. Ceci permet d'étendre comme suit les formules de calcul du pgcd et du ppcm à tous les anneaux factoriels (et plus seulement aux anneaux principaux).

2.4.10. Théorème et définition. *Dans un anneau factoriel, le ppcm et le pgcd peuvent se définir pour tous $x, y \in A^*$ par les formules*

(a) $$\mathrm{ppcm}(x,y) = \prod_{p\in\mathcal{P}} p^{\max(v_p(x),v_p(y))};$$

(b) $$\mathrm{pgcd}(x,y) = \prod_{p\in\mathcal{P}} p^{\min(v_p(x),v_p(y))}.$$

(c) *On a alors l'identité* $\langle \mathrm{pgcd}(x,y)\rangle\langle \mathrm{ppcm}(x,y)\rangle = \langle xy\rangle$.

Démonstration. Les formules (a) et (b) ont déjà été justifiées dans la discussion préliminaire. Pour vérifier (c), on applique (a) et (b), ce qui donne

$$\langle \mathrm{pgcd}(x,y)\rangle\langle \mathrm{ppcm}(x,y)\rangle = \left\langle \prod_{p\in\mathcal{P}} p^{\min(v_p(x),v_p(y)+\max(v_p(x),v_p(y))} \right\rangle.$$

Mais pour des entiers quelconques u, v, il est clair que $\min(u,v)+\max(u,v) = u+v$. Ceci donne

$$\begin{aligned}\langle \mathrm{pgcd}(x,y)\rangle\langle \mathrm{ppcm}(x,y)\rangle &= \left\langle \prod_{p\in\mathcal{P}} p^{v_p(x)+v_p(y)} \right\rangle = \left\langle \prod_{p\in\mathcal{P}} p^{v_p(x)} \prod_{p\in\mathcal{P}} p^{v_p(y)} \right\rangle \\ &= \langle xy\rangle.\end{aligned}$$

□

2.5. Éléments irréductibles de $\mathbb{Q}[X]$, $\mathbb{R}[X]$ et $\mathbb{C}[X]$

On considère ici l'anneau des polynômes $\mathbb{K}[X]$ à une indéterminée à coefficients dans un corps commutatif $\mathbb{K}$. D'après ce que nous avons vu au paragraphe précédent, tout polynôme $F \in \mathbb{K}[X]^*$ de degré $d \geqslant 0$ se décompose de manière unique sous la forme

$$F = a_d\, P_1^{m_1} P_2^{m_2} \cdots P_s^{m_s}, \quad P_j \in \mathcal{P} \ \text{ 2 à 2 distincts}, \ m_j \in \mathbb{N}^*,$$

où $\mathcal{P}$ désigne l'ensemble des polynômes irréductibles unitaires de $\mathbb{K}[X]$, et où $a_d \in \mathbb{K}^*$ est le coefficient dominant de F. L'ensemble des polynômes irréductibles dépend beaucoup du corps sur lequel on se place.

2.5.1. Anneau $\mathbb{C}[X]$. Dans ce cas, on sait que F se scinde en facteurs de degré 1, les éléments de $\mathcal{P}$ sont les $X - w$, $w \in \mathbb{C}$, et F se factorise sous la forme

$$F(X) = a_d(X - w_1)^{m_1}(X - w_2)^{m_2} \cdots (X - w_s)^{m_s}$$

où $w_1, \ldots, w_s \in \mathbb{C}$ sont les racines distinctes, et $m_1, \ldots, m_s$ leurs multiplicités. On a $d = \deg(P) = \sum_{1 \leqslant j \leqslant s} m_j$.

2.5.2. Anneau $\mathbb{R}[X]$. Soit

$$F(X) = a_d X^d + \cdots + a_1 X + a_0, \quad a_j \in \mathbb{R}, \quad a_d \neq 0.$$

Le polynôme F peut avoir des racines réelles r_j, mais il peut aussi avoir des racines complexes $w_j \in \mathbb{C} \smallsetminus \mathbb{R}$. Dans ce cas

$$F(\overline{w}_j) = a_d \overline{w}_j^d + \cdots + a_1 \overline{w}_j + a_0 = \overline{F(w_j)} = 0.$$

Ceci implique que les racines complexes non réelles $w_j \in \mathbb{C} \smallsetminus \mathbb{R}$ se regroupent par paires conjuguées $w_j, \overline{w}_j$. On peut aussi observer aussi que $w_j, \overline{w}_j$ ont les mêmes multiplicités, car les dérivées vérifient

$$F^{(\alpha)}(\overline{w}_j) = \overline{F^{(\alpha)}(w_j)} \quad \text{pour tout } \alpha \in \mathbb{N}.$$

Chaque produit $(X - w_j)(X - \overline{w}_j)$ s'écrit comme un trinôme du second degré $X^2 + \beta_j X + \gamma_j$ dont les coefficients $\beta_j = -w_j - \overline{w}_j = -2 \operatorname{Re} w_j$ et $\gamma_j = |w_j|^2$ sont réels. Le discriminant $\Delta = \beta_j^2 - 4\gamma_j$ est nécessairement < 0 (sinon les racines seraient réelles). Réciproquement, un tel trinôme du second degré de discriminant négatif est bien irréductible dans $\mathbb{R}[X]$. On en déduit que la décomposition de F en facteurs irréductibles est de la forme

$$F(X) = a_d \prod_{1 \leqslant j \leqslant s} (X - r_j)^{m_j} \prod_{1 \leqslant j \leqslant t} (X^2 + \beta_j X + \gamma_j)^{k_j}$$

où les r_j sont les racines réelles, et où les trinômes $X^2 + \beta_j X + \gamma_j$ sont des trinômes de discriminants $\Delta_j = \beta_j^2 - 4\gamma_j < 0$ admettant des racines complexes conjuguées $w_j, \overline{w}_j \in \mathbb{C} \smallsetminus \mathbb{R}$. On a ici $d = \deg(F) = \sum m_j + 2 \sum k_j$.

2.5.3. Exemple. Le polynôme $F(X) = X^4 + 1 \in \mathbb{R}[X]$ qui est de degré 4 sans racines réelles est nécessairement réductible dans $\mathbb{R}[X]$ (puisque les polynômes irréductibles y sont de degrés 1 et 2 seulement). On voit en fait que

$$\begin{aligned} X^4 + 1 &= (X^2+1)^2 - 2X^2 = (X^2+1)^2 - (\sqrt{2}\,X)^2 \\ &= (X^2 + \sqrt{2}\,X + 1)(X^2 - \sqrt{2}\,X + 1). \end{aligned}$$

Il s'agit de la décomposition en facteurs irréductibles, car les deux trinômes du second degré ont pour discriminant $\Delta = 2 - 4 = -2$. On en déduit ainsi qu'on a quatre racines complexes deux à deux conjuguées

$$w_1 = \frac{-\sqrt{2} + i\sqrt{2}}{2}, \quad \overline{w}_1 = \frac{-\sqrt{2} - i\sqrt{2}}{2}, \quad w_2 = \frac{\sqrt{2} + i\sqrt{2}}{2}, \quad \overline{w}_2 = \frac{\sqrt{2} - i\sqrt{2}}{2}.$$

2.5.4. Anneau $\mathbb{Q}[X]$. La situation est ici beaucoup plus compliquée, on montrera plus loin qu'il y a dans $\mathbb{Q}[X]$ des polynômes irréductibles de tous degrés, par exemple $X^d - p$ si p est un nombre premier. Nous nous contenterons de quelques résultats élémentaires, car la théorie générale nécessite des outils plus avancés dont nous ne disposons pas dans ce cours.

2.5.5. Proposition. *Un polynôme $F \in \mathbb{Q}[X]$ de degré 2 ou 3 est irréductible si et seulement si F n'admet pas de racine dans $\mathbb{Q}$.*

Démonstration. Ce fait a déjà été remarqué : si un tel polynôme est réductible dans $\mathbb{Q}[X]$, alors l'un des facteurs au moins est de degré 1 et possède donc une racine rationnelle. □

Nous donnons maintenant une méthode permettant de détecter les racines rationnelles. Remarquons que pour un polynôme $F \in \mathbb{Q}[X]$, on peut toujours multiplier les coefficients par un dénominateur commun de façon à se ramener à un polynôme $F \in \mathbb{Z}[X]$.

2.5.6. Proposition. *Soit $F \in \mathbb{Z}[X]$ un polynôme de degré d,*

$$F(X) = a_d X^d + a_{d-1} X^{d-1} + \cdots + a_1 X + a_0, \quad a_j \in \mathbb{Z}, \quad a_d \neq 0.$$

Supposons aussi $a_0 \neq 0$ (sinon $F(X)$ admet la racine $x = 0$ et on peut factoriser). Si $F(X)$ admet une racine rationnelle $x \in \mathbb{Q}^*$ écrite sous forme d'une fraction réduite $x = k/\ell$, c'est-à-dire telle que $\operatorname{pgcd}(k, \ell) = 1$, alors $k \mid a_0$ et $\ell \mid a_d$.

En pratique, on n'a donc qu'à chercher les diviseurs de a_0 et a_d, et cela ne laisse qu'un nombre fini de possibilités de racines $x \in \mathbb{Q}^*$ à tester. On notera en particulier que si le polynôme F est unitaire ($a_d = 1$), alors $\ell = \pm 1$, donc les racines rationnelles sont nécessairement dans $\mathbb{Z}$.

Démonstration. Si $x = k/\ell \in \mathbb{Q}^*$ est racine, alors

$$\ell^d F(k/\ell) = a_d k^d + a_{d-1} k^{d-1} \ell + \cdots + a_1 k \ell^{d-1} + a_0 \ell^d = 0.$$

Ceci donne

$$a_d k^d = -\ell\big(a_{d-1}k^{d-1} + \cdots + a_1 k\ell^{d-2} + a_0\ell^{d-1}\big),$$

donc $\ell \mid a_d k^d$. Comme $\operatorname{pgcd}(k,\ell) = 1$, le lemme de Gauss implique que $\ell \mid a_d$. De même l'égalité

$$a_0\ell^d = -k\big(a_d k^{d-1} + a_{d-1}k^{d-2}\ell + \cdots + a_1 k\ell^{d-1}\big)$$

implique $k \mid a_0$. □

2.5.7. Un exemple de degré 4. Le polynôme $X^4 + 1$ est irréductible dans l'anneau $\mathbb{Q}[X]$. En effet ce polynôme n'a pas de racine rationnelle (ni même réelle), la seule possibilité serait une décomposition en deux facteurs de degré 2 dans $\mathbb{Q}[X]$. Mais on a déjà vu qu'on avait dans $\mathbb{R}[X]$ une décomposition en deux facteurs irréductibles de degré 2, et c'est la seule décomposition possible (si on prend des polynômes unitaires). Il faudrait donc que cette décomposition soit à coefficients rationnels, ce qui n'est pas le cas, car on a le coefficient $\sqrt{2}$ qui apparaît dans les facteurs irréductibles de $X^4 + 1$ dans $\mathbb{R}[X]$.

2.5.8. Autre exemple. Le polynôme $X^4 + 4$ est en revanche *réductible* dans l'anneau $\mathbb{Q}[X]$, bien que n'ayant aucune racine rationnelle. En effet

$$\begin{aligned} X^4 + 4 &= (X^2+2)^2 - 4X^2 = (X^2+2)^2 - (2X)^2 \\ &= (X^2 + 2X + 2)(X^2 - 2X + 2). \end{aligned}$$

[Accessoirement, les racines complexes sont $z = \pm 1 \pm i$.]

2.5.9. Théorème. *Soit $p \in \mathbb{N}^*$ un nombre premier et $d \in \mathbb{N}^*$. Alors $X^d - p$ est irréductible dans $\mathbb{Q}[X]$.*

Démonstration. Les cas $d = 1, 2, 3$ sont triviaux (pour $d = 2, 3$, on applique 2.5.5). En général, pour $d \geqslant 2$, les racines complexes sont données par $z = p^{1/d}\zeta$ où $\zeta^d = 1$, c'est-à-dire $z = p^{1/d}e^{2\pi ik/d}$, $k = 0, 1, \ldots d-1$, et on a dans $\mathbb{C}[X]$ la factorisation

$$X^d - p = \prod_{k=0}^{p-1}(X - p^{1/d}e^{2\pi ik/d}).$$

Si $X^d - p$ avait une décomposition sous la forme $G(X)H(X)$ dans $\mathbb{Q}[X]$ (avec G, H qu'on peut supposer unitaires, et $0 < \deg(G) < d$, $0 < \deg(H) < d$), alors G (disons) serait un produit de certains des facteurs complexes $(X - p^{1/d}e^{2\pi ik_j/d})$, $1 \leqslant j \leqslant \delta = \deg(G)$. Mais alors le coefficient constant $c \in \mathbb{Q}$ de G serait égal à

$$c = \prod_{1\leqslant j\leqslant\delta}(-p^{1/d}e^{2\pi ik_j/d}) = \pm p^{\delta/d}e^{2\pi i\ell/d}, \quad 1 \leqslant \delta < d.$$

Ceci impliquerait $|c| = p^{\delta/d} = \sqrt[d]{p^\delta} \notin \mathbb{Q}$ puisque $d \nmid \delta$. Or $|c| = \pm c \in \mathbb{Q}$. Cette contradiction démontre l'irréductibilité de $X^d - p$ dans $\mathbb{Q}[X]$. □

2.5.10. Exercice. *Déterminer les racines réelles et la décomposition en facteurs irréductibles de $X^d - p$ dans $\mathbb{R}[X]$, suivant que d est pair ou impair. Redémontrer ainsi l'irréductibilité de $X^d - p$.*

2.5.11. Exercice.** En raffinant le raisonnement du théorème 2.5.9, montrer qu'un polynôme $X^d - a$ avec $a \in \mathbb{Q}_+^*$ est irréductible dans $\mathbb{Q}[X]$ si et seulement si pour tout diviseur d' de d tel que $1 < d' \leqslant d$, on a $\sqrt[d']{a} \notin \mathbb{Q}$ (ce qui revient à dire que si $p_1, \ldots, p_s$ sont les facteurs premiers intervenant dans a, on a $\mathrm{pgcd}(v_{p_1}(a), \ldots, v_{p_s}(a), d) = 1$).

2.6. Congruences et anneaux quotients

On introduit d'abord la notion élémentaire de congruence modulo un idéal.

2.6.1. Définition. *Soit A un anneau (commutatif) et I un idéal de A. On dit que des éléments $x, y \in A$ sont congrus modulo I, et on écrit $x \equiv y \mod I$ si $x - y \in I$.*

Il est clair qu'il s'agit d'une relation d'équivalence. En effet on a :

– réflexivité: $x \equiv x \mod I$, puisque $x - x = 0 \in I$;

– symétrie: $x \equiv y \mod I \Leftrightarrow y \equiv x \mod I$, car $\delta = x - y \in I \Leftrightarrow -\delta = y - x \in I$;

– transitivité : $x \equiv y$ et $y \equiv z \mod I \Rightarrow x \equiv z \mod I$, car $x - z = (x-y)+(y-z) \in I$.

Étant donné $x \in A$, on note $\dot{x}$ la classe d'équivalence pour la relation de congruence modulo I. Comme $y \equiv x \mod I$ équivaut à $y - x = t \in I$, il s'agit simplement de

$$\dot{x} = x + I = \{x + t \,/\, t \in I\}. \tag{2.6.2}$$

Maintenant étant donné deux classes $\dot{x}$ et $\dot{y}$ et $x' = x + s \in \dot{x}$, $y' = y + t \in \dot{y}$, $s, t \in I$, on observe que

$$x' + y' = x + y + (s + t) \quad \text{avec } s + t \in I,$$

$$x'y' = (x + s)(y + t) = xy + (xt + ys + st) \quad \text{avec } xt + ys + st \in I,$$

par conséquent $x' + y' \equiv x + y$ et $x'y' \equiv xy \mod I$. Ceci montre qu'il est possible de poser par définition

$$\dot{x} + \dot{y} := (x + y)^{\bullet} = (x' + y')^{\bullet}, \tag{2.6.3}$$

$$\dot{x}\,\dot{y} := (x\,y)^{\bullet} = (x'\,y')^{\bullet}, \tag{2.6.3$'$}$$

puisque les classes définies par les membres de droite ne dépendent pas des représentants x, y ou x', y' choisis.

2.6.4. Théorème et définition. *L'ensemble des classes d'équivalence $\dot{x}$ des éléments $x \in A$ pour la relation de congruence modulo I se note A/I. Avec les lois définies par (2.6.3) et (2.6.3$'$), on obtient une structure d'anneau commutatif $(A/I, +, \times)$, appelé* anneau quotient *de A par I. De plus, l'application naturelle*

$$\pi_{A,I} : A \to A/I, \quad x \mapsto \dot{x}$$

est un morphisme surjectif d'anneaux.

Démonstration. Les explications données plus haut montre que les lois de composition interne $+$ et $\times$ sont bien définies sur A/I. Tous les axiomes des anneaux (associativité, distributivité, ...) vraies dans A passent "automatiquement" à A/I. La classe $\dot{0}_A$ est élément neutre pour $+$, et la classe $\dot{1}_A$ est élément neutre pour $\times$ dans A/I. La propriété de morphisme est évidente par définition, et la surjectivité de $\pi_{A,I}$ aussi. □

2.6.5. Exemples. (a) On a deux cas assez peu intéressants, à savoir $I = \{0\}$, où chaque classe $\dot{x}$ se réduit au singleton $\{x\}$, de sorte que $A/\{0\}$ peut s'identifier à A lui-même, et le cas $I = A$ où on a $\dot{x} = A = \dot{0}$ pour tout $x \in A$, de sorte que $A/A = \{\dot{0}\}$ est l'anneau trivial (avec $\dot{0} = \dot{1}$).

(b) Dans le cas où $I = \langle g \rangle = gA$ est un idéal principal, la classe d'équivalence $\dot{x}$ consiste en les éléments

$$\dot{x} = \{x + \lambda g \,/\, \lambda \in A\}$$

et l'anneau quotient est souvent noté A/gA. Pour $n \in \mathbb{N}^*$, l'anneau $\mathbb{Z}/n\mathbb{Z}$ est ainsi constitué de n éléments, à savoir les classes

$$\dot{0},\ \dot{1},\ \ldots\ ,(n-1)^{\bullet}$$

où la classe $\dot{x} = \dot{r}$ est obtenu en calculant le reste r de la division de x par n, tel que $x = nq + r$. (On a encore ici les cas "inintéressants" $\mathbb{Z}/0\,\mathbb{Z} = \mathbb{Z}/\{0\} \simeq \mathbb{Z}$ et $\mathbb{Z}/1\,\mathbb{Z} = \mathbb{Z}/\mathbb{Z} = \{\dot{0}\}$). L'anneau $\mathbb{Z}/6\mathbb{Z}$ est bien celui défini par les tables de Pythagore du 1.1.4 (c), et l'anneau $\mathbb{Z}/2\mathbb{Z}$ est la même chose que le corps $\mathbb{F}_2$ déjà mentionné à plusieurs reprises.

2.6.6. Application : "preuves" par 9 et 11. Il s'agit de techniques utilisées autrefois – et peut-être encore aujourd'hui ? – par les écoliers des classes primaires pour vérifier leurs opérations arithmétiques. Bien entendu, les maîtres donnaient en général seulement la recette sans trop en expliquer les raisons ...

(a) *"Preuve" par 9.* On travaille dans l'anneau quotient $\mathbb{Z}/9\mathbb{Z}$. On remarque que l'on a $10 \equiv 1 \bmod 9$ et donc $10^k \equiv 1 \bmod 9$ pour tout $k \in \mathbb{N}$. Par conséquent, si $x \in \mathbb{N}$ est un entier écrit en base 10, on voit que

$$x = a_k\,10^k + a_{k-1}\,10^{k-1} + \cdots + a_1\,10 + a_0 \equiv a_k + a_{k-1} + \cdots + a_1 + a_0 \mod 9,$$

autrement dit x est congru à la somme de ses chiffres modulo 9. Par exemple 4570891 est congru à $4+5+7+8+9+1$, et donc à $7+8+1 = 16$ (puisque $9 \equiv 0 \bmod 9$), 16 étant lui-même congru à $1+6 = 7$. Par suite $4570891 \equiv 7 \bmod 9$. Pour "vérifier" le résultat z d'une multiplication xy, on calcule comme ci-dessus les classes $\dot{x}, \dot{y}, \dot{z}$, et on s'assure que $\dot{x}\,\dot{y} = \dot{z}$. Par exemple, pour vérifier que $23 \times 47 = 1081$ on fait $2+3 = 5$, $4+7 = 11 \equiv 2$, $1+0+8+1 \equiv 1$ et on constate qu'on a bien $5 \times 2 \equiv 1 \bmod 9$. Mais si on avait trouvé $z = 1171$, on ne se serait quand même pas aperçu de l'erreur, la "preuve" par 9 est loin d'être infaillible !

(b) *"Preuve" par 11.* L'idée est la même, on travaille cette fois dans l'anneau quotient $\mathbb{Z}/11\mathbb{Z}$. On a $10 \equiv -1 \bmod 11$ et donc $10^k \equiv (-1)^k \bmod 11$, d'où

$$\begin{aligned} x &= a_k\, 10^k + a_{k-1}\, 10^{k-1} + \cdots + a_1\, 10 + a_0 \\ &\equiv (-1)^k a_k + (-1)^{k-1} a_{k-1} + \cdots - a_1 + a_0 \mod 11, \end{aligned}$$

c'est-à-dire que x est congru à la somme alternée de ses chiffres modulo 11. La "preuve" par 10 est beaucoup plus simple, puisqu'elle dépend seulement du dernier chiffre a_0. On peut aussi faire les "preuves" par 7 ou 13 (ou n'importe quel autre entier n), mais c'est plus compliqué ...

2.6.7. Théorème. *Si A est un anneau principal et $I = gA$ est l'idéal engendré par un élément $g \in A^*$, alors il y a équivalence entre les propriétés suivantes :*

(i) *g est un élément irréductible de A ;*

(ii) *A/gA est intègre non trivial ;*

(iii) *A/gA est un corps.*

Démonstration. Il est évident que (iii) $\Rightarrow$ (ii). D'autre part, si g est décomposable sous la forme $g = xy$ avec $x, y \in A^*$ non inversibles, alors g ne peut diviser x ou y (sinon y, resp. x, serait inversible d'inverse x/g, resp. y/g), donc $\dot{x} \neq \dot{0}$, $\dot{y} \neq \dot{0}$ et

$$\dot{x}\,\dot{y} = \dot{g} = \dot{0} \quad \text{dans } A/gA,$$

de sorte que A/gA est non trivial et possède des diviseurs de zéro. Ceci montre par contraposition que (ii) $\Rightarrow$ (i). Il reste à voir que (i) $\Rightarrow$ (iii). Supposons maintenant g irréductible et soit $\dot{x} \neq \dot{0}$ dans A/gA. Ceci signifie que $g \nmid x$, et d'après le lemme de Gauss 2.4.2 (c), on en déduit que $\operatorname{pgcd}(x, g) = 1$. D'après l'identité de Bézout, il existe alors $\lambda, \mu \in A$ tels que $\lambda x + \mu g = 1$, ce qui donne $\dot{\lambda}\,\dot{x} = \dot{1}$, et on voit que A/gA est bien un corps. □

2.6.8. Corollaire. *L'anneau quotient $\mathbb{Z}/n\mathbb{Z}$ est un corps si et seulement n est un nombre premier. On note $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$; c'est un corps à p éléments, constitué des classes $\dot{0}, \dot{1}, \ldots, (p-1)^{\bullet}$.*

2.6.9. Notion de caractéristique d'un corps. Soit $\mathbb{K}$ un corps, et

$$\varphi_{\text{can}} : \mathbb{Z} \to \mathbb{K}, \quad n \mapsto n_{\mathbb{K}}$$

le morphisme canonique. Le noyau $\operatorname{Ker}(\varphi_{\text{can}}) = \{0\}$ est un idéal de $\mathbb{Z}$, car $\varphi_{\text{can}}(x) = \varphi_{\text{can}}(y) = 0_{\mathbb{K}}$ implique trivialement

$$\varphi_{\text{can}}(\lambda x + \mu y) = \varphi_{\text{can}}(\lambda)\varphi_{\text{can}}(x) + \varphi_{\text{can}}(\mu)\varphi_{\text{can}}(y) = 0_{\mathbb{K}}.$$

Comme $\mathbb{Z}$ est un anneau principal, on a $\operatorname{Ker}(\varphi_{\text{can}}) = p\mathbb{Z}$ pour un certain entier $p \in \mathbb{N}$. On dit alors que $\mathbb{K}$ est un corps de *caractéristique* p. Deux alternatives exclusives l'une de l'autre peuvent se produire.

(a) $p = 0$. Dans ce cas $\mathrm{Ker}(\varphi_{\mathrm{can}}) = \{0\}$, i.e. φ_{can} est injectif, et on peut définir un morphisme canonique de corps en posant

$$\psi_{\mathrm{can}} : \mathbb{Q} \to \mathbb{K}, \quad \frac{a}{b} \mapsto (a_{\mathbb{K}})(b_{\mathbb{K}})^{-1}, \quad a \in \mathbb{Z}, \quad b \in \mathbb{N}^*$$

la vérification facile en est laissée au lecteur. Comme $\psi_{\mathrm{can}}(x) \neq 0$ pour tout $x = a/b \in \mathbb{Q}^*$, on a aussi $\mathrm{Ker}(\psi_{\mathrm{can}}) = \{0\}$, c'est-à-dire que ψ_{can} est injectif et définit un isomorphisme de $\mathbb{Q}$ sur son image $\psi_{\mathrm{can}}(\mathbb{Q}) \subset \mathbb{K}$ (qui est un sous-corps).

