Algèbre Linéaire 2

Comme le titre le suggère, le cours d'Algèbre Linéaire 2 forme une continuation du cours d'Algèbre Linéaire 1. Au delà de ce qui est fait dans ce cours précédent, on vise surtout une étudie approfondie des endomorphismes d'un espace vectoriel E de dimension finie, c'est-à-dire des applications linéaires $E \to E$. Puisque cette étude nécessite une certain aisance dans l'utilisation des nombreuses notions introduites dans le cours de AL1, on commencera avec un rappel des ces notions. Ensuite on introduira le cas relativement simple mais important des endomorphismes dits «diagonalisables». Bien que, du moins pour les espaces vectoriels sur le corps des nombres complexes, ce cas spécial couvre la plupart des endomorphismes, le cas général est plus compliqué. L'étude du cas général repose sur la considération des polynômes en l'endomorphisme considéré, et avant de l'aborder on aura besoin de développer un nombre de propriétés de l'anneau de polynômes sur un corps. Ainsi notre cours aura 4 chapitres:

Chapitre 1 Rappels des notions introduites en Algèbre Linéaire 1.
Chapitre 2 Diagonalisation : vecteurs propres et valeurs propres.
Chapitre 3 Polynômes à coefficients dans le corps K.
Chapitre 4 Réduction d'endomorphismes.

Ce document résume juste les définitions et résultats principaux, pour référence. Il ne peut dispenser des explications, motivations et exemples donnés dans le cours.

Chapitre 1. Rappels des notions introduites en Algèbre Linéaire 1.

1.1. *La notion d'espace vectoriel.*

Dans l'algèbre linéaire on utilise des valeurs de base de deux types: *scalaires* et *vecteurs*.

• *Le corps de base.*

Avant d'introduire des espaces vectoriels, il faut fixer l'ensemble des scalaires dont on se servira. Cet ensemble est le *corps de base* pour l'espace vectoriel, et est en général désigné par la lettre K.

Un corps est un ensemble de valeurs (nombres), qui contient au moins les valeurs particulières 0 et 1, qui est muni des lois internes d'addition, et multiplication, ayant 0 respectivement 1 comme élément neutre, et vérifiant toutes les propriétés algébriques habituelles (commutativité, associativité, distributivité). Pour chaque scalaire a, K contient aussi son opposé $-a$ (symétrique pour l'addition) et, si $a \neq 0$, son inverse a^{-1} (symétrique pour la multiplication).

1.1.1. Exemple. *L'ensemble* $\mathbf{R}$ *des nombre réels est un corps. L'ensemble* $\mathbf{Q}$ *des nombres rationnels, est également un corps car l'addition, soustraction, multiplication, ou division de deux nombres rationnels (sauf division par* 0*) est toujours possible, et donne toujours un nombre rationnel. L'ensemble* $\mathbf{C}$ *des nombres complexes forme aussi un corps, puisque tout nombre complexe non nul* $a + b\mathbf{i}$ *possède un inverse* $\frac{a}{a^2+b^2} - \frac{b}{a^2+b^2}\mathbf{i}$.

Dans ce cours on supposera pour simplicité que K est choisi parmi ces trois exemples de corps bien connus, $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$. Cependant il existe beaucoup d'autres corps, et les énoncés s'appliqueront sans modification aux espaces vectoriels avec de tels corps de base.

1.1.2. Exemple. *Il existe un corps noté* $\mathbf{F}_2$ *qui contient seulement deux éléments, notés* 0 *et* 1.

Dans $\mathbf{F}_2$ on a $1 + 1 = 0$, ce qui est plutôt inhabituel mais ne contredit pas les propriétés générales exigés d'un corps. Ce corps admet des espaces vectoriels particulièrement simples.

1.1 La notion d'espace vectoriel

• *Définition d'un espace vectoriel.*

On suppose désormais un corps K fixé.