(b) $p \in \mathbb{N}^*$, de sorte que $p_K = 0_K$. Comme $1_{\mathbb{K}} \neq 0_{\mathbb{K}}$, on a nécessairement $p > 1$, et d'autre part p doit être un *nombre premier*, sinon on aurait $p = xy$ avec $1 < x, y < p$ et donc $x_{\mathbb{K}}\, y_{\mathbb{K}} = 0_{\mathbb{K}}$ avec $x_{\mathbb{K}}, y_{\mathbb{K}} \neq 0$, ce qui contredirait le fait que $\mathbb{K}$ est un corps. Comme $\varphi_{\mathrm{can}}(x + \lambda p) = \varphi_{\mathrm{can}}(x) = x_{\mathbb{K}}$, on voit que φ_{can} définit par passage au quotient un morphisme de corps

$$\psi_{\mathrm{can}} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{K}, \quad \dot{x} \mapsto x_{\mathbb{K}} \quad (p \text{ premier}),$$

dont le noyau est cette fois réduit à $\{\dot{0}\}$. Par conséquent ψ_{can} est injectif, et définit un isomorphisme de $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ sur $\psi_{\mathrm{can}}(\mathbb{Z}/p\mathbb{Z}) \subset \mathbb{K}$ (qui est un sous-corps).

2.6.10. Théorème. *Soit $\mathbb{K}$ un corps fini. Alors $\mathbb{K}$ est de caractéristique $p > 0$ et $q = \mathrm{card}\,\mathbb{K}$ est une puissance de p, soit $q = p^n$. De plus, pour tout $\alpha \in \mathbb{K}^*$, on a $\alpha^{q-1} = 1_{\mathbb{K}}$, et le polynôme $X^q - X \in \mathbb{K}[X]$ admet la factorisation*

$$X^q - X = \prod_{\alpha \in \mathbb{K}} (X - \alpha).$$

Démonstration. La caractéristique de $\mathbb{K}$ est nécessairement positive, sinon $\mathbb{K}$ contiendrait un sous-corps infini isomorphe à $\mathbb{Q}$. Donc $\mathbb{K}$ contient un sous-corps $\mathbb{K}'$ isomorphe à $\mathbb{F}_p$, et les lois $+$, $\times$ munissent $\mathbb{K}$ d'une structure d'espace vectoriel de dimension finie sur $\mathbb{K}' \simeq \mathbb{F}_p$. Si n est sa dimension, on a alors $\mathbb{K} \simeq (\mathbb{F}_p)^n$, donc $q = \mathrm{card}\,\mathbb{K} = p^n$. Si $\alpha \in \mathbb{K}^*$, l'application $\mathbb{K}^* \to \mathbb{K}^*$, $x \mapsto \alpha x$ est une bijection, dont la bijection inverse est $x \mapsto \alpha^{-1}x$. Comme $\mathrm{card}\,\mathbb{K}^* = q - 1$, le produit (non nul) de tous les éléments de $\mathbb{K}^*$ fournit

$$\prod_{x \in \mathbb{K}^*} x = \prod_{x \in \mathbb{K}^*} \alpha x = \alpha^{q-1} \prod_{x \in \mathbb{K}^*} x.$$

Ceci implique bien que $\alpha^{q-1} = 1$. Comme le polynôme $X^{q-1} - 1$ est de degré $q-1$ et admet les $q - 1$ éléments de $\mathbb{K}^*$ comme racines, on en déduit que ces racines sont simples et qu'on a la factorisation

$$X^{q-1} - 1 = \prod_{\alpha \in \mathbb{K}^*} (X - \alpha).$$

En multipliant par X, on obtient la factorisation annoncée de $X^q - X$. □

2.6.11. Corollaire : petit théorème de Fermat. *Soit p un nombre premier. Alors*

(a) *pour tout $x \in \mathbb{Z}$ non divisible par p, on a $x^{p-1} \equiv 1 \ mod\ p$;*

(b) *pour tout $x \in \mathbb{Z}$, on a $x^p \equiv x \ mod\ p$.*

(c) *dans $\mathbb{F}_p[X]$, on a la factorisation*

$$X^p - X = \prod_{\alpha \in \mathbb{F}_p} (X - \alpha).$$

□

2.7. Théorème des restes chinois

Le problème est de résoudre des congruences simultanées

$$x \equiv a_1 \mod \langle n_1 \rangle, \quad x \equiv a_2 \mod \langle n_2 \rangle, \quad \ldots \quad , \quad x \equiv a_s \mod \langle n_s \rangle,$$

dans $\mathbb{Z}$, ou plus généralement dans un *anneau principal* A. Les références historiques attestent que les mathématiciens chinois se posaient déjà ce genre des questions au début de notre ère, probablement pour traiter des problèmes astronomiques ou calendaires ; les documents trouvés remontent au moins au III$^{\text{e}}$ siècle après J.C. (Sunzi Suanjing, "The mathematical classic of Master Sun"). Il semble que les Grecs aient aussi étudié cette question, peut-être même antérieurement aux Chinois. On commence par traiter le cas plus simple $s = 2$.

2.7.1. Théorème des restes chinois. *On se donne $n_1, n_2 \neq 0$ dans un anneau principal A. Si* $\operatorname{pgcd}(n_1, n_2) = 1$, *les congruences simultanées*

$$x \equiv a_1 \mod \langle n_1 \rangle, \quad x \equiv a_2 \mod \langle n_2 \rangle$$

se résolvent comme l'ensemble des $x \in A$ satisfaisant une unique congruence

$$x \equiv x_0 \mod \langle n_1 n_2 \rangle,$$

où x_0 dépend des données a_1, a_2 (et n_1, n_2).

Démonstration. Démontrons d'abord l'existence d'une solution x_0. D'après l'identité de Bézout, on peut trouver $\lambda_1, \lambda_2 \in A$ tels que $\lambda_1 n_1 + \lambda_2 n_2 = 1$ (et si A est euclidien, on peut utiliser l'algorithme d'Euclide pour trouver λ_1, λ_2). Cette relation implique de façon évidente que

$$\begin{aligned} &\lambda_2 n_2 \equiv 1 \mod \langle n_1 \rangle, \quad \lambda_2 n_2 \equiv 0 \mod \langle n_2 \rangle, \\ &\lambda_1 n_1 \equiv 0 \mod \langle n_1 \rangle, \quad \lambda_1 n_1 \equiv 1 \mod \langle n_2 \rangle. \end{aligned}$$

Par conséquent, si on prend la combinaison linéaire

$$x_0 = a_1(\lambda_2 n_2) + a_2(\lambda_1 n_1),$$

on trouve bien

$$x_0 \equiv a_1 \mod \langle n_1 \rangle, \quad x_0 \equiv a_2 \mod \langle n_2 \rangle.$$

Maintenant, si $x \in A$ est une autre solution, on trouve par différence

$$x - x_0 \equiv 0 \mod \langle n_1 \rangle, \quad x - x_0 \equiv 0 \mod \langle n_2 \rangle,$$

c'est-à-dire que $x - x_0$ doit être multiple de n_1 et n_2. Il doit donc être multiple de $\mathrm{ppcm}(n_1, n_2) = n_1 n_2$ (du fait que $\mathrm{pgcd}(n_1, n_2) = 1$, cf. 2.4.10 (c)). Ceci implique $x \equiv x_0 \mod \langle n_1 n_2 \rangle$, et réciproquement de tels éléments x sont bien des solutions du problème, puisque $x \equiv x_0 \mod \langle n_1 n_2 \rangle$ implique $x \equiv x_0 \equiv a_1 \mod \langle n_1 \rangle$ et $x \equiv x_0 \equiv a_2 \mod \langle n_2 \rangle$. □

2.7.2. Formulation "moderne". Une façon beaucoup plus moderne (20$^{\mathrm{e}}$ siècle), et également plus abstraite, de formuler le théorème des restes chinois est d'écrire que si $\mathrm{pgcd}(n_1, n_2) = 1$, on a un isomorphisme d'anneaux

$$\varphi : A/n_1 n_2 A \to A/n_1 A \times A/n_2 A, \quad \overset{\bullet}{x} \mod \langle n_1 n_2 \rangle \mapsto (\overset{\bullet}{x} \mod \langle n_1 \rangle, \overset{\bullet}{x} \mod \langle n_2 \rangle)$$

[Si A_1 et A_2 sont des anneaux, on munit $A_1 \times A_2$ d'une structure d'anneau en posant $(x, y) + (x', y') = (x + x', y + y')$ et $(x, y)(x', y') = (xx', yy')$; on a $1_{A_1 \times A_2} = (1_{A_1}, 1_{A_2})$]. Il est ici évident que φ est un morphisme, et c'est sa bijectivité qui est équivalente au théorème des restes chinois (la surjectivité exprimant l'existence de la solution x_0, et l'injectivité son unicité modulo $\langle n_1 n_2 \rangle$). La preuve du théorème fournit en outre une formule explicite pour l'isomorphisme inverse φ^{-1} : si $\lambda_1 n_1 + \lambda_2 n_2 = 1$, cet inverse est donné par

$$\begin{aligned} \varphi^{-1} : \quad & A/n_1 A \times A/n_2 A \quad \longrightarrow \quad A/n_1 n_2 A, \\ & (\overset{\bullet}{a_1} \mod \langle n_1 \rangle, \overset{\bullet}{a_2} \mod \langle n_2 \rangle) \longmapsto \big(a_1(\lambda_2 n_2) + a_2(\lambda_1 n_1)\big)^{\bullet} \mod \langle n_1 n_2 \rangle. \end{aligned}$$

2.7.3. Exemples. (a) Dans $\mathbb{Z}$, les conguences simultanées

$$x \equiv 2 \mod \langle 4 \rangle, \quad x \equiv 3 \mod \langle 6 \rangle$$

sont incompatibles. En effet la première implique x pair et la seconde x impair. Mais cela ne contredit pas le théorème des restes chinois puisque $\mathrm{pgcd}(4, 6) = 2 \neq 1$.

(b) Soit à résoudre dans $\mathbb{Q}[X]$ les congruences simultanées

$$P \equiv X \mod \langle X^3 - 1 \rangle, \quad P \equiv -X^2 \mod \langle X^5 + 1 \rangle.$$

On commence par calculer $\mathrm{pgcd}(X^3 - 1, X^5 + 1)$ par l'algorithme d'Euclide :

$$\begin{aligned} X^5 + 1 &= (X^3 - 1)X^2 + (X^2 + 1), \\ X^3 - 1 &= (X^2 + 1)X - (X + 1), \\ X^2 + 1 &= (X + 1)(X - 1) + 2, \\ X + 1 &= 2\left(\frac{1}{2}X + \frac{1}{2}\right) + 0, \end{aligned}$$

ce qui implique bien $\operatorname{pgcd}(X^5+1, X^3-1) = 1$ (à l'élément inversible $2 \in \mathbb{Q}^*$ près ; dans le calcul ci-dessus, on a aussi remplacé $-(X+1)$ par $X+1$, -1 étant inversible). On voit que

$$\begin{aligned} 2 &= (X^2+1) - (X+1)(X-1) = (X^2+1) + \big((X^3-1) - (X^2+1)X\big)(X-1) \\ &= (-X^2+X+1)(X^2+1) + (X-1)(X^3-1) \\ &= (-X^2+X+1)\big((X^5+1) - (X^3-1)X^2\big) + (X-1)(X^3-1) \\ &= (-X^2+X+1)(X^5+1) + (X^4-X^3-X^2+X-1)(X^3-1). \end{aligned}$$

(et on doit diviser par 2 pour obtenir la constante 1). D'après la formule vue dans la démonstration du théorème 2.7.1, une solution du problème est alors

$$\begin{aligned} P_0 &= \frac{1}{2}\Big(X(-X^2+X+1)(X^5+1) + (-X^2)(X^4-X^3-X^2+X-1)(X^3-1)\Big) \\ &= \frac{1}{2}\Big(-X^9+2\,X^7+X^6-X^4+X\Big) \equiv X^7 \mod \langle (X^3-1)(X^5+1)\rangle, \end{aligned}$$

et la solution générale est donc

$$P \equiv X^7 \mod \langle (X^3-1)(X^5+1)\rangle.$$

On peut aisément le vérifier a posteriori :

$$X^7 = X(X^3+1)(X^3-1) + X, \quad X^7 = X^2(X^5+1) - X^2.$$

2.7.4. Généralisation. *On se donne $n_1, \dots, n_s \neq 0$ dans un anneau principal A. Si $\operatorname{pgcd}(n_i, n_j) = 1$ pour tous $i \neq j$, les congruences simultanées*

$$x \equiv a_1 \mod \langle n_1\rangle, \quad x \equiv a_2 \mod \langle n_2\rangle, \quad \dots \quad , \quad x \equiv a_s \mod \langle n_s\rangle$$

se résolvent comme l'ensemble des $x \in A$ satisfaisant une unique congruence

$$x \equiv x_0 \mod \langle n_1 n_2 \cdots n_s\rangle,$$

où x_0 dépend des données $a_1, \dots, a_s$ (et $n_1, \dots, n_s$). En d'autres termes, on a un isomorphisme d'anneaux

$$\begin{aligned} &\varphi : A/(n_1\cdots n_s)A \longrightarrow A/n_1A \times \cdots \times A/n_sA, \\ &\dot{x} \mod \langle n_1\cdots n_s\rangle \longmapsto (\dot{x} \mod \langle n_j\rangle)_{1\leqslant j\leqslant s}. \end{aligned}$$

Démonstration. On raisonne par récurrence sur $s \geqslant 3$, le problème ayant déjà été traité pour $s = 2$. Les deux dernières congruences $x \equiv a_{s-1} \mod \langle n_{s-1}\rangle$, $x \equiv a_s \mod \langle n_s\rangle$ équivalent à une seule congruence $x \equiv a'_{s-1} \mod \langle n_{s-1}n_s\rangle$ d'après le cas $s = 2$, et le lemme de Gauss implique que l'on a bien $\operatorname{pgcd}(n_i, n_{s-1}n_s) = 1$ pour $i < s-1$. On peut donc appliquer l'hypothèse de récurrence pour $s-1$ congruences, respectivement modulo $n_1, n_2, \dots, n_{s-2}, n'_{s-1} = n_{s-1}n_s$, et conclure alors que la solution s'exprime sous forme d'une unique congruence $x \equiv x_0 \mod \langle n_1\cdots n_{s-2}(n_{s-1}n_s)\rangle$. □

2.8. Corps de décomposition*

Nous démontrons ici quelques résultats importants en théorie des corps, qui peuvent être obtenus comme des conséquences assez directes du théorème 2.6.7 dans le cas des anneaux de polynômes. Cette section (quoiqu'en principe accessible sans trop de difficultés au moyen des résultats qui précèdent) est hors programme, et réservée aux étudiants souhaitant étendre un peu leur culture mathématique.

2.8.1. Définition. *Étant donné un corps $\mathbb{K}$, on appelle extension de $\mathbb{K}$ tout corps $\mathbb{L}$ qui contient $\mathbb{K}$ comme sous-corps, et on écrira $\mathbb{L} \supset \mathbb{K}$ pour indiquer cette situation.*

Un exemple important bien connu est $\mathbb{C} \supset \mathbb{R}$.

2.8.2. Théorème. *Soit P un polynôme irréductible de degré d d'un anneau $\mathbb{K}[X]$. Alors l'anneau quotient $\mathbb{L} = \mathbb{K}[X]/\langle P\rangle$ est un corps. De plus $\mathbb{K}$ s'identifie à un sous-corps de $\mathbb{L}$, via le morphisme injectif $\mathbb{K} \hookrightarrow \mathbb{L}$, $c \mapsto \dot{c}$. Le corps $\mathbb{L}$ peut ainsi être considéré comme une extension de $\mathbb{K}$. C'est aussi un $\mathbb{K}$-espace vectoriel de dimension $\dim_{\mathbb{K}} \mathbb{L} = d$, admettant comme base*

$$(\dot{1}, \dot{X}, \dots, \dot{X}^{d-1}).$$

Démonstration. Comme $\mathbb{K}[X]$ est principal, le fait que $\mathbb{L}$ soit un corps résulte du théorème 2.6.7. L'application $\varphi : \mathbb{K} \to \mathbb{L}$, $c \mapsto \dot{c}$ est bien injective car $\mathrm{Ker}(\varphi) = \{0\}$: pour tout $c \in \mathbb{K}^*$, $\dot{c} \neq \dot{0}$, car c ne peut être multiple de P qui est de degré $d \geqslant 1$. Supposons P unitaire (ce n'est pas restrictif), et écrivons

$$P(X) = X^d + a_{d-1}X^{d-1} + \cdots + a_1 X + a_0, \quad d \geqslant 1.$$

En prenant les classes modulo $\langle P\rangle$, on en déduit

$$\dot{X}^d = -\dot{a}_0\dot{1} - \dot{a}_1\dot{X} - \cdots - \dot{a}_{d-1}\dot{X}^{d-1},$$

et en multipliant par $\dot{X}^{n-d}$ pour $n \geqslant d$, on voit que

$$\dot{X}^n = -\dot{a}_0\dot{X}^{n-d} - \dot{a}_1\dot{X}^{n-d+1} - \cdots - \dot{a}_{d-1}\dot{X}^{n-1}, \quad \forall n \geqslant d.$$

Ceci entraîne que modulo P, la classe $\dot{G}$ de tout polynôme $G = \sum_{j=0}^{n} b_j X^j \in \mathbb{K}[X]$ est égale à une combinaison linéaire des éléments de la famille

$$(\dot{1}, \dot{X}, \dots, \dot{X}^{d-1}),$$

de sorte que celle-ci est une famille génératrice de $\mathbb{L}$, vu comme espace vectoriel sur $\mathbb{K}$. Mais si la famille était liée, il existerait des coefficients $\lambda_0, \lambda_1, \dots, \lambda_{d-1} \in \mathbb{K}$ non tous nuls tels que

$$\dot{\lambda}_0\dot{1} + \dot{\lambda}_1\dot{X} + \cdots + \dot{\lambda}_{d-1}\dot{X}^{d-1} = \dot{0},$$

ce qui signifie précisément que $\lambda_0 + \lambda_1 X + \cdots + \lambda_{d-1} X^{d-1}$ est divisible par P. Ceci est impossible puisque P est de degré d. Par conséquent la famille considérée est bien une base, et le théorème est démontré. (On remarquera que le cas $d = 1$ n'est pas très intéressant, on a alors $\mathbb{L} \simeq \mathbb{K}$). □

2.8.3. Exemple. Le polynôme $P(X) = X^2 + 1 \in \mathbb{R}[X]$ est irréductible, donc $\mathbb{R}[X]/\langle X^2+1\rangle$ est un corps. Le théorème ci-dessus montre qu'il admet pour base $(\dot{1}, \dot{X})$ sur $\mathbb{R}$ et qu'on a la relation $(\dot{X})^2 = -\dot{1}$. C'est précisément le corps des complexes ! En fait cette méthode est probablement la méthode la plus fondamentale qui existe pour définir $\mathbb{C}$, fournissant un procédé algébrique naturel pour ajouter une racine manquante à l'équation $X^2 + 1 = 0$.

Le procédé précédent s'applique de manière générale à tout corps $\mathbb{K}$ et tout polynôme $G \in \mathbb{K}[X]$. De façon précise, on va montrer :

2.8.4. Théorème. *Pour tout corps $\mathbb{K}$ et tout polynôme $G \in \mathbb{K}[X]$, il existe une extension $\mathbb{L} \supset \mathbb{K}$ dans laquelle le polynôme G est complètement scindé, à savoir qu'on peut trouver des éléments $w_j \in \mathbb{L}$ tels que*

$$G(X) = a_d \prod (X - w_j)^{m_j} \quad \text{dans } \mathbb{L}[X] \supset \mathbb{K}[X],$$

où a_d est le coefficient dominant de G.

Démonstration. On démontre le résultat par récurrence sur $d = \deg(G)$. Si $d = 1$, le polynôme P est déjà scindé, on prend $\mathbb{L} = \mathbb{K}$ et il n'y a rien à montrer. Supposons le résultat déjà démontré pour $d-1$ et prenons $G \in \mathbb{K}[X]$ de degré $\deg(G) = d$. On commence par le factoriser en ses facteurs irréductibles, soit

$$G = a_d P_1^{k_1} \dots P_s^{k_s}.$$

On pose $\mathbb{L}_1 = \mathbb{K}[Y]/\langle P_1(Y)\rangle$. Ce quotient est une extension $\mathbb{L}_1 \supset \mathbb{K}$ d'après le théorème précédent. Par construction P_1 admet la racine $w_1 = \dot{Y}$ dans $\mathbb{L}_1$, de sorte que $w_1 \in \mathbb{L}_1$ est aussi une racine de $G \in \mathbb{K}[X] \subset \mathbb{L}_1[X]$. Par conséquent, on peut écrire $G(X) = (X - w_1)H(X)$ avec $H \in \mathbb{L}_1[X]$ de degré $d-1$. L'hypothèse de récurrence entraîne qu'il existe une extension $\mathbb{L} \supset \mathbb{L}_1 \supset \mathbb{K}$ dans laquelle H est complètement scindé, et donc G aussi. Le théorème s'ensuit. □

2.8.5. Remarque. la démonstration précédente construit en fait la plus petite extension $\mathbb{L} \supset \mathbb{K}$ dans laquelle $G \in \mathbb{K}[X]$ puisse se scinder. Un tel corps $\mathbb{L}$ est appelé *corps de décomposition* de G. On peut montrer qu'il est unique à isomorphisme de corps près.

2.8.6. Exercice.**

(a) Soit $\mathbb{K}$ un corps de caractéristique $p > 0$. Montrer par la formule du binôme que $F : x \mapsto x^p$ est un morphisme de corps ("morphisme de Frobenius"). Lorsque $\mathbb{K} = \mathbb{F}_p$, montrer que $F = \mathrm{Id}$ (calculer $F(1), F(2) = F(1+1), \dots$) et retrouver ainsi le petit théorème de Fermat.

(b) Soit de nouveau $\mathbb{K}$ un corps de caractéristique $p > 0$. Déduire de (a) que l'application $F^n = F \circ \cdots \circ F : x \mapsto x^{p^n}$ est aussi un morphisme de corps, puis que $\mathbb{K}_n = \{x \in \mathbb{K} \,/\, F^n(x) = x\}$ est un sous-corps de $\mathbb{K}$ possédant au plus p^n éléments.

(c) Montrer que le corps de décomposition $\mathbb{L}$ du polynôme $G = X^{p^n} - X \in \mathbb{F}_p[X]$ coïncide avec $\mathbb{L}_n$ (qui n'est autre que l'ensemble des racines du polynôme G), et que les racines sont simples [raisonner sur la dérivée]. En déduire que $\mathbb{L}$ est un corps ayant exactement $q = p^n$ éléments.

Ce corps est traditionnellement noté $\mathbb{F}_q$.

Nota : le théorème 2.6.10 entraîne assez aisément que tout corps $\mathbb{K}$ à $q = p^n$ éléments est isomorphe à $\mathbb{F}_q$; on a ainsi déterminé tous les corps finis commutatifs ; mais un théorème célèbre démontré en 1905 par le mathématicien écossais Joseph Wedderburn (1882-1948) énonce qu'il n'y a pas de corps finis non commutatifs. Les corps finis sont très importants en pratique, ils sont par exemple utilisés pour les codes correcteurs d'erreurs en informatique et en téléphonie mobile, tels que le code de Reed-Solomon s'appuyant sur le corps $\mathbb{F}_{256}$.

Chapitre 5
Réduction des endomorphismes

L'objet de ce chapitre est d'analyser la structure géométrique des endomorphismes en cherchant des bases de l'espace qui soient susceptibles de donner lieu à des matrices aussi simples que possible. La propriété cruciale est celle de sous-espace stable, et en particulier de droite stable. Ceci conduit aux notions importantes de vecteurs propres et de valeurs propres, lesquelles s'obtiennent au moyen de calculs de déterminants. Un endomorphisme est ainsi "diagonalisable" si et seulement si on peut trouver une base de l'espace formée de vecteurs propres. Dans les autres cas, la structure de l'endomorphisme est plus compliquée, et on doit par exemple étudier les puissances successives de l'endomorphismes et leurs combinaisons linéaires. Une idée importante est d'exploiter les propriétés algébriques des anneaux de polynômes vues au chapitre précédent, en appliquant le calcul polynomial – et notamment l'identité de Bézout – aux endomorphismes à étudier. On peut ainsi montrer que la condition nécessaire et suffisante pour qu'un endomorphisme soit "triangularisable" est que le polynôme caractéristique soit scindé sur le corps de référence, ce qui est toujours le cas pour le corps des complexes. Sur $\mathbb{C}$, on peut toujours trouver une base dans laquelle la matrice de l'endomorphisme est sous "forme réduite de Jordan", du nom du mathématicien français Camille Jordan (1838–1922).

1. Valeurs propres et vecteurs propres

1.1. Polynôme caractéristique et espaces propres

Soit E un espace vectoriel de dimension finie n sur le corps $\mathbb{K}$ (les corps seront toujours supposés commutatifs dans ce chapitre). Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme, c'est à dire une application $\mathbb{K}$-linéaire $f : E \to E$.

Rappelons qu'un sous-espace S de E est dit stable par f si $f(S) \subset S$. D'une certaine manière, le problème de la réduction des endomorphismes consiste essentellement en la détermination des sous-espaces stables. Le cas le plus fondamental est la recherche des droites D de E qui sont stables par f.

1.1.1. Observation de base. *Supposons que D soit une droite stable par un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ et soit v un vecteur directeur de D. Alors il existe un scalaire λ tel que*

$$f(v) = \lambda v, \quad v \neq 0, \quad \lambda \in \mathbb{K}.$$

Réciproquement, si un vecteur $v \neq 0$ vérifie la relation précédente, alors la droite D est stable par f. On dit que v est un vecteur propre *de f et que $\lambda \in \mathbb{K}$ est la valeur propre associée.*

Démonstration. La preuve est immédiate : $D = \{\alpha v \,/\, \alpha \in \mathbb{K}\}$ et donc $f(D) = \{\alpha f(v) \,/\, \alpha \in \mathbb{K}\}$ est engendré par $f(v)$. Dire que $f(D) \subset D$ équivaut à dire que $f(v) \in D$, ce qui signifie qu'il existe $\lambda \in \mathbb{K}$ tel que $f(v) = \lambda v$. □

La condition $f(v) = \lambda v$ équivaut à $(f - \lambda \operatorname{Id}_E)(v) = 0$, ce qui entraîne que v est un vecteur non nul de $\operatorname{Ker}(f - \lambda \operatorname{Id}_E)$. L'endomorphisme $f - \lambda \operatorname{Id}_E$ doit alors être non injectif, ce qui implique que $\det(f - \lambda \operatorname{Id}_E) = 0$ d'après le théorème 3.2.3 du chapitre 3. Réciproquement, si $\det(f - \lambda \operatorname{Id}_E) = 0$, alors $\operatorname{Ker}(f - \lambda \operatorname{Id}_E) \neq 0$, et on peut donc trouver $0 \neq v \in \operatorname{Ker}(f - \lambda \operatorname{Id}_E)$; il s'ensuit que $f(v) = \lambda v$, et la droite $D = \mathbb{K}v$ est stable par f. En résumé :

1.1.2. Théorème. *Pour déterminer les valeurs propres d'un endomorphisme $f \in \operatorname{End}_{\mathbb{K}}(E)$, on calcule le polynôme*

$$\lambda \mapsto \det(f - \lambda \operatorname{Id}_E)$$

et on cherche ses racines. Les vecteurs propres associés à la valeur propre λ sont alors les vecteurs non nuls *du noyau* $\operatorname{Ker}(f - \lambda \operatorname{Id}_E)$.

On notera que $\det(\lambda \operatorname{Id}_E - f) = \det(-(f - \lambda \operatorname{Id}_E)) = (-1)^n \det(f - \lambda \operatorname{Id}_E)$.

1.1.3. Définition. *Le polynôme noté*

$$\chi_f(X) = \det(X \operatorname{Id}_E - f) = (-1)^n \det(f - X \operatorname{Id}_E)$$

s'appelle le polynôme caractéristique *de l'endomorphisme f. Lorsque $\lambda \in \mathbb{K}$ est une valeur propre, c'est-à-dire une racine de $\chi_f(X) \in \mathbb{K}[X]$, on appelle* espace propre associé à la valeur propre λ *le sous-espace* (*nécessairement non nul*)

$$E_{f,\lambda} = \operatorname{Ker}(f - \lambda \operatorname{Id}_E) \subset E.$$

On notera $d_\lambda = \dim_{\mathbb{K}} E_{f,\lambda} = \dim_{\mathbb{K}} \operatorname{Ker}(f - \lambda \operatorname{Id}_E) \geqslant 1$ sa dimension, et $m_\lambda \geqslant 1$ la multiplicité de la racine λ dans $\chi_f(X)$.

Il est important de noter que par définition on doit avoir $E_{f,\lambda} \neq \{0\}$. Si ayant calculé une racine λ on trouve que $\dim_{\mathbb{K}} \operatorname{Ker}(f - \lambda \operatorname{Id}_E) = \{0\}$, c'est qu'il y a une erreur quelque part (les candidats à l'examen sont prévenus !). Par ailleurs, on montrera plus loin qu'on a toujours $d_\lambda \leqslant m_\lambda$ (mais il peut y avoir inégalité stricte).

1.1.4. Calcul du polynôme caractéristique. Soit $\mathcal{B} = (e_1, \ldots, e_n)$ une base de E et $A = (a_{ij})_{1 \leqslant i,j \leqslant n} = \operatorname{Mat}_{\mathcal{B}}^{\mathcal{B}}(f)$. On a

$$\chi_f(X) = \chi_A(X) := \det(XI_n - A) = \begin{vmatrix} X - a_{11} & -a_{12} & \ldots & -a_{1n} \\ -a_{21} & X - a_{22} & \ldots & -a_{2n} \\ \vdots & \vdots & & \vdots \\ -a_{n1} & -a_{n2} & \ldots & X - a_{nn} \end{vmatrix}.$$

On voit alors qu'il s'agit d'un polynôme de degré n. Le déterminant contient un terme diagonal

$$(X-a_{11})(X-a_{22})\cdots(X-a_{nn}) = X^n - \sum_{i=1}^{n} a_{ii}X^{n-1} + \cdots + (-1)^n \prod_{i=1}^{n} a_{ii},$$

et tous les autres termes ont au moins deux facteurs situés en dehors de la diagonale (car si $n-1$ facteurs sont sur la diagonale, cela force le dernier à y être aussi). Tous ces termes sont donc des polynômes de degré au plus $n-2$, qui ne modifient pas les coefficients de X^n et de X^{n-1} déjà fournis par le produit diagonal. Par ailleurs, le coefficient constant de χ_A est $\chi_A(0) = \det(-A) = (-1)^n \det(A)$. Nous pouvons donc énoncer :

1.1.5. Théorème. *Avec les notations précédentes, le polynôme caractéristique $\chi_f(X) \in \mathbb{K}[X]$ est un polynôme unitaire de degré n de la forme*

$$\chi_f(X) = X^n - \operatorname{tr}(f)X^{n-1} + \cdots + (-1)^n \det(f)$$

où $\operatorname{tr}(f) := \operatorname{tr}(A) = \sum_{i=1}^{n} a_{ii}$ s'appelle la trace de A (ou de l'endomorphisme f).

1.1.6. Remarque. Il est possible de donner l'expression de chacun des coefficients du polynôme caractéristique $\chi_A(X)$. En fait, le monôme en X^{n-k} est obtenu en prenant un facteur X dans exactement $n-k$ des termes $(X-a_{ii})$ de la diagonale, et le coefficient correspondant est le déterminant mineur $\Delta_{II}(-A) = (-1)^k \Delta_{II}(A)$ associé à la partie complémentaire I de cardinal k (ici $\Delta_{II}(A) = \Delta_{IJ}(A)$ avec $J = I$). Ceci implique que le monôme de degré $n-k$ de $\chi_A(X)$ est

$$(-1)^k \left(\sum_{I \subset \{1,2,\ldots,n\},\ \operatorname{card}(I)=k} \Delta_{II}(A) \right) X^{n-k},$$

et en particulier le terme en X est $(-1)^{n-1} \operatorname{tr}(\operatorname{comat}(A))X$. Ces formules sont en général trop compliquées pour être utiles en pratique. □

1.1.7. Remarque. Si le polynôme caractéristique $\chi_f(X)$ se scinde entièrement sur le corps $\mathbb{K}$, sous la forme

$$\chi_f(X) = \prod_{j=1}^{s} (X-\lambda_j)^{m_j}, \quad \sum_{j=1}^{s} m_j = n,$$

où les λ_j sont les valeurs propres et les m_j leurs multiplicités, on obtient les relations

$$\sum_{j=1}^{s} m_j\lambda_j = \operatorname{tr}(A), \quad \prod_{j=1}^{s} \lambda_j^{m_j} = \det(A).$$

En effet, il suffit pour cela d'identifier le coefficient de X^{n-1} dans de $\chi_f(X)$, ainsi que son coefficient constant. Ceci donne un autre moyen de vérifier qu'on ne s'est pas trompé dans les calculs. □

1.1.8. Remarque. On sait que le déterminant d'un endomorphisme ne dépend pas de la base utilisée. Il en est donc de même pour le polynôme caractéristique et la trace. On peut vérifier plus explicitement ces propriétés en examinant l'effet d'un changement de base. Soient $\mathcal{B}$, $\mathcal{B}'$ des bases de E, $P = \mathrm{Mat}_{\mathcal{B}}(\mathcal{B}')$ la matrice de passage et

$$A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f), \quad A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$$

les matrices respectives de $f \in \mathrm{End}_{\mathbb{K}}(E)$ dans les bases $\mathcal{B}$, $\mathcal{B}'$. Nous savons que $A' = P^{-1}AP$, et comme $P^{-1}I_nP = I_n$, il vient

$$\begin{aligned}\chi_{A'}(X) &= \det(XI_n - A') = \det(XI_n - P^{-1}AP) = \det\big(P^{-1}(XI_n - A)P\big)\\ &= \det(P)^{-1}\det(XI_n - A)\det(P) = \det(XI_n - A).\end{aligned}$$

Par conséquent on a bien

$$\chi_{A'}(X) = \chi_A(X),$$

et en particulier $\det(A') = \det(A)$ et $\mathrm{tr}(A') = \mathrm{tr}(A)$. Mais pour la trace, on peut aussi observer que l'on a

$$\mathrm{tr}(LM) = \sum_{1\leqslant i,j\leqslant n} \ell_{ij}m_{ji} = \mathrm{tr}(ML)$$

pour des matrices $L, M \in \mathcal{M}_{n\times n}(\mathbb{K})$ quelconques, et donc

$$\mathrm{tr}(A') = \mathrm{tr}((P^{-1}A)P) = \mathrm{tr}(P(P^{-1}A)) = \mathrm{tr}(A). \qquad \square$$

1.1.9. Exemples. (a) Soit E un espace vectoriel euclidien de dimension 3, muni d'une base orthonormée $\mathcal{B} = (e_1, e_2, e_3)$ sur le corps $\mathbb{K} = \mathbb{R}$. On considère la rotation $f : E \to E$ d'angle θ et d'axe $\Delta = \mathbb{R}e_3$, donnée dans la base $\mathcal{B}$ par la matrice

$$A = \begin{pmatrix} \cos\theta & -\sin\theta & 0\\ \sin\theta & \cos\theta & 0\\ 0 & 0 & 1\end{pmatrix}.$$

Nous avons

$$\begin{aligned}\chi_A(X) = \det(XI_3 - A) &= \begin{vmatrix} X-\cos\theta & \sin\theta & 0\\ -\sin\theta & X-\cos\theta & \\ 0 & 0 & X-1\end{vmatrix}\\ &= \big((X-\cos\theta)^2 + (\sin\theta)^2\big)(X-1) = (X-1)(X^2 - 2\cos\theta\, X + 1).\end{aligned}$$

Le trinôme $X^2 - 2\cos\theta\, X + 1$ a pour discriminant $\Delta = 4(\cos\theta)^2 - 4 = -4(\sin\theta)^2$, et admet en général deux racines complexes conjuguées $\lambda = \cos\theta \pm i\sin\theta = e^{\pm i\theta}$. Les racines complexes non réelles sont a priori inutilisables sur le corps $\mathbb{K} = \mathbb{R}$ (voir toutefois la section 1.4). Cependant, il y a le cas particulier $\sin\theta = 0$, i.e. $\theta = k\pi$.

• Pour $\theta \equiv 0 \bmod 2\pi$, on a $A = I_3$, d'où un espace propre $E_{f,1} = \mathrm{Ker}(I_3 - A) = E$ pour $\lambda_1 = 1$ (valeur propre triple).

• Pour $\theta \equiv \pi \bmod 2\pi$, il s'agit d'un demi-tour d'axe $\Delta = \mathbb{R}e_3$, on a $E_{f,1} = \Delta = \mathbb{R}e_3$ pour $\lambda_1 = 1$ (simple) et $E_{f,-1} = \mathrm{vect}(e_1, e_2)$ pour $\lambda_2 = -1$ (double).

• Pour $\theta \not\equiv 0 \bmod \pi$, il y a seulement la valeur propre réelle $\lambda_1 = 1$ et l'espace propre associé $E_{f,1} = \Delta = \mathbb{R}e_3$, le plan $\Pi = \Delta^\perp = \mathrm{vect}(e_1, e_2)$ étant stable par f.

(b) On considère maintenant un espace vectoriel E de dimension 3 muni d'une base $\mathcal{B} = (e_1, e_2, e_3)$ sur le corps $\mathbb{K} = \mathbb{R}$, et l'endomorphisme $f : E \to E$ de matrice

$$A = \begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & 2 \\ 2 & 2 & 1 \end{pmatrix}$$

dans la base $\mathcal{B}$. On calcule ici

$$\begin{aligned} \chi_A(X) = \det(XI_3 - A) &= \begin{vmatrix} X-1 & -2 & -2 \\ -2 & X-1 & -2 \\ -2 & -2 & X-1 \end{vmatrix} \\ &= (X-1)^3 - 16 - 12(X-1) = X^3 - 3X^2 - 9X - 5. \end{aligned}$$

Il est visible que $\lambda = -1$ et $\lambda = 5$ sont racines, ce qui donne la factorisation

$$\chi_A(X) = (X+1)(X^2 - 4X - 5) = (X+1)^2(X-5).$$

On a ainsi une racine double $\lambda_1 = -1$ et une racine simple $\lambda_2 = 5$. Les espaces propres correspondants sont donnés par

$$\mathrm{Ker}(f + \mathrm{Id}_E) : \quad \begin{pmatrix} 2 & 2 & 2 \\ 2 & 2 & 2 \\ 2 & 2 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0, \quad \text{soit} \;\; x_1 + x_2 + x_3 = 0,$$

par conséquent $E_{f,-1}$ est un plan vectoriel, ayant par exemple pour base

$$v_1 \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \quad v_2 \begin{pmatrix} 1 \\ 0 \\ -1 \end{pmatrix} \quad \text{relativement à } \mathcal{B},$$

tandis que

$$\mathrm{Ker}(f - 5\,\mathrm{Id}_E) : \quad \begin{pmatrix} -4 & 2 & 2 \\ 2 & -4 & 2 \\ 2 & 2 & -4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = 0.$$

Un calcul simple montre que $E_{f,5}$ est la droite vectorielle de vecteur directeur

$$v_3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{relativement à } \mathcal{B}.$$

On a ici $f(v_1) = -v_1$, $f(v_2) = -v_2$, $f(v_3) = 5v_3$, et on vérifie aisément que $\mathcal{B}' = (v_1, v_2, v_3)$ est une base de E [$\det_{\mathcal{B}}(\mathcal{B}') = -3$]. Dans cette nouvelle base, l'endomorphisme f a pour matrice

$$A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 5 \end{pmatrix},$$

c'est-à-dire une matrice diagonale, nettement plus simple à étudier que la matrice initiale A dont on était parti. □

1.1.10. Remarque. On a en général la formule

$$\chi_{f+\alpha\,\mathrm{Id}_E}(X) = \chi_f(X-\alpha),$$

en effet $\chi_{f+\alpha\,\mathrm{Id}_E}(X) = \det(X\,\mathrm{Id}_E - (f+\alpha\,\mathrm{Id}_E)) = \det((X-\alpha)\,\mathrm{Id}_E - f)$.

1.2. Endomorphismes diagonalisables

Démontrons d'abord une propriété générale des espaces propres.

1.2.1. Théorème. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme et E_{f,λ_j}, $1 \leqslant j \leqslant s$ ses espaces propres, où les λ_j sont les valeurs propres distinctes (s'il y en a !). Alors les sous-espaces E_{f,λ_j} sont en somme directe, c'est-à-dire que*

$$V = \sum_{1\leqslant j\leqslant s} E_{f,\lambda_j} = \bigoplus_{1\leqslant j\leqslant s} E_{f,\lambda_j} \subset E,$$

ou encore

$$\dim V = \sum_{1\leqslant j\leqslant s} \dim E_{f,\lambda_j} \leqslant \dim E.$$

On remarquera que dans l'exemple 1.1.9 (a) avec $\theta \not\equiv 0 \bmod \pi$, il n'y a pas d'espaces propres, i.e. $s = 0$, et donc $V = \{0\}$. Bien entendu il n'y a rien à démontrer dans un tel cas. Dans l'exemple 1.1.9 (b), on a en revanche $V = E$.

Démonstration. On va montrer que pour tout $p = 1, 2, \ldots, s$ la somme

$$V_p = \sum_{1\leqslant j\leqslant p} E_{f,\lambda_j}$$

est une somme directe. Pour cela, il faut voir que si

$$(*) \qquad x_1 + x_2 + \cdots + x_p = 0 \quad \text{avec } x_j \in E_{f,\lambda_j}, \text{ alors } x_1 = x_2 = \ldots = x_p = 0.$$

On raisonne par récurrence sur p, le résultat étant évident si $p = 1$. Supposons $p \geqslant 2$ et le résultat déjà démontré pour $p-1$. En appliquant f à l'égalité $(*)$, on trouve

$$(**) \qquad f(x_1 + x_2 + \cdots + x_p) = \lambda_1 x_1 + \lambda_2 x_2 + \cdots + \lambda_p x_p = 0.$$

En multipliant $(*)$ par λ_p et en soustrayant de $(**)$, il vient

$$(\lambda_1 - \lambda_p)x_1 + (\lambda_2 - \lambda_p)x_1 + \cdots + (\lambda_{p-1} - \lambda_p)x_{p-1} = 0.$$

Il s'agit d'une somme de $p-1$ vecteurs $(\lambda_j - \lambda_p)x_j \in E_{f,\lambda_j}$, $1 \leqslant j \leqslant p-1$. D'après l'hypothèse de récurrence ces vecteurs sont nuls, et comme $\lambda_j - \lambda_p \neq 0$, on en

déduit $x_1 = x_2 = \ldots = x_{p-1} = 0$. Mais $(*)$ entraîne alors aussi $x_p = 0$, et la propriété est bien démontrée à l'ordre p. □

La définition de ce qu'on entend par endomorphisme diagonalisable est assez peu surprenante :

1.2.2. Définition. *Un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est dit* diagonalisable *s'il existe une base $\mathcal{B}' = (v_1, \ldots, v_n)$ dans laquelle la matrice $A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$ est diagonale, c'est-à-dire de la forme*

$$A' = \begin{pmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \alpha_n \end{pmatrix}, \quad \alpha_j \in \mathbb{K}.$$

Si $\mathcal{B} = (e_1, \ldots, e_n)$ est une base donnée et $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f)$ la matrice associée, alors par définition f est diagonalisable si et seulement il existe une matrice de passage P (inversible, donc) telle que $A' = P^{-1}AP$ soit diagonale.

1.2.3. Théorème. *Soit E un $\mathbb{K}$-espace vectoriel de dimension n, et $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme. Alors les propriétés suivantes sont équivalentes :*

(a) *f est diagonalisable ;*

(b) *E admet une base $\mathcal{B}' = (v_1, \ldots, v_n)$ de vecteurs propres pour f ;*

(c) *Les espaces propres E_{f,λ_j}, $1 \leqslant j \leqslant s$, vérifient* $E = \bigoplus_{1 \leqslant j \leqslant s} E_{f,\lambda_j}$;

(d) *Si $d_j = \dim E_{f,\lambda_j}$, alors* $\sum_{1 \leqslant j \leqslant s} d_j = n.$

Si ces propriétés équivalentes sont satisfaites, le polynôme caractéristique de f est complètement scindé sur $\mathbb{K}$, donné par

$$\chi_f(X) = \prod_{1 \leqslant j \leqslant s} (X - \lambda_j)^{d_j}, \quad \lambda_j \in \mathbb{K}.$$

Démonstration. (a) ⇒ (b). En effet, si f est diagonalisable, il existe une base $\mathcal{B}' = (v_1, \ldots, v_n)$ dans laquelle

$$(*) \qquad \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f) = \begin{pmatrix} \alpha_1 & 0 & \ldots & 0 \\ 0 & \alpha_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \alpha_n \end{pmatrix}, \quad \alpha_j \in \mathbb{K},$$

c'est-à-dire que $f(v_j) = \alpha_j v_j$, i.e. v_j est un vecteur propre de valeur propre α_j. L'implication réciproque (b) ⇒ (a) est évidente.

(b) $\Rightarrow$ (c) Supposons que la matrice de f s'écrive sous la forme diagonale $(*)$ dans $\mathcal{B}'$. Alors les coefficients α_j ne sont pas nécessairement distincts, mais après permutation éventuelle des vecteurs v_j, on peut supposer

$$\alpha_1 = \ldots = \alpha_{d_1} = \lambda_1, \quad \alpha_{d_1+1} = \ldots = \alpha_{d_1+d_2} = \lambda_2, \ \ldots \ , \alpha_{n-d_s+1} = \ldots = \alpha_n = \lambda_s$$

avec des éléments $\lambda_1, \ldots, \lambda_s \in \mathbb{K}$ deux à deux distincts. Ceci donne

$$E_{f,\lambda_1} = \mathrm{vect}(v_1, \ldots, v_{d_1}), \quad E_{f,\lambda_2} = \mathrm{vect}(v_{d_1+1}, \ldots, v_{d_1+d_2}), \ \ldots \ ,$$
$$E_{f,\lambda_s} = \mathrm{vect}(v_{n-d_s+1}, \ldots, v_n),$$

et on a donc bien $E = \bigoplus_{1 \leqslant j \leqslant s} E_{f,\lambda_j}$. Réciproquement, si cette égalité a lieu, alors on obtient une base $\mathcal{B}' = (v_1, \ldots, v_n)$ de vecteurs propres en prenant pour $(v_1, \ldots, v_{d_1})$ une base de E_{f,λ_1}, pour $(v_{d_1+1}, \ldots, v_{d_1+d_2})$ une base de E_{φ,λ_2}, $\ldots$, pour $(v_{n-d_s+1}, \ldots, v_n)$ une base de E_{φ,λ_s}, donc (c) $\Rightarrow$ (b).

(c) $\Rightarrow$ (d) est évident, et (d) $\Rightarrow$ (c) résulte du fait que la somme $\sum_{1 \leqslant j \leqslant s} E_{f,\lambda_j}$ est directe (théorème 1.2.1).

Enfin, si a l'expression $(*)$ pour la matrice de f, le polynôme caractéristique est bien

$$\chi_f(X) = \prod_{1 \leqslant j \leqslant n} (X - \alpha_j) = \prod_{1 \leqslant j \leqslant s} (X - \lambda_j)^{d_j}. \qquad \square$$

1.2.4. Remarque. Dans l'exemple 1.1.9 (b), il était en fait inutile de vérifier que $\mathcal{B}' = (v_1, v_2, v_3)$ était bien une base. En effet, comme on avait pris pour (v_1, v_2) une base de $E_{f,-1}$ et pour v_3 une base de $E_{f,5}$, la propriété "automatique" que $E_{f,-1}$, $E_{f,5}$ soient en somme directe et le fait que $\dim E_{f,-1} + \dim E_{f,5} = 2+1 = 3 = \dim E$ impliquait nécessairement que (v_1, v_2, v_3) soit une base.

1.2.5. Remarque. En général, la somme directe

$$V = \bigoplus_{1 \leqslant j \leqslant s} E_{f,\lambda_j}$$

est un sous-espace stable par f (comme somme de sous-espaces stables), et c'est en fait le plus grand sous-espace stable $S \subset E$ tel que $f_{|S} : S \to S$ soit diagonalisable. En effet, il est clair que $f_{|V}$ est diagonalisable, et réciproquement, si $f_{|S} : S \to S$ est diagonalisable, alors S est somme de sous-espaces propres $E_{f_{|S},\lambda_j}$ qui sont contenus dans les E_{f,λ_j} correspondants, donc

$$S = \bigoplus_j E_{f_{|S},\lambda_j} \subset \bigoplus_j E_{f,\lambda_j} = V. \qquad \square$$

1.3. Calcul par blocs ; multiplicité des valeurs propres

Considérons une matrice $A \in \mathcal{M}_{n\times n}(\mathbb{K})$ triangulaire par blocs (par exemple supérieurement)

$$A = \begin{pmatrix} A_1 & R_{12} & \dots & R_{1s} \\ O & A_2 & \dots & R_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & A_s \end{pmatrix},$$

où les $A_i \in \mathcal{M}_{n_i\times n_i}(\mathbb{K})$ sont des blocs carrés et $n = \sum n_i$. Alors

$$XI_n - A = \begin{pmatrix} XI_{n_1} - A_1 & -R_{12} & \dots & -R_{1s} \\ O & XI_{n_2} - A_2 & \dots & -R_{2s} \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & XI_{n_s} - A_s \end{pmatrix},$$

et les règles de calcul des déterminants impliquent la formule :

1.3.1. Formule. *Si $A \in \mathcal{M}_{n\times n}(\mathbb{K})$ est triangulaire par blocs avec des blocs diagonaux $A_i \in \mathcal{M}_{n_i\times n_i}(\mathbb{K})$, $1 \leqslant i \leqslant s$, alors*

$$\chi_A(X) = \chi_{A_1}(X)\chi_{A_2}(X)\cdots\chi_{A_s}(X).$$

On va déduire de là des résultats importants pour les endomorphismes. Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme d'un espace E de dimension n, muni d'une base $\mathcal{B} = (e_1, \dots, e_n)$. Supposons que f possède un sous-espace stable S, et soit $\mathcal{B}'_S = (v_1, \dots, v_p)$ une base de S. D'après le théorème de la base incomplète, on peut compléter cette famille en une base $\mathcal{B}' = (v_1, \dots, v_n)$ de E, et on obtient alors une décomposition en somme directe

$$E = S \oplus T, \quad S = \mathrm{vect}(v_1, \dots, v_p), \quad T = \mathrm{vect}(v_{p+1}, \dots, v_n).$$

Notons $A = \mathrm{Mat}^{\mathcal{B}}_{\mathcal{B}}(f)$, $A' = \mathrm{Mat}^{\mathcal{B}'}_{\mathcal{B}'}(f)$. Les p premiers vecteurs colonnes de $A' = (a'_{ij})_{1\leqslant i,j\leqslant n}$ sont donnés par $f(v_j) = \sum_{i=1}^{p} a'_{ij} v_i \in S$, $1 \leqslant j \leqslant p$ (car un vecteur de S ne comporte pas de composantes sur les vecteurs $v_{p+1}, \dots, v_n$). Ceci implique que A' est de la forme

$$A' = \begin{array}{c} \begin{matrix} S & T \end{matrix} \\ \left(\begin{array}{c|c} U & V \\ \hline O & W \end{array}\right) \begin{matrix} S \\ T \end{matrix} \end{array} \quad \text{relativement à } \mathcal{B}'.$$

Désignons par $\pi_T : E \to T$ la projection sur T parallèlement à S. Les endomorphismes

$$f_{|S} : S \to S, \quad g = \pi_T \circ f_{|T} : T \to T$$

induits par f admettent les matrices respectives

$$U = \mathrm{Mat}^{\mathcal{B}'_S}_{\mathcal{B}'_S}(f_{|S}), \quad W = \mathrm{Mat}^{\mathcal{B}'_T}_{\mathcal{B}'_T}(g) \quad \text{dans la base } \mathcal{B}'_T = (v_{p+1}, \dots, v_n).$$

On a alors d'après 1.3.1 la relation $\chi_{A'}(X) = \chi_U(X)\chi_W(X)$, ce qui se retraduit comme suit au niveau des endomorphismes.