1.1.3. Définition. *Un espace vectoriel sur K, ou K-espace vectoriel, est un ensemble E muni de deux opérations, l'addition (vectorielle) $E \times E \to E$ notée $(v, w) \mapsto v + w$ et la multiplication scalaire $K \times E \to E$ notée $(\lambda, v) \mapsto \lambda v$, et un élément particulier $0_E \in E$ (le vecteur nul de E), tels que les propriétés suivantes soient vérifiées :*

1. *[Associativité de l'addition vectorielle] $u + (v + w) = (u + v) + w$ pour tout $u, v, w \in E$.*
2. *[Commutativité de l'addition vectorielle] $v + w = w + v$ pour tout $v, w \in E$.*
3. *[Élément neutre pour l'addition vectorielle] $v + 0_E = v$ pour tout $v \in E$.*
4. *[Vecteur opposé] Tout $v \in E$ possède un vecteur opposé $-v \in E$, qui vérifie $v + -v = 0_E$.*
5. *[Associativité de la multiplication scalaire] $\lambda(\mu v) = (\lambda\mu)v$ pour tout $\lambda, \mu \in K$ et $v \in E$.*
6. *[Multiplication scalaire par $1 \in K$] $1v = v$ pour tout $v \in E$.*
7. *[Distributivité par rapport à + dans K] $(\lambda + \mu)v = \lambda v + \mu v$ pour tout $\lambda, \mu \in K$ et $v \in E$.*
8. *[Distributivité par rapport à + dans E] $\lambda(v + w) = \lambda v + \lambda w$ pour tout $\lambda \in K$ et $v, w \in E$.*

Cette liste sert dans la pratique surtout comme référence pour les propriétés dont on peut se servir librement dans les raisonnements et calculs concernant les espaces vectoriels.

1.1.4. Exemple. *Soit X est un ensemble quelconque (de nombres, de positions, de points, de symboles,...). Les vecteurs consisteront d'une affectation qui associe à chaque élément de X un scalaire. Formellement E est l'ensemble K^X des fonctions $X \to K$, et la valeur que $f : X \to K$ affecte à $x \in X$ est $f(x)$. L'addition et multiplication scalaire sont définies en considérant les valeurs associées aux différents x séparément: $f + g : x \mapsto f(x) + g(x)$ et $\lambda f : x \mapsto \lambda(f(x))$. Alors E est un espace vectoriel sur K, avec pour 0_E la fonction constante nulle $0_E : x \mapsto 0$.*

L'espace K^n des n-uplets $(x_1, \ldots, x_n)$ de scalaires, est un cas particulier de cette construction, dans lequel X est l'ensemble des n positions dans le n-uplet. D'autres cas sont l'espace des affectations de valeurs dans K à une liste d'inconnues, l'espace des matrices d'une taille donnée à coefficients dans K, ou l'espace des suites infinies de nombres de K (où on a $X = \mathbf{N}$).

• *Combinaisons linéaires.*

La notion de combinaison linéaire de vecteurs est fondamentale. Elle réunit l'addition et la multiplication scalaire en une notion, évitant ainsi souvent d'avoir à les considérer séparément.

1.1.5. Définition. *Si $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une famille de vecteurs de V, un vecteur $w \in E$ en est une combinaison linéaire s'il existe des scalaires $c_1, \ldots, c_k$ tels que $w = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$.*

On note que 0_E est une combinaison linéaire de n'importe quelle famille de vecteurs (il suffit de prendre tout les $c_i = 0$). C'est même vrai pour la famille vide $[\,]$ (donc avec $k = 0$), puisque par convention la somme vide de vecteurs de E vaut 0_E.

1.1.6. Définition. *Si S est une partie de E, une combinaison linéaire d'éléments de S est une combinaison linéaire d'une famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ de vecteurs pour laquelle $\mathbf{v}_1, \ldots, \mathbf{v}_k \in S$.*

Dans cette définition S peut être un ensemble fini ou infini. On remarque que dans le dernier cas elle évite toute mention de sommes infinies de vecteurs, celles-ci n'étant pas définies.