1.3.2. Théorème. *Si un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ admet un sous-espace stable S, alors si on prend une décomposition $E = S \oplus T$ et qu'on considère les endomorphismes $f_{|S} : S \to X$ et $g = \pi_T \circ f_{|T}$ induits par f, on a la relation*

$$\chi_f(X) = \chi_{f_{|S}}(X)\chi_g(X).$$

En particulier le polynôme caractéristique $\chi_f(X)$ est divisible par $\chi_{f_{|S}}(X)$.

Si nous appliquons ce résultat au cas d'un sous-espace propre

$$S_j = E_{f,\lambda_j} = \mathrm{Ker}(f - \lambda_j \,\mathrm{Id}_E), \quad d_j = \dim S_j$$

(trivialement stable puisque $f(x) = \lambda_j x$ pour tout $x \in S_j$), alors $\chi_f(X)$ est divisible par $\chi_{f_{|S_j}}(X) = (X - \lambda_j)^{d_j}$, ce qui entraîne immédiatement :

1.3.3. Théorème. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$, $\lambda_j \in \mathbb{K}$ une valeur propre de f, m_j sa multiplicité dans le polynôme caractéristique $\chi_f(X)$, et $d_j = \dim E_{f,\lambda_j}$. Alors*

$$d_j \leqslant m_j.$$

1.3.4. Conséquence. *Si $\chi_f(X)$ admet une racine λ_j simple (i.e. si la multiplicité de λ_j est $m_j = 1$), alors on a nécessairement $d_j = m_j = 1$.*

Démonstration. En effet $E_{f,\lambda_j} \neq \{0\}$, par conséquent $1 \leqslant d_j \leqslant m_j \leqslant 1$. □

1.3.5. Exemple. Considérons l'espace $E = \mathbb{K}^n$ muni de sa base canonique $\mathcal{B} = (e_1, \ldots, e_n)$, et l'endomorphisme $f_\alpha : E \to E$ dont la matrice relativement à $\mathcal{B}$ est ce qu'on appelle un *bloc de Jordan (triangulaire inférieur)* de taille n :

$$J_{\alpha;n} = \begin{pmatrix} \alpha & 0 & \ldots & 0 \\ 1 & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 1 & \alpha \end{pmatrix},$$

avec α sur la diagonale principale, 1 sur la diagonale au dessus et 0 ailleurs. On a

$$\chi_{f_\alpha}(X) = \det(XI_n - J_{\alpha;n}) = \begin{vmatrix} X-\alpha & 0 & \ldots & 0 \\ -1 & X-\alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & -1 & X-\alpha \end{vmatrix} = (X-\alpha)^n,$$

et $\chi_{f_\alpha}(X)$ admet une unique valeur propre $\lambda_1 = \alpha$ de multiplicité $m_1 = n$. Comme $f_\alpha - \alpha \,\mathrm{Id}_E = f_0$, on a

$$E_{f_\alpha,\alpha} = \mathrm{Ker}(f_\alpha - \alpha \,\mathrm{Id}_E) = \mathrm{Ker}\, f_0 = \mathbb{K}e_n \quad \Rightarrow \quad d_1 = 1 < m_1 = n.$$ □

Les résultats qui précèdent donnent une nouvelle caractérisation des endomorphismes diagonalisables (qui implique en particulier que les blocs de Jordan de taille $n > 1$ *ne sont pas diagonalisables*).

1.3.6. Théorème. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$, $\dim E = n$. Pour que l'endomorphisme f soit diagonalisable, il faut et il suffit que le polynôme caractéristique $\chi_f(X) \in \mathbb{K}[X]$ soit scindé sur le corps $\mathbb{K}$, c'est-à-dire*

$$\chi_f(X) = \prod_{1 \leqslant j \leqslant s} (X - \lambda_j)^{m_j}, \quad \lambda_j \in \mathbb{K} \ \text{ 2 à 2 distincts},$$

et de plus, que la dimension $d_j = \dim E_{f,\lambda_j}$ soit égale à la multiplicité m_j de la valeur propre λ_j : $d_j = m_j$.

Démonstration. La condition que $\chi_f(X)$ soit scindé et que $d_j = m_j$ est clairement nécessaire pour que f soit diagonalisable, d'après le théorème 1.2.3. Mais réciproquement, si cette condition est satisfaite, on a $\sum d_j = \sum m_j = n$, et le critère 1.2.3 (d) montre que f est diagonalisable. □

1.3.7. Synthèse. On utilise fréquemment la terminologie suivante :

$m_j = $ multiplicité de λ_j dans χ_f = *multiplicité algébrique de λ_j*,
$d_j = $ dimension sous-espace propre E_{f,λ_j} = *multiplicité géométrique de λ_j*.

On a toujours $d_j \leqslant m_j$. Le critère pour qu'un endomorphisme f soit diagonalisable est le suivant :

$f \in \mathrm{End}_{\mathbb{K}}(E)$ diagonalisable $\iff$ χ_f scindé et
$\forall j,\ d_j$ (multiplicité géométrique de λ_j) $= m_j$ (multiplicité algébrique de λ_j).

À l'inverse, lorsque χ_f est scindé, pour voir que f est non diagonalisable, il suffit de trouver une valeur propre λ_j pour laquelle $d_j < m_j$.

On notera en outre que sur $\mathbb{K} = \mathbb{C}$ tout polynôme est scindé ; si les racines de χ_f sont simples, f est donc diagonalisable d'après 1.3.4.

1.4. Valeurs propres complexes d'endomorphismes réels

Un polynôme réel $F \in \mathbb{R}[X]$ de degré $n \geqslant 1$ n'a pas nécessairement de racines réelles, mais il a toujours des racines complexes d'après le théorème de d'Alembert-Gauss. Il est donc intéressant de "savoir quoi faire" des valeurs propres complexes.

1.4.1. Théorème. *Soit f un endomorphisme sur un espace vectoriel E de dimension finie n sur $\mathbb{R}$, représenté par une matrice $A \in \mathcal{M}_{n\times n}(\mathbb{R})$ dans une base $\mathcal{B} = (e_1, \ldots, e_n)$ de E. Si $\lambda \in \mathbb{C} \smallsetminus \mathbb{R}$ est une racine complexe non réelle et si $Z = U + iV$ avec $U, V \in \mathcal{M}_{n\times 1}(\mathbb{R})$ est un vecteur colonne complexe non nul tel que $AZ = \lambda Z$, alors les vecteurs $u, v \in E$ de matrices respectives U, V engendrent un plan $P = \mathrm{vect}(u, v) \subset E$ stable par f.*

Démonstration. On a $\chi_f(X) = \chi_A(X)$, et comme $\mathcal{M}_{n\times n}(\mathbb{R}) \subset \mathcal{M}_{n\times n}(\mathbb{C})$, on peut traiter A comme une matrice complexe. Si $\lambda \in \mathbb{C} \smallsetminus \mathbb{R}$ est une racine complexe non réelle, la théorie générale implique l'existence d'un vecteur colonne complexe Z non nul tel que $AZ = \lambda Z$. Si l'on écrit $Z = U + iV$ et $\lambda = \alpha + i\beta$ avec U, V, α, β réels, il vient

$$AU + iAV = (\alpha + i\beta)(U + iV) \Longleftrightarrow AU = \alpha U - \beta V, \quad AV = \beta U + \alpha V.$$

Les vecteurs colonnes U et V ne peuvent être $\mathbb{R}$-colinéaires, sinon on aurait disons $U \neq 0$ et $V = \gamma U$ avec $\gamma \in \mathbb{R}$, donc $Z = U + i\gamma U = \mu U$ serait $\mathbb{C}$-colinéaire à U, où $\mu = 1 + i\gamma \in \mathbb{C}^*$. Par conséquent $U = \mu^{-1} Z$ serait aussi vecteur propre de A pour la valeur propre λ, mais ceci est contradictoire avec la relation $AU = \lambda U$ où A, U sont réels et λ non réel (le raisonnement est le même si $V \neq 0$ et $U = \gamma V$). Considérons alors le plan vectoriel $P = \mathrm{vect}(u, v) \subset E$ engendré par les vecteurs u, v de coordonnées U, V. On a les relations

$$f(u) = \alpha u - \beta v \in P, \quad f(v) = \beta u + \alpha v \in P \Longrightarrow f(P) \subset P.$$

On voit donc que P est un plan stable de f. □

1.4.2. Interprétation matricielle. On conserve les hypothèses et notations du théorème 1.4.1. Pour expliciter encore un peu plus la situation, complétons (u, v) en une base $\mathcal{B}' = (u, v, t_3, \ldots, t_n)$ de E. Alors f admet dans $\mathcal{B}'$ la matrice

$$A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f) = \begin{pmatrix} \alpha & \beta & * & \ldots & * \\ -\beta & \alpha & * & \ldots & * \\ 0 & 0 & a'_{33} & \ldots & a'_{3n} \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & a'_{n3} & \ldots & a'_{nn} \end{pmatrix}$$

Soit $T = \mathrm{vect}(t_3, \ldots, t_n)$ de sorte que $E = P \oplus T$, et soit $g : T \to T$ l'endomorphisme de matrice $(a'_{ij})_{3 \leqslant i,j \leqslant n}$. Alors

$$\chi_f(X) = \chi_{A'}(X) = \chi_{f_{|P}}(X)\chi_g(X) = ((X - \alpha)^2 + \beta^2)\, \chi_g(X),$$

soit

$$\chi_f(X) = (X - \lambda)(X - \overline{\lambda})\, \chi_g(X). \tag{1.4.3}$$

La conclusion de ces calculs est que le plan stable $P = \mathrm{vect}(u, v)$ associé au vecteur colonne complexe $Z = U + iV$ tel que $AZ = \lambda Z$ fournit dans $\chi_f(X)$ un facteur $\chi_{f|P}(X) = (X - \lambda)(X - \overline{\lambda})$ qui factorise des deux valeurs propres complexes conjuguées $\lambda = \alpha + i\beta$ et $\overline{\lambda} = \alpha - i\beta$.

1.4.4. Corollaire. *Soit $f \in \mathrm{End}_{\mathbb{R}}(E)$ un endomorphisme d'un espace vectoriel réel de dimension $n = \dim_{\mathbb{R}} E \geqslant 1$. Alors f possède au moins une droite stable D (si $\chi_f(X)$ a une racine réelle) ou un plan stable P (si $\chi_f(X)$ a deux racines complexes conjuguées non réelles).*

2. Théorèmes de structure des endomorphismes

2.1. Polynômes d'endomorphismes et polynôme minimal

Soit E un $\mathbb{K}$-espace vectoriel de dimension n. Alors $(\mathrm{End}_{\mathbb{K}}(E), +, \circ)$ est un anneau (non commutatif si $n \geqslant 2$). C'est aussi un $\mathbb{K}$-espace vectoriel de dimension n^2, car si $\mathcal{B} = (e_1, \dots, e_n)$ est une base de E, on a un isomorphisme d'anneaux $\mathbb{K}$-linéaire

$$\varphi_{\mathcal{B}} : \mathrm{End}_{\mathbb{K}}(E) \to \mathcal{M}_{n\times n}(\mathbb{K}), \quad f \mapsto M = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f), \tag{2.1.1}$$

sur l'anneau $(\mathcal{M}_{n\times n}(\mathbb{K}), +, \times)$ des matrices $n \times n$. Cet isomorphisme dépend de la base $\mathcal{B}$, et si on prend une autre base $\mathcal{B}' = (e'_1, \dots, e'_n)$ et $P = \mathrm{Mat}_{\mathcal{B}}(\mathcal{B}')$, on obtient un diagramme d'isomorphismes d'anneaux

$$\begin{array}{ccccc} & \overset{\varphi_{\mathcal{B}}}{\nearrow} & \mathcal{M}_{n\times n}(\mathbb{K}) & & M \\ f \in \mathrm{End}_{\mathbb{K}}(E) & & \downarrow & & \downarrow \\ & \underset{\varphi_{\mathcal{B}'}}{\searrow} & \mathcal{M}_{n\times n}(\mathbb{K}) & & M' = P^{-1}MP. \end{array} \tag{2.1.2}$$

On notera en particulier que l'on a bien $P^{-1}(M_1 + M_2)P = P^{-1}M_1P + P^{-1}M_2P$ et $P^{-1}(M_1M_2)P = (P^{-1}M_1P)(P^{-1}M_2P)$, donc les flèches verticales sont bien des (iso)morphismes d'anneaux. On définit maintenant les polynômes d'endomorphismes et de matrices.

2.1.3. Définition. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ et $M \in \mathcal{M}_{n\times n}(\mathbb{K})$. Si*

$$Q(X) = a_0 + a_1X + \cdots + a_dX^d \in \mathbb{K}[X],$$

on définit

$$Q(f) = a_0\,\mathrm{Id}_E + a_1f + \cdots + a_df^d \in \mathrm{End}_{\mathbb{K}}(E),$$

$$Q(M) = a_0I_n + a_1M + \cdots + a_dM^d \in \mathcal{M}_{n\times n}(\mathbb{K}).$$

Les confusions étant hélas trop fréquentes, le lecteur prendra garde au fait qu'à l'élément unité 1 de l'anneau $\mathbb{K}[X]$ pour la multiplication, doit correspondre l'élément unité $\mathrm{Id}_E \in \mathrm{End}_{\mathbb{K}}(E)$ et l'élément unité $I_n \in \mathcal{M}_{n\times n}(\mathbb{K})$. Si $v \in E$ est un vecteur, l'image par l'endomorphisme $Q(f)$ du vecteur v est ainsi

$$Q(f)(v) = a_0v + a_1f(v) + \cdots + a_df^d(v)$$

(et la notation $Q(f(v))$ n'a aucun sens !). L'observation suivante est presque immédiate, mais néanmoins fondamentale.

2.1.4. Théorème. *L'endomorphisme f et la matrice $M \in \mathcal{M}_{n\times n}(\mathbb{K})$ étant fixés, on a des morphismes d'anneaux*

$$\mathbb{K}[X] \to \mathrm{End}_{\mathbb{K}}(E), \quad Q \mapsto Q(f),$$

$$\mathbb{K}[X] \to \mathcal{M}_{n\times n}(\mathbb{K}), \quad Q \mapsto Q(M).$$

Démonstration. Soient $Q = \sum_i a_i X^i, R = \sum_j b_j X^j \in \mathbb{K}[X]$. Alors $Q + R = \sum_i (a_i + b_i) X^i$ et $QR = \sum_k c_k X^k$ avec $c_k = \sum_{i+j=k} a_i b_j$. En convenant que $f^0 = \mathrm{Id}_E$ (et de même $M^0 = I_n$), ceci donne

$$(Q+R)(f) = \sum_i (a_i + b_i) f^i = \sum_i a_i f^i + \sum b_i f^i = Q(f) + R(f)$$

et

$$\begin{aligned}(QR)(f) &= \sum_k c_k f^k = \sum_k \left(\sum_{i+j=k} a_i b_j \right) f^k \\ &= \sum_{i,j} a_i b_j f^{i+j} = \sum_{i,j} a_i b_j f^i \circ f^j \\ &= \sum_i a_i f^i \circ \sum_j b_j f^j = Q(f) \circ R(f).\end{aligned}$$

La vérification est quasi-identique pour les matrices (au remplacement près de f par M, et de la loi $\circ$ par la loi $\times$), et sera donc omise. □

On peut composer ces morphismes d'anneaux avec les morphismes $\varphi_{\mathcal{B}}$ et aussi avec le changement de base. Si $f \in \mathrm{End}_{\mathbb{K}}(E)$ est donné et $M = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f)$, $M' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$, on obtient un diagramme de morphismes d'anneaux

(2.1.5)
$$\begin{array}{ccccc} & & \varphi_{\mathcal{B}} \nearrow & \mathcal{M}_{n\times n}(\mathbb{K}) & Q(M) \\ Q \in \mathbb{K}[X] & \longmapsto & Q(f) \in \mathrm{End}_{\mathbb{K}}(E) & \downarrow & \downarrow \\ & & \varphi_{\mathcal{B}'} \searrow & \mathcal{M}_{n\times n}(\mathbb{K}) & Q(M'), \end{array}$$

et on a ici

$$Q(M') = Q(P^{-1}MP) = \sum_i a_i (P^{-1}MP)^i = \sum_i a_i P^{-1} M^i P = P^{-1} Q(M) P.$$

Prière de relire au moins 3 fois tout ce paragraphe et de ne pas "s'emmêler les pinceaux" ! Un autre fait important est que les polynômes d'endomorphismes commutent, bien que l'anneau des endomorphismes ne soit en général pas commutatif – la raison étant que celui des polynômes l'est.

2.1.6. Propriété. *Pour tous polynômes $Q, R \in \mathbb{K}[X]$, on a commutation des polynômes d'endomorphismes*

$$Q(f) \circ R(f) = R(f) \circ Q(f).$$

Démonstration. En effet $Q(f) \circ R(f) = (QR)(f) = (RQ)(f) = R(f) \circ Q(f)$. □

On exploite souvent cette commutation au moyen du lemme suivant.

2.1.7. Proposition. *Soit $f, g \in \mathrm{End}_{\mathbb{K}}(E)$ tel que $f \circ g = g \circ f$. Alors* $\mathrm{Ker}\, g$ *et* $\mathrm{Im}\, g$ *sont des sous-espaces stables par f, c'est-à-dire que*

$$f \circ g = g \circ f \quad \Longrightarrow \quad f(\mathrm{Ker}\, g) \subset \mathrm{Ker}\, g, \quad f(\mathrm{Im}\, g) \subset \mathrm{Im}\, g.$$

Démonstration. Pour tout $x \in \mathrm{Ker}\, g$ on a $g(f(x)) = f(g(x)) = f(0) = 0$, donc $f(x) \in \mathrm{Ker}\, g$, et $\mathrm{Ker}\, g$ est bien stable par f. D'autre part, pour tout $y = g(x) \in \mathrm{Im}\, g$, on a $f(y) = f(g(x)) = g(f(x)) \in \mathrm{Im}\, g$, de sorte que $\mathrm{Im}\, g$ est également stable par f. □

Nous pouvons maintenant combiner les deux propriétés précédentes en prenant $R(X) = X$, et donc $R(f) = f$.

2.1.8. Corollaire. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$. Alors, pour tout $Q \in \mathbb{K}[X]$, les sous-espaces* $\mathrm{Ker}\, Q(f)$ *et* $\mathrm{Im}\, Q(f)$ *sont stables par f.* □

2.1.9. Polynômes annulateurs et polynôme minimal. Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme. D'après ce qui précède, on a un morphisme d'anneaux

$$\psi_f : \mathbb{K}[X] \to \mathrm{End}_{\mathbb{K}}(E), \quad Q \mapsto Q(f),$$

et le noyau

$$\mathrm{Ker}\, \psi_f = \big\{ Q \in \mathbb{K}[X] \,/\, Q(f) = 0 \big\}$$

consiste en ce qu'on appelle les *polynômes annulateurs* de f. Comme on a $\dim \mathbb{K}[X] = +\infty$ et $\dim \mathrm{End}_{\mathbb{K}}(E) = n^2$, le morphisme ψ_f ne peut pas être injectif, et on a donc $\mathrm{Ker}\, \psi_f \neq \{0\}$. Le noyau $\mathrm{Ker}\, \psi_f$ est en fait un idéal de $\mathbb{K}[X]$: en effet, si $Q(f) = 0$ et $R(f) = 0$ alors $(Q + R)(f) = 0$, et pour tout $G \in \mathbb{K}[X]$, $(GQ)(f) = G(f) \circ Q(f) = 0$, donc on a bien

$$\begin{aligned} &\forall Q, R \in \mathrm{Ker}\, \psi_f, \quad Q + R \in \mathrm{Ker}\, \psi_f, \\ &\forall G \in \mathbb{K}[X], \quad \forall Q \in \mathrm{Ker}\, \psi_f, \quad GQ \in \mathrm{Ker}\, \psi_f. \end{aligned}$$

Comme l'anneau $\mathbb{K}[X]$ est principal, l'idéal $\mathrm{Ker}\, \psi_f$ est un nécessairement un idéal principal $\langle \mu_f(X) \rangle$, engendré par un polynôme non nul de degré minimal de $\mathrm{Ker}\, \psi_f$, uniquement déterminé si on le choisit unitaire. L'idéal $\mathrm{Ker}\, \psi_f$ est appelé *idéal annulateur* de f.

2.1.10. Définition. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$. On appelle polynôme minimal de f, noté $\mu_f(X)$, un polynôme unitaire $Q(X)$ de degré minimal tel que $Q(f) = 0$. L'ensemble $\{Q \in \mathbb{K}[X] \,/\, Q(f) = 0\}$ des polynômes annulateurs coïncide alors avec l'idéal $\langle \mu_f(X) \rangle$ des multiples de $\mu_f(X)$.*

Écrivons

$$\mu_f(X) = c_0 + c_1 X + \cdots + c_{d-1} X^{d-1} + X^d.$$

Comme $\mu_f(f) = c_0\, \mathrm{Id}_E + c_1 f + \cdots + c_{d-1} f^{d-1} + f^d$, on obtient

$$f^d = -c_0\, \mathrm{Id}_E - c_1 f - \cdots - c_{d-1} f^{d-1},$$

et en multipliant par f^{j-d} pour $j \geqslant d$ on obtient

$$f^j = -c_0 f^{j-d} - c_1 f^{j-d+1} - \cdots - c_{d-1} f^{j-1}, \quad j \geqslant d.$$

Par récurrence, les $(f^j)_{j\geqslant d}$, sont combinaisons linéaires de $\mathrm{Id}_E, f, \ldots, f^{d-1}$, d'où

$$\mathrm{Im}\,\psi_f = \mathrm{vect}(f^j)_{j\in\mathbb{N}} = \mathrm{vect}(\mathrm{Id}_E, f, \ldots, f^{d-1}). \tag{2.1.11}$$

On remarquera que les endomorphismes $\mathrm{Id}_E, f, \ldots, f^{d-1}$ sont nécessairement linéairement indépendants, sinon on obtiendrait un polynôme Q de degré $\leqslant d-1$ tel que $Q(f) = 0$, ce qui contredirait la minimalité de μ_f. On en déduit

$$\dim \mathrm{Im}\,\psi_f = \dim \mathrm{vect}(f^j)_{j\in\mathbb{N}} = d = \deg \mu_f(X). \tag{2.1.12}$$

Comme $\mathrm{Im}\,\psi_f \subset \mathrm{End}_{\mathbb{K}}(E)$, on en déduit déjà $d \leqslant \dim \mathrm{End}_{\mathbb{K}}(E) = n^2$, mais on verra plus loin que l'on a en fait toujours $d \leqslant n$ (voir § 2.3).

2.1.13. Propriétés fondamentales du polynôme minimal.

(a) Une homothétie vectorielle $f = \alpha\,\mathrm{Id}_E$ admet comme polynôme minimal $\mu_f(X) = X - \alpha$ (car $Q(X) = X - \alpha$ donne $Q(f) = f - \alpha\,\mathrm{Id}_E$), et dans ce cas l'image

$$\mathrm{Im}\,\psi_f = \mathrm{vect}(f^j)_{j\in\mathbb{N}} = \mathbb{K}\,\mathrm{Id}_E$$

est de dimension 1. Ces propriétés caractérisent les homothéties vectorielles.

(b) Considérons une matrice diagonale par blocs

$$A = \begin{pmatrix} A_1 & O & \ldots & O \\ O & A_2 & \ldots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \ldots & A_s \end{pmatrix},$$

où les $A_i \in \mathcal{M}_{n_i\times n_i}(\mathbb{K})$ sont des blocs carrés. Alors pour tout polynôme $Q \in \mathbb{K}[X]$, $Q(X) = \sum_i c_i X^i$, on a

$$A^i = \begin{pmatrix} A_1^i & O & \ldots & O \\ O & A_2^i & \ldots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \ldots & A_s^i \end{pmatrix} \;\Rightarrow\; Q(A) = \begin{pmatrix} Q(A_1) & O & \ldots & O \\ O & Q(A_2) & \ldots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \ldots & Q(A_s) \end{pmatrix},$$

donc Q est un polynôme annulateur de A si et seulement si Q est multiple de chacun des polynômes minimaux μ_{A_i}. On en déduit dans ce cas que

$$\mu_A = \mathrm{ppcm}(\mu_{A_1}, \mu_{A_2}, \ldots, \mu_{A_s}).$$

(c) En particulier si A est une matrice diagonale ayant des coefficients diagonaux $\alpha_1, \alpha_2, \ldots, \alpha_n$ deux à deux distincts on a $\mu_A(X) = \chi_A(X) = \prod_{1\leqslant i\leqslant n}(X - \alpha_i)$,

mais si certaines valeurs sont répétées, disons $\alpha_i = \lambda_1$ avec multiplicité m_1, ..., $\alpha_i = \lambda_s$ avec multiplicité m_s (et les λ_j deux à deux distincts), on a

$$\chi_A(X) = \prod_{1\leqslant i\leqslant s} (X-\lambda_i)^{m_i}, \quad \mu_A(X) = \prod_{1\leqslant i\leqslant s} (X-\lambda_i),$$

et donc $\deg \mu_A = s < n = \deg \chi_A$ dès que l'une des multiplicité m_i est > 1.

(d) Dans le cas d'une matrice triangulaire (disons supérieure) par blocs, on voit de même que

$$A = \begin{pmatrix} A_1 & R_{1,2} & \dots & R_{1,s} \\ O & A_2 & \dots & R_{2,s} \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & A_s \end{pmatrix} \quad \Rightarrow \quad Q(A) = \begin{pmatrix} Q(A_1) & * & \dots & * \\ O & Q(A_2) & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & Q(A_s) \end{pmatrix},$$

par conséquent il est nécessaire que Q soit multiple des μ_{A_i} pour avoir $Q(A) = 0$. On en déduit dans ce cas que

$$\mu_A \text{ est multiple de } \operatorname{ppcm}(\mu_{A_1}, \mu_{A_2}, \dots, \mu_{A_s}).$$

(e) Si $f \in \mathrm{End}_{\mathbb{K}}(E)$ est un endomorphisme quelconque, on a $Q(f + \alpha \operatorname{Id}_E) = 0$ si et seulement si $Q(X+\alpha)$ est multiple de $\mu_f(X)$, ce qui équivaut à dire que $Q(X)$ est multiple de $\mu_f(X-\alpha)$, comme on le voit en effectuant la substitution $X \mapsto X - \alpha$. On obtient par conséquent une formule analogue à celle déjà vue pour le polynôme caractéristique :

$$\mu_{f+\alpha \operatorname{Id}_E}(X) = \mu_f(X-\alpha).$$

(f) Considérons l'endomorphisme $h : \mathbb{K}^n \to \mathbb{K}^n$ dont la matrice dans la base canonique $(e_1, \dots, e_n)$ est

$$N = \begin{pmatrix} 0 & 1 & \dots & 0 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 0 \end{pmatrix}.$$

On a donc $h(e_j) = e_{j-1}$ si $j > 1$ et $h(e_1) = 0$. Les puissances h^k sont données par $h^k(e_j) = e_{j-k}$ si $j > k$ et $f^k(e_j) = 0$ si $1 \leqslant j \leqslant k$, de sorte que N^k est la matrice triangulaire supérieure formée de coefficients 1 sur la k-ième diagonale au dessus de la diagonale principale et de 0 ailleurs. On a ainsi en particulier

$$N^{n-1} = \begin{pmatrix} 0 & 0 & \dots & 1 \\ 0 & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 0 \end{pmatrix}, \quad N^n = 0,$$

et on voit que $I_n, N, N^2, \ldots, N^{n-1}$ sont linéairement indépendantes. Il en résulte aussitôt que

$$\mu_h(X) = \mu_N(X) = X^n.$$

(Un tel endomorphisme est appelé endomorphisme nilpotent, voir §2.6 pour l'étude générale de ces endomorphismes). Dans le cas d'un bloc de Jordan triangulaire supérieur

$$J = \begin{pmatrix} \alpha & 1 & \ldots & 0 \\ 0 & \alpha & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 0 & \ldots & 0 & \alpha \end{pmatrix} = N + \alpha I_n,$$

on obtient ainsi

$$\mu_J(X) = \mu_N(X-\alpha) = (X-\alpha)^n = \chi_J(X)$$

Notons que J est ici une matrice triangulaire formée de blocs diagonaux $A_i = (\alpha)$ de taille 1×1, et le polynôme minimal de A_i est (à l'évidence) $\mu_{A_i}(X) = X-\alpha$, mais μ_J ne coïncide pas avec $\mathrm{ppcm}(\mu_{A_1}, \ldots, \mu_{A_n}) = X-\alpha$, il en est seulement multiple. Ceci montre qu'on ne peut pas en général conclure que $\mu_A = \mathrm{ppcm}(\mu_{A_1}, \ldots, \mu_{A_n})$ dans le cas triangulaire par blocs.

(g) Par transposition, il est clair que ${}^tQ(A) = Q({}^tA)$ pour toute matrice A. On en déduit la formule $\mu_{{}^tA} = \mu_A$ (et on a d'ailleurs de même $\chi_{{}^tA} = \chi_A$).

(h) Si A est une matrice diagonale par blocs $A = A_1 \boxplus \cdots \boxplus A_s$ telle que A_i est un bloc de Jordan $J_{n_i;\alpha_i}$ (ou un transposé d'un tel bloc), la combinaison des propriétés précédentes permet de calculer aisément le polynôme minimal

$$\mu_A(X) = \mathrm{ppcm}(\mu_{A_1}, \ldots, \mu_{A_s}) = \mathrm{ppcm}((X-\alpha_1)^{n_1}, \ldots, (X-\alpha_s)^{n_s}),$$

et en particulier $\mu_A(X) = \prod_{i=1}^s (X-\alpha_i)^{n_i}$ si les valeurs propres α_i sont 2 à 2 distinctes. □

2.1.14. Exemple. Considérons

$$A = \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix}, \quad B = \begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 0 & 0 \\ 0 & 0 & \lambda & 0 \\ 0 & 0 & 0 & \mu \end{pmatrix}$$

dans $\mathcal{M}_{4\times 4}(\mathbb{K})$ avec $\lambda \neq \mu$, on vérifie facilement d'après les propriétés qui précèdent (B étant formé de 3 blocs de Jordan de taille 2×2, 1×1, 1×1) que

$$\chi_A(X) = \chi_B(X) = (X-\lambda)^3(X-\mu),$$
$$\mu_A(X) = (X-\lambda)(X-\mu), \quad \mu_B(X) = (X-\lambda)^2(X-\mu).$$ □

Les propriétés et exemples ci-dessus montrent qu'il y a des liens très forts entre le polynôme caractéristique et le polynôme minimal, bien qu'il ne coïncident pas toujours. Voici un autre résultat dans cette direction.

2.1.15. Théorème. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ et $Q \in \mathbb{K}[X]$ un polynôme annulateur, i.e. $Q(f) = 0$. Alors pour toute valeur propre $\lambda \in \mathbb{K}$ de f, on a $Q(\lambda) = 0$.*
En particulier, ceci s'applique au polynôme minimal $Q = \mu_f$, donc $\mu_f(X)$ est divisible par $\prod_{j=1}^{s}(X - \lambda_j)$ où les λ_j sont les valeurs propres distinctes.

Démonstration. Supposons $f(v) = \lambda v$ avec $v \neq 0$, et écrivons $Q(X) = \sum_{i=0}^{d} a_i X^i$. La relation $Q(f) = 0$ implique

$$0 = Q(f)(v) = \sum_{i=0}^{d} a_i f^i(v) = \sum_{i=0}^{d} a_i \lambda^i v = Q(\lambda)\, v \quad \Rightarrow \quad Q(\lambda) = 0. \qquad \square$$

2.2. Endomorphismes cycliques et matrices compagnons

La recherche du polynôme minimal d'un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ peut (au moins en théorie) se faire en calculant les puissances successives $(f^j)_{j\in\mathbb{N}}$ et en déterminant le plus grand entier $d \geqslant 1$ tel que $(1, f, \ldots, f^{d-1})$ soient linéairement indépendants. Ceci amène aux notions voisines de sous-espace cyclique et d'endomorphisme cyclique.

2.2.1. Théorème et définition. *Un sous-espace vectoriel $S \subset E$ est dit cyclique (relativement à l'endomorphisme f) si S est engendré par un vecteur $v_0 \in E$ et ses images successives $f^j(v_0)$, c'est à dire si*

$$S = \mathrm{vect}(f^j(v_0))_{j\in\mathbb{N}},$$

et on dit alors que v_0 est un générateur du sous-espace cyclique S pour f. Tout sous-espace cyclique S relativement à f est stable par f.

Démonstration. Par définition, S consiste en les combinaisons linéaires finies

$$x = \alpha_0 v_0 + \alpha_1 f(v_0) + \cdots + \alpha_j f^j(v_0),$$

de sorte que

$$f(x) = \alpha_0 f(v_0) + \alpha_1 f^2(v_0) + \cdots + \alpha_j f^{j+1}(v_0) \in S.$$

Ceci implique que $f(S) \subset S$. $\square$

Si l'on prend l'entier p maximum tel que $v_0, f(v_0), \ldots, f^{(p-1)}(v_0)$ soient linéairement indépendants, alors par définition de p il existe une relation de dépendance linéaire avec des coeffcients α_i non tous nuls

$$\alpha_0 v_0 + \alpha_1 f(v_0) + \cdots + \alpha_{p-1} f^{p-1}(v_0) + \alpha_p f^p(v_0) = 0.$$

On a nécessairement $\alpha_p \neq 0$, donc $f^p(v_0) = -\dfrac{\alpha_0}{\alpha_p} v_0 - \cdots - \dfrac{\alpha_{p-1}}{\alpha_p} f^{p-1}(v_0)$, et en reprenant l'image par f^{j-p}, $j \geqslant p$, il vient

$$f^j(v_0) = -\frac{\alpha_j}{\alpha_p} f^{j-p}(v_0) - \cdots - \frac{\alpha_{p-1}}{\alpha_p} f^{j-1}(v_0).$$

On voit ainsi par récurrence sur j que tous les $f^j(v_0)$, $j \geqslant p$, sont combinaisons linéaires de $v_0, f(v_0), \ldots, f^{(p-1)}(v_0)$, d'où

$$S = \mathrm{vect}(v_0, f(v_0), \ldots, f^{p-1}(v_0)) \quad \text{et} \quad \dim S = p.$$

On obtient ainsi la conséquence suivante.

2.2.2. Proposition. *Si le polynôme minimal μ_f est de degré $d < n$, alors tout sous-espace cyclique S pour f vérifie* $\dim S \leqslant d < n$.

Démonstration. En effet, si $d = \deg \mu_f$, on a vu que f^d est combinaison linéaire de $\mathrm{Id}_E, f, \ldots, f^{d-1}$, donc pour tout vecteur $v_0 \in E$ et $i \geqslant d$, $f^i(v_0)$ est combinaison linéaire de $v_0, f(v_0), \ldots, f^{d-1}(v_0)$. □

2.2.3. Définition. *On dit que l'endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est cyclique si l'espace E lui-même est cyclique pour f, c'est-à-dire s'il existe $v_0 \in E$ tel que*

$$(v_0, f(v_0), \ldots, f^{n-1}(v_0)) \quad \textit{soit une base de } E.$$

2.2.4. Matrice compagnon et polynôme caractéristique d'un endomorphisme cyclique. Soit f un endomorphisme de E (éventuellement défini par une matrice $A = \mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f)$ dans une base $\mathcal{B} = (e_1, \ldots, e_n)$). On suppose que l'endomorphisme f est cyclique, et on choisit un vecteur $v_0 \in E$ tel que les itérés $(f^i(v_0))_{0 \leqslant i \leqslant n-1}$ définissent une base $\mathcal{B}' = (v_0, f(v_0), \ldots, f^{n-1}(v_0))$ de E. Le vecteur $f^n(v_0)$ est alors combinaison linéaire de $v_0, f(v_0), \ldots, f^{n-1}(v_0)$, ce qu'on écrira sous la forme

$$a_0 v_0 + a_1 f(v_0) + \cdots + a_{n-1} f^{n-1}(v_0) + f^n(v_0) = 0, \quad a_j \in \mathbb{K}.$$

Comme $f(f^i(v_0)) = f^{i+1}(v_0)$ pour $i = 0, 1, \ldots, n-2$ et

$$f(f^{n-1}(v_0)) = f^n(v_0) = -a_0 v_0 - a_1 f(v_0) - \cdots - a_{n-1} f^{n-1}(v_0),$$

la matrice $M = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$ est donnée par

(2.2.5)
$$M = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & -a_{n-2} \\ 0 & 0 & \ldots & 0 & 1 & -a_{n-1} \end{pmatrix}.$$

2.2.6. Définition. *On dit que la matrice M définie par* 2.2.5 *est la "matrice compagnon" du polynôme unitaire*

$$Q(X) = a_0 + a_1 X + \cdots + a_{n-1} X^{n-1} + X^n.$$

Le polynôme Q est choisi en sorte que l'on ait la relation (évidente) $Q(f)(v_0) = 0$. Par commutativité des polynômes d'endomorphismes, on en déduit

$$Q(f)(f^i(v_0)) = (Q(f) \circ f^i)(v_0) = (f^i \circ Q(f))(v_0) = f^i(Q(f)(v_0)) = 0.$$

Comme $\mathcal{B}' = (f^i(v_0))_{0 \leqslant i \leqslant n-1}$ est une base de E, ceci implique $Q(f) = 0$. Pour la même raison, on a $R(f)(v_0) \neq 0$ pour tout polynôme $R \neq 0$ de degré $\deg R \leqslant n-1$, donc $R(f) \neq 0$. On obtient ainsi :

2.2.7. Fait fondamental. *L'endomorphisme f et la matrice compagnon M du polynôme Q ont précisément pour polynôme minimal* $\mu_f = \mu_M = Q$. □

Il nous reste à calculer le polynôme caractéristique de M. Nous avons

$$\chi_M(X) = \det(XI_n - M) = \begin{vmatrix} X & 0 & 0 & \dots & 0 & a_0 \\ -1 & X & 0 & \dots & 0 & a_1 \\ 0 & -1 & X & \dots & 0 & a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -1 & X & a_{n-2} \\ 0 & 0 & \dots & 0 & -1 & X + a_{n-1} \end{vmatrix}.$$

On calcule ce déterminant en développant par rapport à la dernière colonne. Le terme du bas donne une contribution $(X + a_{n-1})X^{n-1}$, tandis que le terme a_i donne une contribution

$$(-1)^{n-1-i}a_i \begin{vmatrix} X & 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ -1 & X & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \dots & -1 & X & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & -1 & X & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 0 & -1 & X & \dots & 0 \\ \vdots & & \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & & -1 & X \\ 0 & \dots & 0 & 0 & 0 & \dots & \dots & 0 & -1 \end{vmatrix}.$$

Le déterminant $(n-1) \times (n-1)$ ci-dessus est formé de deux blocs carrés, le premier triangulaire inférieur de taille $i \times i$ et de coefficients diagonaux égaux à X, le second triangulaire supérieur de taille $(n-1-i) \times (n-1-i)$ et de coefficients diagonaux égaux à -1. Le terme a_i fournit donc une contribution $(-1)^{n-1-i}a_i \times X^i \times (-1)^{n-1-i} = a_i X^i$, d'où

$$\chi_M(X) = \sum_{i=0}^{n-2} a_i X^i + (X + a_{n-1})X^{n-1} = Q(X),$$

de sorte que $\chi_M = \mu_M = Q$. Ce résultat peut se reformuler comme suit :

2.2.8. Théorème. *Si l'endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est cyclique, le polynôme caractéristique χ_f et le polynôme minimal μ_f coïncident : on a*

$$\chi_f = \mu_f = Q,$$

où Q est le polynôme compagnon de la matrice M de f exprimée dans une base $\mathcal{B}' = (v_0, f(v_0), \ldots, f^{n-1}(v_0))$ associée à un vecteur cyclique v_0. Il en résulte que

$$\chi_f(f) = Q(f) = 0, \quad \chi_M(M) = Q(M) = 0.$$

□

2.3. Théorème de Cayley-Hamilton

Ce théorème est attribué au mathématicien britannique Arthur Cayley (1821–1895) et au mathématicien et physicien irlandais William Hamilton (1805–1865), qui ont tous deux éminemment contribué au développement de l'algèbre linéaire. Mais en fait la première démonstration complète en dimension > 3 semble avoir été donnée en 1878 par le mathématicien allemand Ferdinand Georg Frobenius (1849–1917), à qui l'on doit aussi la notion de polynôme minimal.

2.3.1. Théorème (Cayley-Hamilton). *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme d'un espace vectoriel de dimension finie. Alors le polynôme caractéristique χ_f est un polynôme annulateur de f, c'est-à-dire que l'on a* $\chi_f(f) = 0$.

Démonstration. Le résultat a été démontré à la section précédente dans le cas d'un endomorphisme cyclique, et on va voir qu'on peut se ramener à ce cas. Posons $\theta = \chi_f(f) \in \mathrm{End}_{\mathbb{K}}(E)$. Il s'agit de voir que $\theta(v) = 0$ pour tout vecteur $v \in E$. C'est évident si $v = 0$. Si $v \neq 0$, le vecteur v engendre un sous-espace cyclique

$$S = \mathrm{vect}(v, f(v), \ldots, f^i(v), \ldots) \subset E$$

non réduit à $\{0\}$ et stable par f. Par définition, l'endomorphisme induit $f_{|S} : S \to S$ est cyclique, avec les $f^i_{|S}(v)$ qui engendrent S. On déduit donc du théorème 2.2.8 que $\chi_{f_{|S}}(f_{|S}) = 0$. En particulier, comme $v \in S$, on a

$$\chi_{f_{|S}}(f)(v) = \chi_{f_{|S}}(f_{|S})(v) = 0.$$

Mais si on prend un sous-espace $T \subset E$ supplémentaire tel que $E = S \oplus T$, on obtient une décomposition par blocs

$$f = \begin{array}{c} \begin{matrix} S & \quad T \end{matrix} \\ \left(\begin{array}{c|c} f_{|S} & h \\ \hline 0 & g \end{array}\right) \begin{matrix} S \\ T \end{matrix} \end{array} \tag{2.3.2}$$

avec $g = \pi_T \circ f_{|T}$ (cf. théorème 1.3.2), d'où $\chi_f(X) = \chi_g(X)\,\chi_{f_{|S}}(X)$ (produit commutatif). On en déduit

$$\theta(v) = \chi_f(f)(v) = \chi_g(f) \circ \chi_{f_{|S}}(f)(v) = 0.$$

Le théorème est démontré. □

La théorème de Cayley-Hamilton (et la preuve qui en a été donnée) entraînent également la conséquence importante qui suit.

2.3.3. Théorème. *Pour tout endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ d'un espace vectoriel de dimension finie n, le polynôme caractéristique χ_f et le polynôme minimal μ_f satisfont les relations de divisibilité mutuelles*

$$\mu_f \mid \chi_f \mid (\mu_f)^n,$$

en particulier $\deg \mu_f \leqslant \deg \chi_f = n = \dim E$.

Démonstration. (a) La relation $\mu_f \mid \chi_f$ est une conséquence directe du théorème de Cayley-Hamilton, et on en déduit bien $\deg \mu_f \leqslant n$.

(b) On vérifie maintenant que $\chi_f \mid (\mu_f)^n$ par récurrence sur $n = \dim E$. Si $n = 1$, f a pour matrice $A = (\alpha)$ et on a $\chi_A(X) = \mu_A(X) = X - \alpha$. Supposons le résultat démontré pour tout endomorphisme d'un espace E' de dimension $\dim E' < n$. On utilise alors la décomposition (2.3.2) dans laquelle $f_{|S}$ est cyclique, ce qui nous permet déjà de conclure que $\chi_{f_{|S}} = \mu_{f_{|S}}$ par le théorème 2.2.8. Si $T = \{0\}$, il vient $S = E$ et le résultat s'ensuit (on a même $\chi_f = \mu_f$). Sinon, on est dans le cas où $0 < \dim T < n = \dim E$ et on peut appliquer l'hypothèse de récurrence à $E' = T$ et $g \in \mathrm{End}_{\mathbb{K}}(T)$, ce qui donne $\chi_g \mid (\mu_g)^{\dim T}$, d'où

$$\chi_f = \chi_{f_{|S}}\, \chi_g \quad \Rightarrow \quad \chi_f \mid \mu_{f_{|S}}\, (\mu_g)^{\dim T}.$$

Mais $\mu_{f_{|S}}$ et μ_g divisent tous deux μ_f (propriété 2.1.13 (d)), donc

$$\chi_f \mid (\mu_f)^{1+\dim T} \mid (\mu_f)^n. \qquad \square$$

2.3.4. Remarque. Il résulte du théorème 2.3.3 que χ_f et μ_f ont toujours les mêmes racines (avec des multiplicités éventuellement différentes). Dans le cas $f = \alpha\,\mathrm{Id}_E$, on a précisément $\mu_f(X) = X - \alpha$ et $\chi_f(X) = (X-\alpha)^n = \mu_f(X)^n$, donc la relation $\chi_f \mid (\mu_f)^n$ ne peut pas être améliorée.

2.4. Lemme des noyaux et décomposition par blocs

On utilise ici de manière essentielle le fait que l'anneau $\mathbb{K}[X]$ est un anneau principal (et donc factoriel).

2.4.1. Lemme des noyaux. *Soient $Q_1, \ldots, Q_s \in \mathbb{K}[X]$ des polynômes premiers entre eux deux à deux et $f \in \mathrm{End}_{\mathbb{K}}(E)$. Alors*

$$\mathrm{Ker}(Q_1 \ldots Q_s(f)) = \bigoplus_{j=1}^{s} \mathrm{Ker}(Q_j(f)).$$

Démonstration. Commençons par le cas $s = 2$ (le résultat étant vide si $s = 1$). Par hypothèse $\mathrm{pgcd}(Q_1, Q_2) = 1$; d'après l'identité de Bézout, il existe des polynômes

R_1, R_2 tels que $R_1Q_1 + R_2Q_2 = 1$, et donc $R_1(f)\circ Q_1(f) + R_2(f)\circ Q_2(f) = \mathrm{Id}_E$. Si $v \in \mathrm{Ker}(Q_1(f)) \cap \mathrm{Ker}(Q_2(f))$, on obtient

$$v = \mathrm{Id}_E(v) = R_1(f)\circ Q_1(f)(v) + R_2(f)\circ Q_2(f)(v) = 0,$$

ce qui montre que $\mathrm{Ker}(Q_1(f)) \cap \mathrm{Ker}(Q_2(f)) = \{0\}$, et nos deux noyaux sont bien en somme directe. D'autre part, comme

$$(Q_1Q_2)(f) = Q_1(f)\circ Q_2(f) = Q_2(f)\circ Q_1(f),$$

il est clair que

$$\mathrm{Ker}(Q_1(f)) \subset \mathrm{Ker}(Q_1Q_2(f)) \quad \text{et} \quad \mathrm{Ker}(Q_2(f)) \subset \mathrm{Ker}(Q_1Q_2(f)),$$

donc

$$\mathrm{Ker}(Q_1(f)) \oplus \mathrm{Ker}(Q_2(f)) \subset \mathrm{Ker}(Q_1Q_2(f)).$$

Maintenant, si $v \in \mathrm{Ker}(Q_1Q_2(f)$, on peut écrire comme ci-dessus

$$v = \mathrm{Id}_E(v) = R_1(f)\circ Q_1(f)(v) + R_2(f)\circ Q_2(f)(v)$$

et on a par commutativité

$$Q_2(f)\circ R_1(f)\circ Q_1(f)(v) = R_1(f)\circ Q_2(f)\circ Q_1(f)(v) = R_1(f)\circ (Q_1Q_2)(f)(v) = 0,$$

donc $w_2 = R_1(f)\circ Q_1(f)(v) \in \mathrm{Ker}(Q_2(f))$, et on vérifie de la même manière que $w_1 = R_2(f)\circ Q_2(f)(v) \in \mathrm{Ker}(Q_1(f))$. Ceci implique

$$v = w_1 + w_2 \in \mathrm{Ker}(Q_1(f)) \oplus \mathrm{Ker}(Q_2(f)),$$

donc $\mathrm{Ker}((Q_1Q_2)(f)) \subset \mathrm{Ker}(Q_1(f)) \oplus \mathrm{Ker}(Q_2(f))$ et l'égalité s'ensuit pour $s = 2$. Le cas général s'en déduit par récurrence sur $s \geqslant 3$ en observant que

$$Q_1 \ldots Q_{s-2}Q_{s-1}Q_s = Q_1 \ldots Q_{s-2}(Q_{s-1}Q_s),$$

avec $\mathrm{pgcd}(Q_i, Q_{s-1}Q_s) = 1$ par le lemme de Gauss. □

Le lemme des noyaux a plusieurs conséquences importantes, notamment les théorèmes de décomposition fondamentaux que nous étudierons dans les sections suivantes. En voici le point de départ.

2.4.2. Théorème de décomposition par blocs. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme quelconque.*

(a) *On a dans $\mathbb{K}[X]$ des décompositions en polynômes irréductibles*

$$\chi_f(X) = \prod_{j=1}^{s} P_j(X)^{m_j}, \quad \mu_f(X) = \prod_{j=1}^{s} P_j(X)^{m'_j} \quad \textit{avec } 1 \leqslant m'_j \leqslant m_j \leqslant n.$$

En particulier χ_f et μ_f ont le même ensemble de racines, et si les polynômes χ_f et μ_f sont scindés sur $\mathbb{K}$, on a

$$\chi_f(X) = \prod_{j=1}^{s} (X-\lambda_j)^{m_j}, \quad \mu_f(X) = \prod_{j=1}^{s} (X-\lambda_j)^{m'_j} \quad \textit{avec } 1 \leqslant m'_j \leqslant m_j \leqslant n\,,$$

où $\lambda_1, \ldots, \lambda_s \in \mathbb{K}$ sont les valeurs propres distinctes.