1.1.7. Proposition. *Toute combinaison linéaire d'éléments d'un ensemble S de combinaisons linéaires de $\mathbf{v}_1, \ldots, \mathbf{v}_k \in E$ est elle-même combinaison linéaire de $\mathbf{v}_1, \ldots, \mathbf{v}_k$.*

Les combinaison linéaires de $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ et de $\{\mathbf{v}_1, \ldots, \mathbf{v}_k\}$ sont les mêmes. Pour une famille infinie $[\mathbf{v}_1, \mathbf{v}_2, \ldots]$ on *définit* ses combinaisons linéaires comme celles de l'ensemble $\{\mathbf{v}_1, \mathbf{v}_2, \ldots\}$.

Attention au fait que le terme «combinaison linéaire» est parfois utilisé pour désigner une *expression* de la forme $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$, et non pas la valeur de cette expression (qui elle est un vecteur de E). La distinction est importante car deux expressions différentes peuvent avoir la même valeur. Ce sens du terme s'applique notamment dans la phrase «combinaison linéaire non triviale», qui désigne une telle expression dans laquelle au moins un des coefficients c_i est différent de $0 \in K$; cela n'empêche pas forcément la valeur de l'expression d'être 0_E.

1.2. *Sous-espaces vectoriels.*

En algèbre linéaire on considère souvent des *ensembles* de vecteurs plutôt que des vecteurs individuelles. Hormis les listes finies de vecteurs (qui seront appelés familles), il s'agit le plus souvent d'ensembles infinies mais d'une structure très simple, les sous-espaces vectoriels (s.e.v.).

1.2.1. Définition. *Une partie V d'un K-espace vectoriel E est un sous-espace vectoriel si*
1. $0_E \in V$;
2. $v + w \in V$ *pour tout* $v, w \in V$;
3. $\lambda v \in V$ *pour tout* $\lambda \in K$ *et* $v \in V$.

On pourra résumer cette définition en une seul condition: une partie V de E est s.e.v. de E si toute combinaison linéaire d'éléments de V est de nouveau élément de V.

1.2.2. Exemple. *L'ensemble $\{0_E\}$ (dit sous-espace nul) et l'ensemble E sont des s.e.v. de E.*

1.2.3. Exemple. *Pour toute famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ dans E, il découle de la proposition 1.1.7 que l'ensemble V des combinaisons linéaires de $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est un s.e.v. de E. Plus généralement les combinaisons linéaires d'une partie donnée S de E forment toujours un s.e.v. de E.*

Un autre type d'exemple de s.e.v. peut être décrit de façon informelle : si dans un espace E de fonctions (par exemple des affectations de valeurs scalaires à un certain nombre d'inconnues) on pose un système d'équations linéaires homogènes, alors l'ensemble de solutions du système forme un s.e.v. de E. Cet exemple repose essentiellement sur la propriété simple suivante :

1.2.4. Proposition. *Si V_1, V_2 sont des s.e.v. de E, alors leur intersection $V_1 \cap V_2$ est aussi un s.e.v. de E. Ceci reste vrai pour l'intersection d'un nombre plus grand (même infini) de s.e.v.*

Souvent des problèmes en algèbre linéaire font intervenir des s.e.v. dans leur formulation ou dans leur solution. Mais la notion peut aussi servir pour obtenir l'espace vectoriel lui-même:

1.2.5. Proposition. *Si V est un sous-espace vectoriel de E, alors V, muni des restrictions à V des opérations d'addition vectorielle et multiplication scalaire de E, et avec 0_E comme vecteur nul, forme lui-même un K-espace vectoriel.*

C'est cette proposition qui permet souvent de définir des espaces vectoriels sans vérification fastidieuse des conditions de la définition 1.1.3, à partir d'un K-espace (beaucoup) plus grand.