(b) *On a une décomposition en somme directe*

$$E = V_1 \oplus \cdots \oplus V_s \quad \text{où} \quad V_j = \operatorname{Ker} P_j(f)^{m'_j} = \operatorname{Ker} P_j(f)^{k_j}, \quad \forall k_j \geqslant m'_j.$$

Les sous-espaces V_j sont stables par f, et les endomorphismes induits par restriction $f_j = f_{|V_j} : V_j \to V_j$ fournissent relativement à des bases des V_j et la base correspondante $\mathcal{B}'$ de E une décomposition en blocs de la matrice A' de f :

$$f = f_1 \boxplus \cdots \boxplus f_s, \quad A' = \begin{pmatrix} A'_1 & \dots & O \\ \vdots & \ddots & \vdots \\ O & \dots & A'_s \end{pmatrix} \quad \text{où} \quad A'_j = \operatorname{Mat}(f_j).$$

(c) *Pour tout $j = 1, \dots, s$, l'endomorphisme $f_j \in \operatorname{End}_{\mathbb{K}}(V_j)$ est tel que*

$$\chi_{f_j} = P_j^{m_j}, \quad \mu_{f_j} = P_j^{m'_j}, \quad \chi_f = \prod_{j=1}^{s} \chi_{f_j}, \quad \mu_f = \prod_{j=1}^{s} \mu_{f_j}.$$

(d) *Il existe un polynôme $\Pi_j \in \mathbb{K}[X]$ tel que*

$$\Pi_j(f) = \pi_j = \textit{projection de } E \textit{ sur } V_j \textit{ parallèlement à } \bigoplus_{i \neq j} V_i.$$

Démonstration. (a) On sait que μ_f divise χ_f et que χ_f divise μ_f^n, donc ces polynômes ont les mêmes facteurs irréductibles P_j, avec des multiplicités éventuellement différentes. Le fait que $\mu_f \mid \chi_f$ implique $m'_j \leqslant m_j$, et le fait que $\deg \chi_f = n$ implique $m_j \leqslant n$. L'assertion relative au cas scindé est une conséquence immédiate.

(b) Nous avons par définition $\mu_f(f) = 0$, et donc aussi $Q(f) = 0$ pour tout multiple $Q = \prod_{j=1}^{s} P_j^{k_j}$ avec $k_j \geqslant m'_j$. Comme les polynômes $P_j^{k_j}$ sont premiers entre eux deux à deux, le lemme des noyaux implique

$$E = \operatorname{Ker} \mu_f(f) = \bigoplus_{j=1}^{s} V_j, \quad V_j = \operatorname{Ker} P_j(f)^{m'_j},$$

$$E = \operatorname{Ker} \mu_f(f) = \bigoplus_{j=1}^{s} V'_j, \quad V'_j = \operatorname{Ker} P_j(f)^{k_j}.$$

Or, pour tout $v \in E$,

$$P_j(f)^{m'_j}(v) = 0 \quad \Rightarrow \quad P_j(f)^{k_j}(v) = P_j(f)^{k_j - m'_j} \circ P_j(f)^{m'_j}(v) = 0,$$

donc $V_j \subset V'_j$, et on en déduit que l'on a nécessairement $V_j = V'_j$ puisque $\dim E = \sum \dim V_j = \sum \dim V'_j$. On sait de plus par le corollaire 2.1.8 que les sous-espaces V_j sont stables par f. Les affirmations du (b) en découlent.

(c) Comme $P_j(f_j)^{m'_j} = P_j(f)^{m'_j}_{|V_j} = 0$ par définition de V_j, on voit que μ_{f_j} divise $P_j^{m'_j}$. En particulier les polynômes μ_{f_j} sont premiers entre eux deux à deux, et la propriété 2.1.13 (b) donne

$$\mu_f = \mathrm{ppcm}(\mu_{f_1}, \ldots, \mu_{f_s}) = \prod_{j=1}^{s} \mu_{f_j}.$$

La seule possibilité (par unicité de la factorisation) est que $\mu_{f_j} = P_j^{m'_j}$. Maintenant, on sait aussi que $\chi_f = \prod \chi_{f_j}$, et comme χ_{f_j} divise $\mu_{f_j}^{n_j} = P_j^{n_j m'_j}$ où $n_j = \dim V_j$, on voit que χ_{f_j} est une puissance de P_j. La seule décomposition possible de χ_f est par suite $\chi_{f_j} = P_j^{m_j}$. On trouve ainsi que $n_j = \deg \chi_{f_j} = m_j \deg P_j$.

(d) D'après le théorème des restes chinois appliqué dans l'anneau principal $\mathbb{K}[X]$, il existe un polynôme Π_j tel que $\Pi_j \equiv 1 \bmod \langle P_j^{m'_j} \rangle$ et $\Pi_j \equiv 0 \bmod \langle P_i^{m'_i} \rangle$ pour $i \neq j$. Il est facile de voir que $\pi_j = \Pi_j(f) \in \mathrm{End}_{\mathbb{K}}(E)$ est la projection de E sur V_j dans la décomposition $E = V_1 \oplus \cdots \oplus V_s$. En effet, l'unicité des solutions du théorème des restes chinois montre que

$$\Pi_j^2 - \Pi_j \equiv 0 \bmod \langle P_1^{m'_1} \cdots P_s^{m'_s} \rangle = \langle \mu_f \rangle$$

donc $\pi_j^2 - \pi_j = 0$ et π_j est un projecteur ; d'autre part (et par définition), Π_j est divisible par $\prod_{i \neq j} P_i^{m'_i}$ et $\Pi_j - 1$ est divisible par $P_j^{m'_j}$, ce qui implique de nouveau par le lemme des noyaux que

$$\mathrm{Ker}(\pi_j) \supset \bigoplus_{i \neq j} \mathrm{Ker}\, P_i^{m'_i}(f) = \bigoplus_{i \neq j} V_i \quad \text{et} \quad \mathrm{Ker}(\pi_j - \mathrm{Id}_E) \supset \mathrm{Ker}\, P_j^{m'_j}(f) = V_j. \qquad \square$$

2.4.3. Corollaire. *Un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est diagonalisable si et seulement si le polynôme minimal μ_f est scindé et à racines simples, i.e. $\mu_f(X) = \prod_{j=1}^{s}(X - \lambda_j)$ où les λ_j sont les valeurs propres distinctes.*

Démonstration. Si f est diagonalisable, le fait que μ_f soit scindé à racines simples résulte de la discussion 2.1.13 (c). Réciproquement, si $\mu_f(X) = \prod_{j=1}^{s}(X - \lambda_j)$ est scindé à racines simples, le théorème de décomposition par blocs appliqué à $P_j(X) = X - \lambda_j$, montre que les espaces propres $V_j = \mathrm{Ker}(f - \lambda_j \mathrm{Id}_E)$ forment une décomposition en somme directe $E = V_1 \oplus \cdots \oplus V_s$, donc f est diagonalisable. $\square$

Une autre application est la caractérisation des endomorphismes cycliques (la démonstration est un peu plus délicate, réservée aux étudiants aventureux !).

2.4.4. Théorème. *Un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est cyclique si et seulement si $\deg \mu_f = n = \dim E$, autrement dit, si et seulement si $\mu_f = \chi_f$.*

Démonstration.* Si f est cyclique, on ne peut avoir $d = \deg \mu_f < n$, sinon on en conclurait par la proposition 2.2.2 que tout sous-espace cyclique S est de dimension $\dim S \leqslant d$, ce qui est contradictoire pour $S = E$; par suite $\deg \mu_f = n$.

C'est la réciproque qui est le point délicat. Supposons $\deg \mu_f = n$, et soit

$$\mu_f(X) = \chi_f(X) = \prod_{j=1}^{s} P_j(X)^{m_j}, \quad m'_j = m_j \geqslant 1$$

la décomposition de $\mu_f(X)$ en facteurs irréductibles. On considère la somme directe $E = V_1 \oplus \cdots \oplus V_s$ donnée par le théorème 2.4.2 (b), et les restrictions $f_j = f_{|V_j}$. Le résultat 2.4.2 (c) montre que $\mu_{f_j} = \chi_{f_j} = P_j^{m_j}$, et on va d'abord montrer que f_j est cyclique. Comme $P_j(f_j)^{m_j-1} \neq 0$, il existe un vecteur $v_j \in V_j$ tel que $P_j(f_j)^{m_j-1}(v_j) \neq 0$. Soit $W_j \subset V_j$ le sous-espace cyclique engendré par les $f_j^i(v_j) = f^i(v_j)$. Alors W_j est stable par f_j, et le polynôme minimal de $f_{|W_j}$ divise nécessairement $P_j^{m_j}$. Mais comme P_j est irréductible, on en conclut que $\mu_{f_{|W_j}} = P_j^{k_j}$ avec $k_j \leqslant m_j$. Le fait que $P_j(f_j)^{m_j-1}(v_j) \neq 0$ implique que l'on a nécessairement $k_j = m_j$. Il s'ensuit $\dim W_j \geqslant \deg \mu_{f_{|W_j}} = \deg \chi_{f_j} = \dim V_j$, donc $W_j = V_j$, et f_j est bien cyclique, avec $V_j = \mathrm{vect}(f^i(v_j))$. Pour terminer, on montre que f est cyclique, avec $E = \mathrm{vect}(f^i(v))_{i\geqslant 0}$ où $v = v_1 + \cdots + v_s$. Grâce à 2.4.2 (d), on a

$$f^i(v_j) = f^i(\pi_j(v)) = f^i \circ \Pi_j(f)(v) = Q_{i,j}(f)(v) \quad \text{avec} \quad Q_{i,j}(X) = X^i \Pi_j(X),$$

ce qui montre que $\mathrm{vect}(f^i(v))_{i\geqslant 0} = \{Q(f)(v) \,/\, Q \in \mathbb{K}[X]\}$ contient la somme $V_1 \oplus \cdots \oplus V_s = E$. Par conséquent f est cyclique. □

2.5. Sous-espaces caractéristiques

Si $\lambda_j \in \mathbb{K}$ est valeur propre d'un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$, $P_j(X) = X - \lambda_j$ est l'un des facteurs irréductibles des décompositions de $\chi_f(X)$ et $\mu_f(X)$. Ceci amène à la définition suivante.

2.5.1. Définition. *Si $\lambda_j \in \mathbb{K}$ est une valeur propre de $f \in \mathrm{End}_{\mathbb{K}}(E)$, on définit le* sous-espace caractéristique *associé à λ_j comme étant*

$$C_{f,\lambda_j} = \mathrm{Ker}(f - \lambda_j \,\mathrm{Id}_E)^{m'_j}$$

où m'_j est la multiplicité de $X - \lambda_j$ dans $\mu_f(X)$.

2.5.2. Proposition. *$V_j = C_{f,\lambda_j}$ est un sous-espace stable par f, et la restriction $f_j = f_{|V_j} \in \mathrm{End}_{\mathbb{K}}(V_j)$ vérifie*

$$(f_j - \lambda_j \,\mathrm{Id}_E)^{m'_j} = 0.$$

2.5.3. Remarque. On sait d'après 2.4.2 (b) que l'on a en fait l'égalité $C_{f,\lambda_j} = \mathrm{Ker}(f - \lambda_j \,\mathrm{Id}_E)^k$ pour tout $k \geqslant m'_j$, et il est souvent plus facile de calculer C_{f,λ_j} par la formule

$$C_{f,\lambda_j} = \mathrm{Ker}(f - \lambda_j \,\mathrm{Id}_E)^{m_j}$$

où m_j est la multiplicité de $X-\lambda_j$ dans χ_f, car χ_f est en général plus directement accessible que μ_f. □

L'observation suivante est également utile pour étudier la structure de f. La démonstration en sera donnée à la section suivante.

2.5.4. Proposition. *On a des sous-espaces stables emboîtés*

$$E_{f,\lambda_j} = \mathrm{Ker}(f-\lambda_j\,\mathrm{Id}_E) \subsetneq \mathrm{Ker}(f-\lambda_j\,\mathrm{Id}_E)^2 \subsetneq \cdots \subsetneq \mathrm{Ker}(f-\lambda_j\,\mathrm{Id}_E)^{m'_j} = C_{f,\lambda_j}$$

et ensuite les noyaux des puissances ultérieures n'augmentent plus. Lorsque $m'_j = 1$, le sous-espace propre E_{f,λ_j} coïncide avec le sous-espace caractéristique C_{f,λ_j}.

Comme conséquence du théorème général 2.4.2, on obtient en particulier le théorème de décomposition suivant, toujours applicable si $\mathbb{K}=\mathbb{C}$:

2.5.5. Théorème. *Si*

$$\chi_f(X) = \prod_{j=1}^{s}(X-\lambda_j)^{m_j}, \quad \mu_f(X) = \prod_{j=1}^{s}(X-\lambda_j)^{m'_j}$$

sont scindés sur $\mathbb{K}$, on a une décomposition par blocs associés aux sous-espaces caractéristiques

$$E = C_{f,\lambda_1} \oplus \cdots \oplus C_{f,\lambda_s}, \quad f = f_1 \boxplus \cdots \boxplus f_s,$$

et on a $(f_j-\lambda_j\,\mathrm{Id})^{m'_j}=0$ pour chaque bloc.

2.6. Endomorphismes nilpotents

Les résultats précédents montrent qu'il est crucial d'étudier les endomorphismes ayant une puissance nulle.

2.6.1. Définition. *Un endomorphisme $h \in \mathrm{End}_{\mathbb{K}}(E)$ est dit nilpotent s'il existe un exposant $k \in \mathbb{N}^*$ tel que $h^k = 0$.*

Il revient au même de dire que le polynôme minimal est de la forme $\mu_h(X) = X^{m'}$ avec $m' \in \mathbb{N}^*$. On sait que l'on a toujours $\deg \mu_h = m' \leqslant n$, donc si h est nilpotent, la plus petite puissance qui soit nulle est $h^{m'} = 0$ avec $m' \leqslant n$.

2.6.2. Exemple. Toute matrice $A \in \mathcal{M}_{n\times n}(\mathbb{K})$ strictement triangulaire (i.e. avec des coefficients diagonaux nuls) est nilpotente. Ainsi, si $A = (a_{ij})_{1\leqslant i,j\leqslant n}$ est strictement triangulaire inférieure (avec $a_{ij}=0$ pour $i \leqslant j$), soit

$$A = \begin{pmatrix} 0 & 0 & \ldots & 0 & 0 \\ a_{21} & 0 & \ldots & 0 & 0 \\ \vdots & \ddots & \ldots & 0 & 0 \\ \vdots & a_{ij} & \ddots & 0 & 0 \\ a_{n1} & a_{n2} & \ldots & a_{nn-1} & 0 \end{pmatrix} \quad i>j,$$

on voit que l'endomorphisme associé $h : \mathbb{K}^n \to \mathbb{K}^n$, relativement à la base canonique $(e_j)_{1\leqslant j\leqslant n}$ de $\mathbb{K}^n$, satisfait

$$h(e_j) \in \mathrm{Vect}(e_{j+1},\dots,e_n), \quad h(\mathrm{Vect}(e_j,\dots,e_n)) \subset \mathrm{Vect}(e_{j+1},\dots,e_n).$$

Par récurrence sur i, on en déduit

$$h^i(\mathrm{Vect}(e_1,\dots,e_n)) \subset \mathrm{Vect}(e_{i+1},\dots,e_n).$$

Ceci entraîne que h^i a une matrice triangulaire dont la diagonale principale et les $i-1$ diagonales situées en dessous sont nulles, et en particulier $h^n = 0$. De même, toute matrice strictement triangulaire supérieure est nilpotente. □

De manière générale, on a le résultat suivant qui implique la proposition 2.5.4 ci-dessus, en prenant $h = f_j - \lambda_j \mathrm{Id}_{V_j}$ sur l'espace $V_j = C_{f,\lambda_j}$.

2.6.3. Théorème. *Si $h \in \mathrm{End}_{\mathbb{K}}(E)$ est nilpotent d'exposant $m' = \deg \mu_h$, on a des noyaux emboîtés*

$$\{0\} = \mathrm{Ker}\, h^0 \subsetneq \mathrm{Ker}\, h \subsetneq \mathrm{Ker}\, h^2 \subsetneq \cdots \subsetneq \mathrm{Ker}\, h^{m'} = E,$$

et les sous-espaces $S_k = \mathrm{Ker}\, h^k$ vérifient

$$h(S_k) \subset S_{k-1} \subsetneq S_k \quad \textit{pour } k \geqslant 1.$$

Démonstration. Comme $h^0 = \mathrm{Id}_E$ et $h^{m'} = 0$, on a bien $S_0 = \mathrm{Ker}\, h^0 = \{0\}$ et $S_{m'} = \mathrm{Ker}\, h^{m'} = E$. Comme par hypothèse m' est l'exposant minimal, on a $h^{m'-1} \neq 0$ et donc $S_{m'-1} = \mathrm{Ker}\, h^{m'-1} \neq S_{m'} = E$. D'autre part, l'implication évidente

$$h^k(v) = 0 \quad \Rightarrow \quad h^{k+1}(v) = 0,$$

montre que $S_k \subset S_{k+1}$, et comme $h^k(v) = h^{k-1}(h(v))$, on voit aussi que $h(S_k) \subset S_{k-1}$. L'égalité $S_k = S_{k+1}$ signifierait que l'on aurait une équivalence

$$\forall v \in E, \quad h^k(v) = 0 \quad \Leftrightarrow \quad h^{k+1}(v) = 0,$$

et en prenant $v = h(w)$, on en déduirait alors $S_{k+1} = S_{k+2}$, de sorte que $S_\ell = S_k$ pour tout $\ell \geqslant k$. Comme $S_{m'-1} \subsetneq S_{m'} = E$, on a donc bien aussi $S_k \subsetneq S_{k+1}$ pour $k \leqslant m'-1$. □

2.6.4. Endomorphismes nilpotents cycliques. Conformément aux définition déjà données, un endomorphisme $h \in \mathrm{End}_{\mathbb{K}}(E)$ est nilpotent et cyclique s'il est nilpotent et s'il existe un vecteur v_0 tel que $\mathcal{B}' = (v_0, h(v_0), \dots, h^{n-1}(v_0))$ soit une base de E. Dans ce cas le polynôme minimal est $\mu_h(X) = X^n$ et la matrice de h dans $\mathcal{B}'$ est la matrice $A' \in \mathrm{Mat}_{n\times n}(\mathbb{K})$ donnée par

$$A' = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \dots & 1 & 0 \end{pmatrix},$$

qui n'est autre que la matrice compagnon du polynôme $Q(X) = X^n$. Une telle matrice est appelée *bloc de Jordan* nilpotent cyclique de taille n. Dans ce cas, $S_k = \operatorname{Ker} h^k$ est l'espace de dimension k

$$S_k = \operatorname{vect}(h^{n-k}(v_0), h^{n-k+1}(v_0), \ldots, h^{n-1}(v_0)),$$

et on voit facilement que $(A')^k$ est la matrice dont la diagonale de coefficients 1 est la k-ième diagonale sous la diagonale principale, par exemple

$$(A')^k = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 \\ 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & \ddots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & 0 & 0 \\ 0 & 0 & \ldots & 1 & 0 & 0 \end{pmatrix} \quad \text{pour } k = 2,$$

le noyau étant formé par les k derniers vecteurs de $\mathcal{B}'$ (qui donnent les k colonnes de 0 à droite). Remarquons aussi, ce qui nous sera utile plus loin, que

$$\text{(2.6.5)} \qquad \operatorname{Im} h^k = \operatorname{vect}(h^k(v_0), h^{k+1}(v_0), \ldots, h^{n-1}(v_0)) = S_{n-k} = \operatorname{Ker} h^{n-k}.$$

Remarquons enfin qu'il est équivalent de travailler dans la base $\mathcal{B}''$ obtenue en renversant l'ordre des vecteurs, soit $\mathcal{B}'' = (f^{n-1}(v_0), f^{n-2}(v_0), \ldots, f(v_0), v_0)$, et que la matrice de f dans la base $\mathcal{B}''$ est alors la matrice triangulaire supérieure

$$A'' = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 1 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 1 \\ 0 & 0 & 0 & \ldots & 0 & 0 \end{pmatrix}. \qquad \square$$

Le résultat suivant élucide entièrement la structure des endomorphismes nilpotents, à partir de celle déjà décrite des endomorphismes nilpotents cycliques.

2.6.6. Théorème. *Soit $h \in \operatorname{End}_{\mathbb{K}}(E)$ un endomorphisme nilpotent de polynôme minimal $\mu_h(X) = X^{m'}$ sur un $\mathbb{K}$ espace vectoriel E de dimension n. Alors il existe des sous-espaces cycliques $W_j = \operatorname{vect}(v_j, h(v_j), \ldots, h^{n_j-1}(v_j))$ de dimensions respectives décroissantes*

$$n_1 \geqslant \cdots \geqslant n_s, \quad \text{où} \quad n_1 = m' \quad \text{et} \quad \sum_i n_i = n,$$

fournissant une décomposition par blocs

$$E = W_1 \oplus \cdots \oplus W_s, \quad h = h_1 \boxplus \cdots \boxplus h_s,$$

avec des endomorphismes nilpotents cycliques $h_1 \in \operatorname{End}_{\mathbb{K}}(W_1), \ldots, h_s \in \operatorname{End}_{\mathbb{K}}(W_s)$.

2.6.7. Interprétation du théorème. Tout endomorphisme nilpotent peut être représenté dans une base convenable, au choix, par une matrice A' triangulaire inférieure, ou une matrice A'' triangulaire supérieure, de la forme

$$A' = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 \\ \alpha_1 & 0 & 0 & \dots & 0 & 0 \\ 0 & \alpha_2 & 0 & \dots & 0 & 0 \\ 0 & 0 & \alpha_3 & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & 0 \\ 0 & 0 & 0 & \dots & \alpha_{n-1} & 0 \end{pmatrix}, \quad A'' = \begin{pmatrix} 0 & \alpha_1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \alpha_2 & \dots & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \ddots & \alpha_{n-2} & 0 \\ \vdots & \vdots & \vdots & \ddots & 0 & \alpha_{n-1} \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix},$$

avec une diagonale de coefficients $\alpha_j = 0$ ou 1, formée d'une suite de $n_1 - 1$ chiffres 1, d'un 0, puis de $n_2 - 1$ chiffres 1, d'un 0, ..., de $n_s - 1$ chiffres 1. Ci-dessous, par exemple, une matrice nilpotente 7×7 triangulaire inférieure, formée d'un bloc A_1 cyclique de taille 4, d'un bloc A_2 cyclique de taille 3 et d'un bloc cyclique A_3 de taille 1 :

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

On a $A_1^4 = 0$, $A_2^3 = 0$, $A_3^1 = 0$, et donc $A^4 = 0$, $\mu_A(X) = X^4$, $\chi_A(X) = X^8$. Si $(e_1, \dots, e_8)$ désigne la base canonique des vecteurs colonnes, le sous-espace propre $\operatorname{Ker} A = E_{A,0}$ est égal à $\operatorname{vect}(e_4, e_7, e_8)$. □

Démonstration du théorème 2.6.6**. On raisonne par récurrence sur $n = \dim E$, le théorème étant trivial si $n = 1$. On suppose le théorème démontré pour toute dimension $n' < n$ et on va le démontrer pour $\dim E = n$ (récurrence "forte").

Posons $n_1 = m'$. Puisque $\mu_h(X) = X^{n_1}$, on a par hypothèse $h^{n_1} = 0$ et $h^{n_1-1} \neq 0$, donc il existe un vecteur $v_1 \in E$ tel que $h^{n_1-1}(v_1) \neq 0$. On considère le sous-espace cyclique

$$W_1 = \operatorname{vect}(v_1, h(v_1), \dots, h^{n_1-1}(v_1)).$$

Il est à noter que les vecteurs ci-dessus sont nécessairement linéairement indépendants, sinon on aurait un polynôme Q de degré $< n_1$ tel que $Q(h)(v_1) = 0$, et donc $Q(h_{|W_1}) = 0$, mais $\mu_{h_{|W_1}}$ doit diviser X^{n_1} et est en fait forcément égal à X^{n_1} en vertu de notre choix $h^{n_1-1}(v_1) \neq 0$. Par suite $\dim W_1 = n_1$ et $h_1 = h_{|W_1}$ est nilpotent cyclique. Soit $\widetilde{E}$ un supplémentaire de W_1 de sorte que $E = W_1 \oplus \widetilde{E}$, soient $\pi_1, \tilde{\pi}$ les projections de E sur W_1 et $\widetilde{E}$ respectivement, et $h = \pi_1 \circ h_{|\widetilde{E}}$,

$\tilde h = \tilde\pi \circ h_{|\widetilde E}$. Ceci donne une décomposition triangulaire par blocs de h

$$h = \begin{array}{c} \begin{matrix} W_1 & \widetilde E \end{matrix} \\ \left(\begin{array}{c|c} h_1 & g \\ \hline O & \tilde h \end{array}\right) \begin{matrix} W_1 \\ \widetilde E \end{matrix} \end{array}$$

(qu'on peut interpréter si on veut comme une décomposition matricielle dans des bases fixées de W_1 et de $\widetilde E$). Pour tout entier $k \in \mathbb{N}$, on a

$$(*) \qquad h^k = \begin{array}{c} \begin{matrix} W_1 & \widetilde E \end{matrix} \\ \left(\begin{array}{c|c} h_1^k & ? \\ \hline O & \tilde h^k \end{array}\right) \begin{matrix} W_1 \\ \widetilde E \end{matrix} \end{array}$$

et on en déduit en particulier que $\tilde h^{m'} = 0$, d'où $\mu_{\tilde h}(X) = X^{n_2}$ avec $n_2 \leqslant n_1 = m'$. D'après l'hypothèse de récurrence appliquée à l'endomorphisme $\tilde h \in \mathrm{End}_{\mathbb{K}}(\widetilde E)$ avec $\tilde n = \dim \widetilde E = n - n_1$, $\tilde n < n$, nous obtenons des sous-espaces cycliques

$$\widetilde W_j = \mathrm{vect}(\tilde v_j, \tilde h(\tilde v_j), \ldots, \tilde h^{\,n_j-1}(\tilde v_j)) \subset \widetilde E, \quad j = 2, \ldots, s,$$

où $\dim \widetilde W_j = n_j$, $n_2 \geqslant \cdots \geqslant n_s$, $\sum_{j=2}^s n_j = \tilde n = n - n_1$, et une décomposition par blocs

$$\widetilde E = \widetilde W_2 \oplus \cdots \oplus \widetilde W_s, \quad \tilde h = \tilde h_2 \boxplus \cdots \boxplus \tilde h_s$$

avec des endomorphismes $\tilde h_j \in \mathrm{End}_{\mathbb{K}}(\widetilde W_j)$ nilpotents cycliques. Ceci ne démontre pas encore le théorème à cause de la présence de l'endomorphisme g dans h et du terme "?" dans h^k, qu'il nous faut éliminer. Or, par hypothèse de récurrence, on a $\tilde h^{\,n_j}(\tilde v_j) = 0$, c'est-à-dire $h^{n_j}(\tilde v_j) \in W_1$ d'après $(*)$. Comme $h^{n_1} = 0$, on obtient $h^{n_1-n_j}(h^{n_j}(\tilde v_j)) = 0$, i.e. $h^{n_j}(\tilde v_j) \in \mathrm{Ker}\, h_1^{n_1-n_j}$. Mais on sait d'après (2.6.5) que $\mathrm{Ker}\, h_1^{n_1-n_j} = \mathrm{Im}\, h_1^{n_j}$, donc il existe un vecteur $t_j \in W_1$ tel que

$$h^{n_j}(\tilde v_j) = h_1^{n_j}(t_j) = h^{n_j}(t_j).$$

On pose maintenant $v_j = \tilde v_j - t_j$. On observe que l'on a par construction $h^{n_j}(v_j) = 0$, et on considère alors les sous-espaces cycliques

$$W_j = \mathrm{vect}(v_j, h(v_j), \ldots, h^{n_j-1}(v_j)).$$

Grâce à $(*)$, les projections de $h^k(v_j)$ et $\tilde h^k(\tilde v_j)$ sur $\widetilde E$ coïncident, et la matrice de passage de la base

$$\widetilde{\mathcal{B}} = \big(v_1, h(v_1), \ldots, h^{n_1-1}(v_1), \tilde v_2, \tilde h(\tilde v_2), \ldots, \tilde h^{n_2-1}(\tilde v_2), \ldots, \tilde v_s, \tilde h(\tilde v_s), \ldots, \tilde h^{n_s-1}(\tilde v_s)\big)$$

au système de vecteurs

$$\mathcal{B}' = \big(v_1, h(v_1), \ldots, h^{n_1-1}(v_1), v_2, h(v_2), \ldots, h^{n_2-1}(v_2), \ldots, v_s, h(v_s), \ldots, \tilde h^{n_s-1}(v_s)\big)$$

est triangulaire supérieure à coefficients diagonaux égaux à 1, donc $\mathcal{B}'$ est une base, et on obtient une décomposition en somme directe

$$E = W_1 \oplus \cdots \oplus W_s.$$

Il est clair par construction que h admet la décomposition par blocs voulue

$$h = h_1 \boxplus \cdots \boxplus h_s \quad \text{où} \quad h_j = h_{|W_j}. \qquad \square$$

2.7. Triangularisation et réduction de Jordan

Cette réduction a été établie par Jordan dans son célèbre "*Traité des substitutions et des équations algébriques*", paru en 1870, qui lui vaudra le prix Poncelet de l'Académie des Sciences. Nous avons fait l'essentiel du travail dans les sections précédentes, il ne nous reste donc plus qu'à faire la synthèse.

2.7.1. Définition. *Un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est dit* triangularisable (*on dit aussi parfois* "trigonalisable") *s'il existe une base $\mathcal{B}' = (v_1, \dots, v_n)$ dans laquelle la matrice $A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$ est triangulaire, c'est-à-dire par exemple triangulaire inférieure*

$$A' = \begin{pmatrix} \alpha_1 & 0 & \dots & 0 \\ * & \alpha_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & * & \dots & \alpha_n \end{pmatrix}, \quad \alpha_j \in \mathbb{K}.$$

(*Notons qu'on peut passer de la forme triangulaire inférieure à la forme triangulaire supérieure et vice-versa en renversant simplement l'ordre des vecteurs de la base $\mathcal{B}'$ choisie*).

On dit de même qu'une matrice A est triangularisable s'il existe une matrice de passage P telle que $A' = P^{-1}AP$ soit triangulaire.

2.7.2. Théorème de réduction de Jordan. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme. Les propriétés suivantes sont équivalentes* :

(a) *f est triangularisable*;

(b) *le polynôme caractéristique χ_f est scindé sur $\mathbb{K}$, $\chi_f(X) = \prod_{j=1}^{s} (X - \lambda_j)^{m_j}$, avec des $\lambda_j \in \mathbb{K}$ deux à deux distincts.*

Lorsque le polynôme caractéristique χ_f est scindé sur $\mathbb{K}$ comme indiqué en (b), *on peut trouver une décomposition par blocs*

$$E = V_1 \oplus \cdots \oplus V_s, \quad f = f_1 \boxplus \cdots \boxplus f_s, \quad f_j = f_{|V_j}$$

où les $V_j = C_{f,\lambda_j} = \mathrm{Ker}(f - \lambda_j \mathrm{Id}_E)^{m_j}$ sont les sous-espaces caractéristiques, qui ont pour dimensions $\dim C_{f,\lambda_j} = m_j$. De plus, il existe une base $\mathcal{B}'$ de E formée de la juxtaposition de bases des V_j, telle que $A' = \mathrm{Mat}_{\mathcal{B}'}^{\mathcal{B}'}(f)$ soit une "matrice triangulaire de Jordan"

$$A' = \begin{pmatrix} A'_1 & O & \dots & O \\ O & A'_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & A'_s \end{pmatrix}, \quad A'_j = \begin{pmatrix} \lambda_j & 0 & 0 & \dots & 0 & 0 \\ \alpha_{j,1} & \lambda_j & 0 & \dots & 0 & 0 \\ 0 & \alpha_{j,2} & \lambda_j & \dots & 0 & 0 \\ 0 & 0 & \alpha_{j,3} & \ddots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \lambda_j & 0 \\ 0 & 0 & 0 & \dots & \alpha_{j,n_j-1} & \lambda_j \end{pmatrix}$$

où $(\alpha_{j,k})_{1 \leqslant k \leqslant n_j - 1}$ est une suite de 0 et de 1 (un bloc A'_j comme ci-dessus est appelé bloc de Jordan, *on peut aussi le prendre triangulaire supérieur*).

Enfin, la multiplicité m'_j de λ_j dans le polynôme minimal $\mu_f(X) = \prod_{j=1}^{s}(X-\lambda_j)^{m'_j}$ correspond au maximum des tailles des sous-blocs cycliques du bloc A'_j associé à cette valeur propre (= maximum du nombre de 1 consécutifs plus 1).

Voici par exemple un bloc de Jordan de taille 8, de valeur propre λ, composé de 3 sous-blocs cycliques de tailles respectives $4, 3, 1$, et les multiplicités associées :

$$J_{\lambda\,;\,4,3,1} = \begin{pmatrix} \lambda & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \lambda & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & \lambda & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \lambda & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & \lambda & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{pmatrix}, \quad m_\lambda = 8, \quad m'_\lambda = 4.$$

Démonstration. (a) $\Rightarrow$ (b). Si f admet une matrice triangulaire A' comme dans la définition 2.7.1, alors $\chi_f(X) = \prod_{j=1}^{n}(X-\alpha_j)$ donc χ_f est scindé.

(b) $\Rightarrow$ (a). Si $\chi_f(X) = \prod_{j=1}^{n}(X-\lambda_j)^{m_j}$ est scindé, on sait déjà par le théorèmes 2.5.5 que l'on a une décomposition par blocs au moyen des sous-espaces caractéristiques $V_j = C_{f,\lambda_j} = \mathrm{Ker}(f-\lambda_j\,\mathrm{Id}_E)^{m_j}$. Mais alors

$$h_j = (f-\lambda_j\,\mathrm{Id}_E)_{|V_j} = f_{|V_j} - \lambda_j\,\mathrm{Id}_{V_j}$$

satisfait $h_j^{m_j} = 0$, i.e. h_j est nilpotent. Il existe donc une base $\mathcal{B}'_j$ de V_j dans laquelle la matrice de h_j est un bloc triangulaire nilpotent comme dans 2.6.7, ce qui implique alors que $f_j = f_{|V_j} = h_j + \lambda_j\,\mathrm{Id}_{V_j}$ est un bloc de Jordan de valeur propre λ_j. L'assertion concernant les multiplicités m'_j résulte du théorème 2.6.6. □

2.7.3. Corollaire. *Sur le corps $\mathbb{K} = \mathbb{C}$, tout endomorphisme $f \in \mathrm{End}_{\mathbb{C}}(E)$ est trianguarisable, et on peut trouver une base $\mathcal{B}'$ de E dans laquelle $A' = \mathrm{Mat}^{\mathcal{B}'}_{\mathcal{B}'}(f)$ est une matrice triangulaire de Jordan.*

2.7.4. Corollaire. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme et $A \in \mathcal{M}_{n\times n}(\mathbb{K})$ sa matrice dans une base de E. Si $\mathbb{L} \supset \mathbb{K}$ est le corps de décomposition du polynôme caractéristique $\chi_f = \chi_A$ (chapitre 4, théorème 2.8.4), il existe une matrice de passage $P \in \mathcal{M}_{n\times n}(\mathbb{L})$ telle que $A' = P^{-1}AP \in \mathcal{M}_{n\times n}(\mathbb{L})$ soit une matrice triangulaire de Jordan.*

2.7.5. Remarque. Dans le théorème de réduction de Jordan, on peut décomposer chaque bloc A'_j en $A'_j = D'_j + N'_j$ avec $D'_j = \lambda_j I_{n_j}$ diagonale et N'_j triangulaire nilpotente (et cyclique) ; de façon évidente, D'_j et N'_j commutent. En assemblant ces blocs, on obtient une décomposition $A' = D' + N'$ où D' est une matrice diagonale, et N' est une matrice triangulaire nilpotente (i.e. avec des 0 sur la diagonale), telles que D' et N' commutent. En revenant à la matrice $A = PA'P^{-1}$ via la matrice de passage P, on trouve la décomposition dite de *Jordan-Chevalley*

$$A = A_{\mathrm{ss}} + A_{\mathrm{nil}} \quad \text{avec } A_{\mathrm{ss}} := PD'P^{-1},\ \ A_{\mathrm{nil}} = PN'P^{-1} \text{ qui commutent,}$$

telles que $A_{\rm ss}$ soit *semi-simple* (c'est-à-dire par définition diagonalisable sur une extension $\mathbb{L} \supset \mathbb{K}$), et $A_{\rm nil}$ *nilpotente*. Son intérêt est que pour un corps $\mathbb{K}$ "parfait" tel que $\mathbb{R}$ on a bien $A_{\rm ss}, A_{\rm nil} \in \mathcal{M}_{n\times n}(\mathbb{K})$: cette propriété sera démontrée au §2.8.

La réduction de Jordan permet de résoudre le problème de la congugaison des matrices. Rappelons d'abord quelques définitions.

2.7.6. Définition. (a) *Deux matrices $A, B \in \mathcal{M}_{n\times p}(\mathbb{K})$ sont dites équivalentes, $A \equiv B$, s'il existe des matrices inversibles $P \in \mathcal{M}_{n\times n}(\mathbb{K})$ et $Q \in \mathcal{M}_{p\times p}(\mathbb{K})$ telles que $B = PAQ$.*

(b) *Deux matrices carrées $A, B \in \mathcal{M}_{n\times n}(\mathbb{K})$ sont dites conjuguées, $A \simeq B$, s'il existe une matrice inversible $P \in \mathcal{M}_{n\times n}(\mathbb{K})$ telle que $B = PAP^{-1}$.*

On vérifie aisément qu'il s'agit de relations d'équivalence. Du point de vue des applications linéaires, la relation (a) signifie qu'on a affaire à la *même application linéaire* $f : E \to F$ quitte à changer les bases de E et F, tandis que la relation (b) signifie qu'on a affaire au *même endomorphisme* $f : E \to E$ quitte à changer la base de référence dans E. La relation $A \equiv B$ est facile à décrire : il faut et il suffit que $\mathrm{rang}(A) = \mathrm{rang}(B)$. En effet, si $f : E \to F$ est de rang r, en prenant une base $(e_1, \dots, e_p)$ de E telle que $(e_{r+1}, \dots, e_p)$ soit une base de $\mathrm{Ker}(f)$ et une base $(e'_1, \dots, e'_n)$ de f obtenue en complétant $e'_1 = f(e_1), \dots, e'_r = f(e_r)$, on se ramène toujours à ce que la matrice de f soit une matrice rectangulaire de la forme

$$A = \begin{pmatrix} 1 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 0 & 0 & \dots & 0 & \dots & 0 \end{pmatrix}$$

avec r coefficients 1 sur la diagonale principale. Pour la conjugaison, on a le résultat beaucoup plus subtil suivant – au moins sur le corps des complexes. (Un énoncé similaire serait valable sur tout corps $\mathbb{K}$, à condition de passer à une extension $\mathbb{L} \supset \mathbb{K}$ convenable).

2.7.7. Théorème. *Deux matrices $A, B \in \mathcal{M}_{n\times n}(\mathbb{C})$ sont conjuguées si et seulement si elles ont le même polynôme caractéristique*

$$\chi_A(X) = \chi_B(X) = \prod_{1\leqslant j\leqslant s} (X - \lambda_j)^{m_j}$$

et si les sous-espaces caractéristiques se décomposent en sous-blocs cycliques de mêmes tailles (autrement dit, si elles ont la même forme réduite de Jordan, à l'ordre près des sous-blocs cycliques associés aux différentes valeurs propres).

Démonstration. Si les matrices A et B sont conjuguées, elles peuvent être associées au même endomorphisme f dans des bases différentes. On a donc par construction

une même forme réduite de Jordan, celle associée à f. Réciproquement, si A et B ont même polynôme caractéristique et même forme réduite de Jordan J, il existe des matrices de passage P, Q inversibles telles que $P^{-1}AP = J = Q^{-1}BQ$, donc $B = (PQ^{-1})^{-1}A(PQ^{-1})$, et A, B sont conjuguées. □

2.8. Décomposition de Jordan-Chevalley***

Cette décomposition a été établie par Claude Chevalley (1909-1984) dans les années 1950. Elle s'applique directement aux endomorphismes, indépendamment de tout choix de base, et généralise la réduction de Jordan dans le cas où le polynôme caractéristique n'est plus nécessairement scindé sur le corps $\mathbb{K}$ considéré – ce qui se produit souvent sur $\mathbb{Q}$ ou sur $\mathbb{R}$. (Assez curieusement, certains livres de cours en langue française utilisent la terminologie de "décomposition de Dunford", mais c'est semble-t-il à tort, les travaux de Nelson Dunford étant postérieurs à ceux de Chevalley et pas nécessairement en lien direct avec les résultats en question ...). On commencera par introduire les notions utiles d'endomorphisme simple et semi-simple, et on en décrira les propriétés fondamentales.

Nota : cette section est conceptuellement assez difficile – et hors-programme !

2.8.1. Théorème et définition. *Pour $f \in \mathrm{End}_{\mathbb{K}}(E)$, $n = \dim E \geqslant 1$, les deux propriétés suivantes sont équivalentes :*

(a) *Les seuls sous-espaces $S \subset E$ qui sont stables par f sont $S = \{0\}$ et $S = E$.*

(b) *Le polynôme caractéristique $\chi_f(X) \in \mathbb{K}[X]$ est irréductible.*

On dit alors que f est un endomorphisme simple. *Pour un tel endomorphisme f, on a l'égalité $\mu_f = \chi_f$.*

Démonstration. (b) $\Rightarrow$ (a). Si on avait un sous-espace S stable par f distinct de $\{0\}$ et E, on en déduirait une décomposition $\chi_f = \chi_{f|S}\,\chi_g$ en facteurs de degrés $\geqslant 1$ (théorème 1.3.2), ce qui contredirait l'hypothèse que χ_f est irréductible.

(a) $\Rightarrow$ (b). L'hypothèse (a) entraîne nécessairement que f est cyclique, car on a nécessairement $S = \mathrm{vect}(f^i(v))_{i\geqslant 0} = E$ si $v \neq 0$. Par suite $\mu_f = \chi_f$. Si μ_f n'est pas irréductible, deux cas peuvent se présenter : ou bien $\mu_f = P^m$ est une puissance d'un irréductible avec $m \geqslant 2$, ou bien μ_f contient des facteurs irréductibles distincts. Dans ce deuxième cas, le théorème de décomposition par blocs montrerait que E est une somme directe de plusieurs sous-espaces stables non triviaux, ce qui est contradictoire. Mais dans le cas $\mu_f = P^m$, $m \geqslant 2$, le sous-espace $S = \mathrm{Ker}\, P(f)^{m-1}$ serait non nul et distinct de E, contradiction également.

Si f est simple, μ_f est un polynôme de degré $\geqslant 1$ qui divise χ_f. L'irréductibilité de χ_f implique $\mu_f = \chi_f$. □

2.8.2. Exemples. (a) Si $n = 1$, tout endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est simple (et de la forme $f = \lambda\,\mathrm{Id}_E$). Réciproquement si $\mathbb{K} = \mathbb{C}$, les seuls polynômes irréductibles unitaires sont les $\chi_f(X) = X - \lambda$, et donc les seuls endomorphismes simples sont les homothéties $f = \lambda\,\mathrm{Id}_E$ sur un $\mathbb{C}$-espace vectoriel E de dimension 1.

(b) Sur $\mathbb{K} = \mathbb{R}$, la matrice 2×2

$$A = \begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad b \neq 0,$$

admet pour polynôme caractéristique $\chi_A(X) = (X-a)^2 + b^2$ qui est irréductible sur $\mathbb{R}$, donc A est simple. Géométriquement si on identifie $\mathbb{R}^2$ à $\mathbb{C}$, il s'agit d'une similitude directe $z \mapsto \lambda z$, de rapport α et d'angle θ avec $\lambda = \alpha e^{i\theta} = a + ib$. Il n'y a pas dans $\mathbb{R}^2$ de sous-espace vectoriel invariant si $b \neq 0$, i.e. $\theta \not\equiv 0 \bmod \pi$ (mais si $b = 0$, $A = aI_2$, $\chi_A(X) = (X-a)^2$ et toute droite est invariante). Comme les seuls polynômes irréductibles de $\mathbb{R}[X]$ sont de degré 1 et 2, les seuls endomorphismes simples sur le corps des réels apparaissent en dimension 1 et 2 (et on les a tous décrits à isomorphisme près).

(c) Sur $\mathbb{K} = \mathbb{Q}$, la matrice 3×3

$$A = \begin{pmatrix} 0 & 0 & 2 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

est la matrice compagnon du polynôme irréductible $Q(X) = X^3 - 2$, donc A est simple. La théorie nous dit que les seuls sous-espaces stables S de l'endomorphisme associé $f \in \mathrm{End}_{\mathbb{Q}}(\mathbb{Q}^3)$ sont $\{0\}$ et $\mathbb{Q}^3$. Sur $\mathbb{K}' = \mathbb{R}$ ou $\mathbb{C}$, on a en revanche la droite propre $\mathbb{K}'v$ de valeur propre $\sqrt[3]{2}$ avec

$$v = \begin{pmatrix} \sqrt[3]{4} \\ \sqrt[3]{2} \\ 1 \end{pmatrix}.$$

(d) Plus généralement, soit $P \in \mathbb{K}[X]$ un polynôme irréductible unitaire de degré d,

$$P(X) = a_0 + a_1 X + \cdots + a_{d-1} X^{d-1} + X^d.$$

On a vu au chapitre 4 (théorème 2.8.2), que $\mathbb{L} = \mathbb{K}[X]/\langle P \rangle$ est un corps, et que c'est aussi un $\mathbb{K}$-espace vectoriel de dimension $\dim_{\mathbb{K}} \mathbb{L} = d$, admettant comme base

$$\mathcal{B} = (\dot{1}, \dot{X}, \ldots, \dot{X}^{d-1}).$$

Alors $\lambda = \dot{X}$ est une racine du polynôme P dans $\mathbb{L}[X]$, i.e. $P(\lambda) = 0$. L'endomorphisme "évident"

$$f_P : \mathbb{L} \to \mathbb{L}, \quad v \mapsto \lambda v, \quad \text{i.e.} \quad f_P : \mathbb{L} \to \mathbb{L}, \quad \dot{Q} \mapsto (XQ)^{\bullet},$$

qui a trivialement pour matrice $(\lambda) \in \mathcal{M}_{1\times 1}(\mathbb{L})$ lorsqu'on se place sur le corps $\mathbb{L}$, admet, lorsque $\mathbb{L}$ est vu comme espace vectoriel sur $\mathbb{K}$, la matrice

$$\mathrm{Mat}_{\mathcal{B}}^{\mathcal{B}}(f_P) = \begin{pmatrix} 0 & 0 & 0 & \ldots & 0 & -a_0 \\ 1 & 0 & 0 & \ldots & 0 & -a_1 \\ 0 & 1 & 0 & \ldots & 0 & -a_2 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 & -a_{d-2} \\ 0 & 0 & \ldots & 0 & 1 & -a_{d-1} \end{pmatrix} \in \mathcal{M}_{d\times d}(\mathbb{K}).$$

C'est précisément la matrice compagnon du polynôme P : en effet l'image de la base $\mathcal{B}$ par f_P est $f_P(\mathcal{B}) = (\dot{X}, \dot{X}^2, \dots, \dot{X}^d)$, et modulo $\langle P\rangle$ on a

$$\dot{X}^d = -a_0\dot{1} - a_1\dot{X} - \cdots - a_{d-1}\dot{X}^{d-1}.$$

Sur le corps $\mathbb{K}$, on a donc $\chi_{f_P}(X) = P(X)$, et f_P est simple. □

On va maintenant caractériser les endomorphismes dont le polynôme minimal μ_f est irréductible.

2.8.3. Proposition. *Soit $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme dont le polynôme minimal $\mu_f(X) = P(X) \in \mathbb{K}[X]$ est irréductible. Alors on a une décomposition en somme directe $E = W_1 \oplus \cdots \oplus W_m$ et une décomposition en blocs correspondante $f = f_1 \boxplus \cdots \boxplus f_m$ avec des endomorphismes simples $f_j = f_{|W_j} \in \mathrm{End}_{\mathbb{K}}(W_j)$ tels que $\chi_{f_j} = P$. Dans ces conditions, on a $\chi_f = P^m$, et il existe des isomorphismes $\mathbb{K}$-linéaires*

$$\varphi_j : \mathbb{L} = \mathbb{K}[X]/\langle P\rangle \to W_j$$

tels que $f_j = \varphi_j \circ f_P \circ \varphi_j^{-1}$. Les blocs f_j sont donc tous "isomorphes" entre eux et isomorphes à l'exemple 2.8.2 (d).

Démonstration. On sait que χ_f divise $\mu_f^{\dim E}$, donc χ_f est une puissance de P. Soit v_1 un vecteur non nul de E. On considère le morphisme

$$\psi_1 : \mathbb{K}[X] \to E, \quad Q \mapsto Q(f)(v).$$

Son image n'est autre que le sous-espace cyclique $W_1 = \mathrm{vect}(f^i(v_1))_{i\geqslant 0}$, stable par f. Il est d'autre part évident que $\mathrm{Ker}\,\psi_1 = \langle Q_1\rangle$ est un idéal de $\mathbb{K}[X]$ qui contient $\langle P\rangle$ mais qui n'est pas égal à $\mathbb{K}[X]$ tout entier ($1 \notin \mathrm{Ker}\,\psi_1$ puisque $Q(f)(v_1) = v_1$ pour $Q = 1$). Comme P est irréductible, $Q_1 \mid P$ et Q_1 non inversible, on a nécessairement $\mathrm{Ker}\,\psi_1 = \langle P\rangle$. Par passage au quotient, on obtient par conséquent un isomorphisme

$$\varphi_1 : \mathbb{K}[X]/\langle P\rangle \longrightarrow W_1,$$

et en particulier $\dim W_1 = \deg P = d$. On a ici par définition

$$\varphi_1(\dot{Q}) = \psi_1(Q) = Q(f)(v),$$
$$\varphi_1(f_P(\dot{Q}) = \varphi_1(\dot{X}\dot{Q}) = \psi_1(XQ) = f \circ Q(f)(v) = f(\varphi_1(\dot{Q}))$$

pour tout $Q \in \mathbb{K}[X]$, ce qui signifie que $\varphi_1 \circ f_P = f_1 \circ \varphi_1$ où $f_1 = f_{|W_1}$ est la restriction de f à W_1. Par suite on a bien $f_1 = \varphi_1 \circ f_P \circ \varphi_1^{-1}$. Par ailleurs χ_{f_1} qui divise χ_f est aussi une puissance de P, mais comme $\deg \chi_{f_1} = \dim W_1 = d$, on doit avoir $\chi_{f_1} = P$, donc f_1 est simple. Ce résultat vaut pour tout vecteur non nul $v_1 \in E$. Si $W_1 = E$, on a fini (avec $m = 1$). Sinon on prend $v_2 \in E \smallsetminus W_1$ et on construit W_2 comme pour W_1. On voit que $W_1 \cap W_2 \subsetneq W_2$ est stable par f, donc $W_1 \cap W_2 = \{0\}$ et W_1, W_2 sont en somme directe. Si $E = W_1 \oplus W_2$,

on a fini (avec $m = 2$). Sinon, on prend $v_3 \in E \smallsetminus (W_1 \oplus W_2)$, on constate que $W_3 \cap (W_1 \oplus W_2) \subsetneq W_3$ est stable, donc réduit à $\{0\}$, par suite W_1, W_2, W_3 sont en somme directe. On peut continuer ainsi jusqu'à ce que $E = W_1 \oplus \cdots \oplus W_m$. Il vient alors $\chi_f = \prod \chi_{f_j} = P^m$, et l'affirmation concernant l'isomorphisme φ_j a déjà été démontrée. □

Ce qui précède permet entre autres de décrire ce qui se passe lorsque le polynôme caractéristique $\chi_f(X)$ admet une racine dans une extension $\mathbb{L}$ du corps $\mathbb{K}$ dans lequel on travaille – c'est une généralisation du théorème 1.4.1, qui concernait le cas de l'extension $\mathbb{C} \supset \mathbb{R}$ et d'une racine complexe non réelle de $\chi_f(X)$.