1.2.6. Exemple. *L'espace $K^{\mathbf{N}}$ des fonctions $\mathbf{N} \to K$ s'identifie à celle des suites infinies $(c_0, c_1, \ldots)$ de scalaires, une telle suite correspondant à la fonction $i \mapsto c_i$. Une telle suite est ultimement nulle si l'ensemble $\{\, i \in \mathbf{N} \mid c_i \neq 0 \,\}$ est fini. On vérifie facilement que la partie $V \subset K^{\mathbf{N}}$ des suites ultimement nulles forme un s.e.v. de $K^{\mathbf{N}}$, et V est donc un espace vectoriel lui-même. Cet espace V s'identifie à l'espace noté $K[X]$ des polynômes en une indéterminée X à coefficients dans K : si $(c_0, c_1, \ldots)$ est ultimement nulle il existe $N \in \mathbf{N}$ tel que $c_i = 0$ pour tout $i > N$, et la suite correspond au polynôme $c_0 + c_1X + c_2X^2 + \cdots + c_NX^N$. Si l'on fixe auparavant un $N \in \mathbf{N}$, les polynômes pour lesquels on peut utiliser cet N (qui sont donc de degré N au plus), forme un s.e.v. de $K[X]$, qui sera noté $K[X]_{\leq N}$, ou encore $K[X]_{<N+1}$.*

1.3. *Familles génératrices, relations, familles liées ou libres.*

• *Espace engendré, famille génératrice, dimension finie.*

1.3.1. Définition. *Si* $\mathbf{v}_1, \ldots, \mathbf{v}_k \in E$*, le sous-espace des combinaisons linéaires de* $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ *de l'exemple 1.2.3 est noté* $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$*. Il est appelé le s.e.v. engendré par* $\mathbf{v}_1, \ldots, \mathbf{v}_k$*.*

1.3.2. Proposition. *Un sous-espace* W *de* E *qui contient* $\mathbf{v}_1, \ldots, \mathbf{v}_k$*, contient aussi tous les vecteurs de* $V = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$*, et* V *est (dans ce sens fort) le plus petit de tous les s.e.v. de* E *qui contiennent les vecteurs* $\mathbf{v}_1, \ldots, \mathbf{v}_k$*.*

1.3.3. Définition. *Dans la définition précédente,* $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ *est appelé une famille génératrice du s.e.v.* $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$*. En particulier si* $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = E$ *alors elle est une famille génératrice de* E*. Cela veut dire que tout vecteur de* E *est combinaison linéaire de* $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$*.*

1.3.4. Exemple. *Dans l'espace* K^3 *des triplets de scalaires, les trois vecteurs* $\mathbf{e}_1 = (1, 0, 0)$*,* $\mathbf{e}_2 = (0, 1, 0)$ *et* $\mathbf{e}_3 = (0, 0, 1)$ *forment une famille génératrice de* E*, car* $(x, y, z) = x\mathbf{e}_1 + y\mathbf{e}_2 + z\mathbf{e}_3$ *pour tout* $x, y, z \in K$*. Il existe beaucoup d'autres familles génératrices de* K^3*, et en général une famille génératrice d'un certain espace vectoriel n'est presque jamais sa seule famille génératrice.*

On verra plus tard qui son on choisit au hasard trois vecteurs de K^3, ils forment la plupart du temps (mais pas toujours!) une famille génératrice de K^3, mais qu'aucune famille génératrice n'existe avec moins de 3 vecteurs. Par contre plein de familles génératrices plus grandes existent :

1.3.5. Propriété. *Si* $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ *est une famille génératrice de* V*, toute autre famille obtenue en rajoutant de nouveaux vecteurs de* V *à* $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ *sera aussi famille génératrice de* V*.*

C'est parce qu'on peut toujours associer aux vecteurs rajoutés des coefficients 0 pour étendre une combinaison linéaire de $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ à une combinaison linéaire de la famille étendue.