2.8.4. Théorème. *Soit E un $\mathbb{K}$-espace vectoriel, $f \in \mathrm{End}_{\mathbb{K}}(E)$ un endomorphisme, et P un facteur irréductible de degré d du polynôme caractéristique $\chi_f(X)$.*

(a) *Il existe un sous-espace stable $W \subset E$, de dimension d, et un isomorphisme $\mathbb{K}$-linéaire $\varphi : \mathbb{L} = \mathbb{K}[X]/\langle P\rangle \to W$ tel que $f_{|W} = \varphi \circ f_P \circ \varphi^{-1}$.*

(b) *En particulier, tout endomorphisme simple $f \in \mathrm{End}_{\mathbb{K}}(E)$ de polynôme caractéristique $\chi_f = P$ est "isomorphe" à l'exemple f_P du 2.8.2 (d), i.e. on a $f = \varphi \circ f_P \circ \varphi^{-1}$ pour un isomorphisme $\mathbb{K}$-linéaire $\varphi : \mathbb{L} = \mathbb{K}[X]/\langle P\rangle \to E$.*

Démonstration. (a) Écrivons $\mu_f(X) = \prod_{j=1}^{s} P_j(X)^{m'_j}$ avec $P = P_1$. On pose $S = \mathrm{Ker}\, P(f)$. Nécessairement $S \neq \{0\}$, sinon $P(f) = P_1(f)$ serait inversible et la relation $\mu_f(f) = 0$ impliquerait aussi que $\prod_{j=2}^{s} P_j(f)^{m'_j} = 0$, ce qui est contradictoire. Par construction $f_{|S}$ est annulé par P, donc $\mu_{f|S} = P$. Mais alors le théorème 2.8.3 fournit des sous-espaces $W_j \subset S$ ayant la propriété annoncée (en fait la preuve de 2.8.3 montre que tout sous-espace cyclique $W = \mathrm{Vect}(f^i(v))_{i\geqslant 0}$ engendré par $v \in S$ satisfait (a)).

(b) Si f est simple, on a nécessairement $W = E$ et (b) résulte de (a). □

Dans la suite, nous aurons besoin d'exclure certains corps ayant un comportement "pathologique". Ceci amène à la définition et au lemme qui suivent.

2.8.5. Définition. *Un corps $\mathbb{K}$ est dit parfait si sa caractéristique est nulle, ou bien s'il est de caractéristique $p > 0$ et que tout élément de $\mathbb{K}$ admet une racine p-ième dans $\mathbb{K}$ (autrement dit si le morphisme dit de Frobenius $x \mapsto x^p$ est surjectif).*

2.8.6. Exemples. (a) Les corps $\mathbb{Q}, \mathbb{R}, \mathbb{C}$, et plus généralement tous les sous-corps de $\mathbb{C}$ sont parfaits (il y en a beaucoup d'autres que $\mathbb{Q}, \mathbb{R}$, par exemple $\mathbb{Q}[i]$, $\mathbb{Q}[\sqrt{2}\,]$).

(b) Les corps finis sont parfaits. En effet, il est facile de voir en caractéristique $p > 0$ que $F : x \mapsto x^p$ est un morphisme de corps (l'additivité venant de la formule du binôme et du fait que $\binom{p}{k}$ est divisible par p si $0 < k < p$), et $F(x) = 0$ implique $x = 0$, donc F est injectif. Mais si $\mathbb{K}$ est fini, l'injectivité implique trivialement la surjectivité. En particulier $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$ est parfait ; le petit théorème de Fermat nous dit d'ailleurs que $F : \mathbb{F}_p \to \mathbb{F}_p$ est l'application identique.

(c) Si $\mathbb{K}$ est de caractéristique p, le fait que $\binom{p}{k}$ soit divisible par p si $0 < k < p$ implique de nouveau que pour tout polynôme $Q(X) = a_0 + a_1X + \cdots + a_dX \in \mathbb{K}[X]$,

on a

$$Q(X)^p = \left(\sum_{j=0}^{d} a_j X^j\right)^p = \sum_{j=0}^{d} a_j^p X^{pj}.$$

Ceci entraîne par exemple que le corps des fractions rationnelles $\mathbb{K}(X)$ à coefficients dans $\mathbb{K}$ *n'est pas parfait*, car la puissance p-ième d'une fraction rationnelle $R(X)$ est une fraction de la forme $R'(X^p)$, donc $F : \mathbb{K}(X) \to \mathbb{K}(X^p) \subsetneq \mathbb{K}(X)$.

(d) Si $\mathbb{K}$ est un corps quelconque de caractéristique $p > 0$, il est facile de voir que l'extension $\mathbb{K}' = \mathbb{K}[a^{1/p^m}]_{a\in\mathbb{K},m\in\mathbb{N}}$ obtenue en adjoignant à $\mathbb{K}$ toutes les racines p^m-ièmes de tous les éléments $a \in \mathbb{K}$ est un corps parfait.

2.8.7. Lemme. *Soit $\mathbb{K}$ un corps parfait. Alors*

(a) *Si $P \in \mathbb{K}[X]$ est irréductible, on a $P'(X) \neq 0$ et* $\operatorname{pgcd}(P, P') = 1$.

(b) *Un polynôme non nul $Q \in \mathbb{K}[X]$ vérifie* $\operatorname{pgcd}(Q, Q') = 1$ *si et seulement si $Q = cP_1 \cdots P_m$ est un produit de polynômes irréductibles deux à deux distincts, avec $m \in \mathbb{N}$ et $c \in \mathbb{K}^*$.*

Pour illustrer la nécessité de supposer $\mathbb{K}$ parfait, observons que si $\mathbb{K}$ est un corps non parfait et si on choisit $a \in \mathbb{K}$ n'admettant pas de racine p-ième dans $\mathbb{K}$, alors $P(X) = X^p - a$ est irréductible dans $\mathbb{K}[X]$ comme on peut le voir, mais $P'(X) = pX^{p-1} = 0$, et donc on a malencontreusement $\operatorname{pgcd}(P, P') = P \neq 1$. Encore plus problématique pour un polynôme irréductible, on constate que l'on a $P(X) = (X - a^{1/p})^p$ dans l'extension $\mathbb{L} = \mathbb{K}[a^{1/p}]$, donc $a^{1/p} \in \mathbb{L}$ est une racine multiple.

Démonstration. (a) Soit $d = \deg P \geqslant 1$. Si $P' \neq 0$, alors $\deg P' \leqslant d - 1$, donc P ne peut diviser P', et par conséquent $\operatorname{pgcd}(P, P') = 1$. Il reste à étudier ce qui se passe si $P' = 0$. Or, si $P(X) = \sum a_j X^j$, on a $P'(X) = \sum j a_j X^{j-1}$, et P' ne peut donc être nul que si $ja_j = 0$ pour tout j, ce qui implique que le corps est de caractéristique $p > 0$ et que $p \mid j$ chaque fois que $a_j \neq 0$. Le polynôme P est donc de la forme $P(X) = \sum a_{pj} X^{pj}$. Mais comme $\mathbb{K}$ est supposé parfait, il existe $b_j \in \mathbb{K}$ tel que $a_{pj} = b_j^p$, ce qui entraîne

$$P(X) = \sum a_{pj} X^{pj} = \sum b_j^p X^{pj} = \Big(\sum b_j X^j\Big)^p,$$

de sorte que P ne serait pas irréductible. On a donc bien $P' \neq 0$ sous l'hypothèse que P est irréductible.

(b) Supposons $Q = cP_1 \cdots P_m$ comme indiqué. Si $\operatorname{pgcd}(Q, Q') \neq 1$, il existe un polynôme irréductible P qui divise à la fois Q et Q'. Alors P est nécessairement l'un des P_j, par exemple $P = P_1$. Dans ces conditions l'hypothèse

$$P = P_1 \mid Q' = (P_1(P_2 \cdots P_n))' = P_1(P_2 \cdots P_n)' + P_1' P_2 \cdots P_n$$

implique que $P_1 \mid P_1' P_2 \cdots P_n$. Mais $\operatorname{pgcd}(P_1, P_j) = 1$ pour $j \geqslant 2$, et donc $P_1 \mid P_1'$ d'après le lemme de Gauss. Ceci contredit (a), de sorte qu'on a nécessairement

$\mathrm{pgcd}(Q, Q') = 1$. Si Q n'est pas de la forme indiquée, sa décomposition en facteurs premiers contient une puissance P^k, $k \geqslant 2$ d'un polynôme irréductible. Mais alors P divise à la fois Q et Q', et $\mathrm{pgcd}(Q, Q') \neq 1$. □

2.8.8. Théorème et définition. *Soit $\mathbb{K}$ un corps parfait. Pour $f \in \mathrm{End}_{\mathbb{K}}(E)$, $n = \dim E \geqslant 1$, les trois propriétés suivantes sont équivalentes :*

(a) *On a une décomposition en somme directe $E = V_1 \oplus \cdots \oplus V_s$ et une décomposition en blocs correspondante $f = f_1 \boxplus \cdots \boxplus f_s$ avec des endomorphismes simples $f_j \in \mathrm{End}_{\mathbb{K}}(V_j)$.*

(b) *Le polynôme minimal μ_f est un produit $P_1 \cdots P_\ell$ de facteurs irréductibles distincts.*

(c) *Il existe une extension $\mathbb{L} \supset \mathbb{K}$ telle que la matrice A de f relativement à une base quelconque de E soit diagonalisable dans $\mathcal{M}_{n\times n}(\mathbb{L})$ (en considérant que $\mathcal{M}_{n\times n}(\mathbb{L}) \supset \mathcal{M}_{n\times n}(\mathbb{K})$).*

On dit alors que f est un endomorphisme semi-simple.

Démonstration. (a) ⇒ (b). C'est clair, car μ_f est le ppcm des polynômes irréductibles $\mu_{f_j} = \chi_{f_j}$, donc c'est le produit de tous les éléments distincts

$$P_i \in \{\chi_{f_j} \,/\, j = 1, \ldots, s\}.$$

(b) ⇒ (c). Si (*b*) est vérifié, on prend pour extension $\mathbb{L} \supset \mathbb{K}$ le corps de décomposition du polynôme μ_f (chapitre 4, théorème 2.8.4). Le lemme 2.8.7 (b) implique $\mathrm{pgcd}(\mu_f, \mu_f') = 1$. Ceci entraîne que les racines de $\mu_f(X)$ dans $\mathbb{L}$ sont simples, et le corollaire 2.4.3 montre alors que la matrice A de f dans n'importe quelle base est diagonalisable dans $\mathcal{M}_{n\times n}(\mathbb{L})$.

(c) ⇒ (b). Si A est diagonalisable dans $\mathcal{M}_{n\times n}(\mathbb{L})$ avec $\mathbb{L} \supset \mathbb{K}$, on sait (corollaire 2.4.3) que le polynôme minimal $\mu_f = \mu_A$ est scindé sur $\mathbb{L}$ et a toutes ses racines simples. Ceci implique que dans $\mathbb{K}[X]$ on a $\mu_f = P_1 \cdots P_\ell$ avec des facteurs irréductibles sans multiplicités.

(b) ⇒ (a). Si $\mu_f = P_1 \cdots P_\ell$, le théorème des noyaux montre que f admet une décomposition par blocs $f = g_1 \boxplus \cdots \boxplus g_\ell$ suivant les sous-espaces $V_j = \mathrm{Ker}\, P_j(f)$, et si $g_j = f_{|V_j}$, on a $\mu_{g_j} = P_j$. Mais dans ce cas, d'après la proposition 2.8.3, g_j admet une décomposition $g_j = f_{j,1} \boxplus \cdots \boxplus f_{j,k_j}$ avec des $f_{j,k}$ simples, et la preuve est terminée. □

2.8.9. Lemme. *Soit $T \in \mathbb{K}[X]$ un polynôme de degré $d \geqslant 1$.*

(a) *Il existe une écriture*

$$T(X+Y) = T(X) + T_1(X)Y + \cdots + T_d(X)Y^d$$

avec $T_j(X) \in \mathbb{K}[X]$ de degré $\leqslant d - j$, et on a en particulier $T_1 = T'$.

(b) *Si $\mathrm{pgcd}(T, T') = 1$, il existe pour tout $m \geqslant 1$ des polynômes $Q_m(X) \in \mathbb{K}[X]$ tels que*

$$T\big(X + Q_m(X)T(X)\big) \equiv 0 \mod \langle T(X)^m \rangle.$$

Démonstration. (a) résulte simplement de la formule du binôme de Newton appliquée à chaque monôme $a_i X^i$ de T, ou, si on veut, de la formule de Taylor qui donne $T_j(X) = \frac{1}{j!}T^{(j)}(X)$ (mais cette écriture suppose que $\mathbb{K}$ soit de caractéristique > 0, hypothèse qui n'est pas nécessaire ici). On a en tout cas par définition $T_1 = T'$.

(b) On raisonne par récurrence sur m. Si $m = 1$, on peut prendre simplement $Q_1 = 0$. Supposons Q_m déjà construit pour $m \geqslant 1$, tel que

$$(*) \qquad T(X - Q_m(X)T(X)) = U(X)T(X)^m.$$

On cherche Q_{m+1} sous la forme $Q_{m+1} = Q_m(X) + V(X)T(X)^{m-1}$. La formule (a) appliquée deux fois avec $Y = Q_m(X)T(X)$ et $Y = Q_{m+1}(X)T(X)$ donne

$$\begin{aligned} T(X + Q_{m+1}(X)T(X)) - T(X + Q_m(X)T(X)) & \\ &= \sum_{j=1}^{d} T_j(X)T(X)^j \Big(Q_{m+1}(X)^j - Q_m(X)^j\Big) \\ &\equiv T'(X)V(X)T(X)^m \mod \langle T(X)^{m+1}\rangle, \end{aligned}$$

car les termes d'indice $j \geqslant 2$ sont multiples de $T(X)^{m+1}$, vu que la différence $Q_{m+1}(X)^j - Q_m(X)^j$ contient un facteur $Q_{m+1}(X) - Q_m(X) = V(X)T(X)^{m-1}$. Par conséquent

$$T(X - Q_{m+1}(X)T(X)) \equiv \big(U(X) + T'(X)V(X)\big)T(X)^m \mod \langle T(X)^{m+1}\rangle$$

Nous voulons que le membre de droite soit congru à 0 mod $\langle T(X)^{m+1}\rangle$, et pour cela il suffit que que $U(X) + T'(X)V(X)$ soit divisible par $T(X)$. Or, l'identité de Bézout implique l'existence de polynômes $A, B \in \mathbb{K}[X]$ tels que $A(X)T(X) + B(X)T'(X) = 1$. Si on prend $V(X) = -B(X)U(X)$, il vient alors

$$U(X) + T'(X)V(X) = U(X)\big(1 - B(X)T'(X)\big) = U(X)A(X)T(X).$$

Ceci implique l'existence d'un polynôme Q_{m+1} satisfaisant la condition $(*)$ à l'ordre $m+1$, et l'étape de récurrence est démontrée. □

2.8.10. Théorème de Jordan-Chevalley. *Soit $\mathbb{K}$ un corps parfait et f un endomorphisme quelconque de $\mathrm{End}_{\mathbb{K}}(E)$. On peut alors trouver une décomposition*

$$f = f_{\mathrm{ss}} + f_{\mathrm{nil}}$$

avec des endomorphismes $f_{\mathrm{ss}}, f_{\mathrm{nil}} \in \mathrm{End}_{\mathbb{K}}(E)$ ayant les propriétés suivantes :

(a) *f_{ss} et f_{nil} commutent : $f_{\mathrm{ss}} \circ f_{\mathrm{nil}} = f_{\mathrm{nil}} \circ f_{\mathrm{ss}}$;*

(b) *f_{ss} est semi-simple et f_{nil} est nilpotent.*

Sous les hypothèses (a) *et* (b)*, la décomposition $f = f_{\mathrm{ss}} + f_{\mathrm{nil}}$ est unique. De plus :*

(c) *f_{ss} et f_{nil} peuvent être écrits comme des polynômes $f_{\mathrm{ss}} = Q(f)$, $f_{\mathrm{nil}} = R(f)$ de l'endomorphisme f, où $Q, R \in \mathbb{K}[X]$ ne dépendent que de $\mu_f(X)$;*

(d) *si* $\mathbb{L} \supset \mathbb{K}$ *est une extension de* $\mathbb{K}$ *dans laquelle le polynôme caractéristique* $\chi_f(X)$ *est scindé, de sorte que la matrice* $A \in \mathcal{M}_{n\times n}(\mathbb{K})$ *de* f *admet une forme réduite de Jordan* $A' = P^{-1}AP \in \mathcal{M}_{n\times n}(\mathbb{L})$ *pour une certaine matrice de passage* $P \in \mathcal{M}_{n\times n}(\mathbb{L})$*, alors les matrices* $A_{\text{ss}}, A_{\text{nil}} \in \mathcal{M}_{n\times n}(\mathbb{K})$ *de* $f_{\text{ss}}, f_{\text{nil}}$ *sont données par la décomposition*

$$\begin{aligned}&A = A_{\text{ss}} + A_{\text{nil}} \quad \textit{où}\\ &A' = D' + N',\ \ D' = \textit{partie diagonale de } A',\ \ N' = \textit{partie triangulaire de } A',\\ &A_{\text{ss}} = PD'P^{-1},\quad A_{\text{nil}} = PN'P^{-1}.\end{aligned}$$

Démonstration. Existence de la décomposition. Soit $\mu_f = \prod_{j=1}^{s} P_j^{m'_j}$ la décomposition de μ_f en facteurs irréductibles et soit $T = \prod_{j=1}^{s} P_j$. Comme $\mathbb{K}$ est parfait, nous avons $\operatorname{pgcd}(T, T') = 1$ grâce au lemme 2.8.7 (b). Prenons $m \geqslant \max\{m'_j\}$ de façon que $\mu_f(X) \mid T(X)^m$ (on peut par exemple prendre pour m le maximum $\max\{m_j\}$ des multiplicités de χ_f, qui est mieux connu, ou encore $m = n = \dim E$). Le point important est que l'on ait $T(f)^m = 0$. Le lemme 2.8.9 (b) fournit des polynômes $Q_m, R_m \in \mathbb{K}[X]$ tels que

$$(*) \qquad T\big(X + Q_m(X)T(X)\big) = R_m(X)T(X)^m.$$

On pose

$$f_{\text{ss}} = f + Q_m(f)\circ T(f), \quad f_{\text{nil}} = -Q_m(f)\circ T(f),$$

de sorte que f_{ss} et f_{nil} commutent (en tant que polynômes de l'endomorphisme f), et $f = f_{\text{ss}} + f_{\text{nil}}$. L'égalité $(*)$ implique

$$T(f_{\text{ss}}) = R_m(f)\circ T(f)^m = 0,$$

et d'autre part

$$(f_{\text{nil}})^m = (-1)^m Q_m(f)^m \circ T(f)^m = 0.$$

On voit donc que f_{nil} est nilpotent et que $\mu_{f_{\text{ss}}}$ divise T, ce qui entraîne que $\mu_{f_{\text{ss}}}$ a des facteurs irréductibles 2 à 2 distincts, et donc, d'après le critère 2.8.8 (b), que f_{ss} est bien semi-simple. La propriété (c) est une conséquence immédiate de ce qui précède.

Pour démontrer l'unicité de la décomposition, on a besoin de quelques lemmes assez élémentaires.

2.8.11. Lemme. *Si* $g_1, g_2 \in \operatorname{End}_{\mathbb{K}}(E)$ *sont semi-simples et commutent, alors* $g = g_1 + g_2$ *est semi-simple.*

Démonstration. On se place dans une extension $\mathbb{L} \supset \mathbb{K}$ dans laquelle μ_{g_1} et μ_{g_2} sont scindés. En considérant les matrices de g_1, g_2 et en changeant de corps si nécessaire, on voit d'après le critère 2.8.8 (c) qu'on peut supposer g_1 et g_2 diagonalisable dans $\mathcal{M}_{n\times n}(\mathbb{K})$. Mais alors, on a une décomposition $E = V_1 \oplus \cdots \oplus V_s$ suivant les sous-espaces propres V_j de g_1 qui, en outre, sont stables par g_2. Dans ces

conditions, $g_{1|V_j} = \lambda_j \operatorname{Id}_{V_j}$ et $g_{2|V_j}$ est diagonalisable dans V_j, ce qui fournit une base commune $\mathcal{B}'_j$ de vecteurs propres pour g_1 et g_2 dans chaque V_j, et par suite une base commune $\mathcal{B}'$ de vecteurs propres dans E. On voit alors que $g_1 + g_2$ est diagonalisable dans la base $\mathcal{B}'$ de E, et le critère 2.8.8 (c) implique que $g_1 + g_2$ est semi-simple. □

2.8.12. Lemme. *Si $h_1, h_2 \in \mathrm{End}_{\mathbb{K}}(E)$ sont nilpotents et commutent, alors $h = h_1 + h_2$ est nilpotent.*

Démonstration. Si $h_1^{m_1} = 0$ et $h_2^{m_2} = 0$, la formule du binôme de Newton montre que $(h_1 + h_2)^{m_1+m_2} = 0$. □

2.8.13. Lemme. *Si un endomorphisme $f \in \mathrm{End}_{\mathbb{K}}(E)$ est à la fois nilpotent et semi-simple, alors $f = 0$.*

Démonstration. On doit avoir à la fois que $\mu_f(X) = X^m$ pour un certain $m \in \mathbb{N}^*$ et que μ_f est un produit de polynômes irréductibles distincts (d'après 2.8.8 (b)). Ceci impose $\mu_f(X) = X$, et donc $f = \mu_f(f) = 0$. □

Fin de la démonstration du théorème de Jordan-Chevalley 2.8.10. Il nous faut d'abord démontrer l'unicité. Si l'on avait une deuxième décomposition

$$f = f_{\mathrm{ss}} + f_{\mathrm{nil}} = g + h$$

avec g semi-simple, h nilpotent et $g \circ h = h \circ g$, alors g commuterait avec $f = g + h$ et donc aussi avec $f_{\mathrm{ss}} = Q(f)$, $f_{\mathrm{nil}} = R(f)$ d'après 2.8.10 (c). De même h commuterait avec f_{ss} et f_{nil}. Mais alors on aurait l'égalité

$$g - f_{\mathrm{ss}} = -(h - f_{\mathrm{nil}})$$

où $g - f_{\mathrm{ss}}$ est semi-simple et $-(h - f_{\mathrm{nil}})$ nilpotent (lemmes 2.8.11 et 2.8.12), ce qui entraîne $g - f_{\mathrm{ss}} = h - f_{\mathrm{nil}} = 0$ (lemme 2.8.13).

(d) En remplaçant $\mathbb{K}$ par une extension $\mathbb{L} \supset \mathbb{K}$ dans laquelle χ_f est scindé, la réduction de Jordan de la section 2.7 permet d'obtenir une matrice conjuguée $A' = P^{-1}AP \in \mathcal{M}_{n\times n}(\mathbb{L})$ qui est une matrice triangulaire de Jordan, ce qui donne une décomposition $A' = D' + N'$ avec une partie diagonale D' et une partie nilpotente N' qui commutent. Par suite

$$A = PA'P^{-1} = PD'P^{-1} + PN'P^{-1},$$

où $PD'P^{-1}, PN'P^{-1}$ commutent, $PD'P^{-1}$ est semi-simple et $PN'P^{-1}$ nilpotente. L'unicité de la décomposition dans $\mathcal{M}_{n\times n}(\mathbb{L})$ implique que l'on a bien $A_{\mathrm{ss}} = PD'P^{-1}$ et $A_{\mathrm{nil}} = PN'P^{-1}$. □

2.8.14. Remarque. Pour une matrice A triangulaire, la décomposition en parties semi-simple et nilpotente est bien plus subtile que le procédé qui consisterait à prendre la diagonale et la partie triangulaire restante. Par exemple, la matrice 2×2

$$A = \begin{pmatrix} \alpha & \gamma \\ 0 & \beta \end{pmatrix}$$

est digonalisable si $\beta \neq \alpha$, donc dans ce cas la partie nilpotente est nulle, et la partie semi-simple coïncide avec A.

2.8.15. Remarque. Sur $\mathbb{K} = \mathbb{R}$, la "décomposition"

$$\begin{pmatrix} \alpha & -\beta & u & w \\ \beta & \alpha & v & t \\ 0 & 0 & \alpha & -\beta \\ 0 & 0 & \beta & \alpha \end{pmatrix} = \begin{pmatrix} \alpha & -\beta & 0 & 0 \\ \beta & \alpha & 0 & 0 \\ 0 & 0 & \alpha & -\beta \\ 0 & 0 & \beta & \alpha \end{pmatrix} + \begin{pmatrix} 0 & 0 & u & w \\ 0 & 0 & v & t \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

n'est la décomposition de Jordan-Chevalley que si les matrices $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ et $\begin{pmatrix} u & v \\ w & t \end{pmatrix}$ commutent (ce qui est le cas par exemple si $\beta = 0$ ou si $w = -u$ et $t = u$). Un exercice recommandable est de chercher la véritable décomposition dans le cas où $\beta \neq 0$ et w, t sont quelconques.