Pour un s.e.v. V (qui est en général infini), le fait d'en donner une famille génératrice $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une façon commode (car finie) de décrire V, à savoir $V = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. Mais attention : comme on a remarqué dans l'exemple, une telle famille génératrice, et donc l'écriture $V = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, est loin d'être unique.

1.3.6. Définition. *Un* K*-espace vectoriel qui admet (au moins) une famille génératrice finie est appelé un* K*-espace de dimension finie ; dans le cas contraire on parle d'un* K*-espace de dimension infinie.*

Dans ce cours on considérera presque exclusivement des espace vectoriels de dimension finie, ce qui simplifiera certaines considérations théoriques, et qui permettra surtout des méthodes explicites qui ne s'étendent pas aux espaces de dimension infinie. On pourra montrer que les espaces de fonctions K^X sont de dimension infinie si (et seulement si) X est infinie, et que même $K[X]$ (pourtant strictement un s.e.v. de $K^{\mathbf{N}}$) est de dimension infinie ; en revanche les espaces K^n pour $n \in \mathbf{N}$ sont de dimension finie (comme dans l'exemple de K^3 ci-dessus), ainsi que tous les s.e.v. de $K[X]$ de la forme $K[X]_{<n}$ (car $[1, X, X^2, \ldots, X^{n-1}]$ en est une famille génératrice).

• *Relations de dépendance entre vecteurs; familles liées et familles libres.*

Non seulement une famille génératrice pour un (sous) espace vectoriel donné n'est pas unique, il s'avère même pratiquement impossible de désigner parmi ces familles une de « famille génératrice préférée ». Mais de l'autre côté, certaines familles génératrices sont en tant que telle clairement redondantes. Si par exemple le dernier vecteur $\mathbf{v}_k$ est déjà combinaison linéaire des vecteurs $[\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}]$ qui le précèdent, alors $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k) = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1})$ et la famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est redondant comme famille génératrice. Cela mène à la notion d'une famille liée.

1.3.7. Proposition/Définition. *Pour une famille $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ de vecteurs de E, les conditions suivantes sont équivalentes, et si elle sont vérifiées on appelle $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ une famille liée :*

(i) Parmi $\mathbf{v}_1, \dots, \mathbf{v}_k$, un vecteur (au moins) est combinaison linéaire des autres, c'est-à-dire il existe i tel que $\mathbf{v}_i \in \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_k)$.

(ii) Il existe des coefficients $c_1, \dots, c_k \in K$, dont au moins un coefficient non nul, tels que $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = 0_E$.

L'équation du point (ii) est appelée une relation (de dépendance linéaire) entre $\mathbf{v}_1, \dots, \mathbf{v}_k$; c'est une combinaison linéaire non triviale de ces vecteurs dont la valeur est le vecteur nul. L'équivalence énoncée dans la proposition découle du fait qu'une équation de la forme $\mathbf{v}_i = c_1\mathbf{v}_1 + \cdots + c_{i-1}\mathbf{v}_{i-1} + c_{i+1}\mathbf{v}_{i+1} + \cdots + c_k\mathbf{v}_k$ donne une relation $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = 0_E$ dans laquelle on a $c_i = -1 \neq 0$, et que d'autre part toute relation $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = 0_E$ avec $c_i \neq 0$ peut être rendue dans cette forme après division de toute l'équation par $-c_i$.

Comme on veut plutôt éviter la redondance, c'est la notion contraire est plus intéressante :

1.3.8. Définition. *Une famille est libre (ou linéairement indépendante) si elle n'est pas liée.*

On note que contrairement á la définition 1.3.3, ces définitions ne mentionnent pas de s.e.v.

1.3.9. Proposition. *Une famille $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ de vecteurs est libre si et seulement si pour des scalaires inconnues $c_1, \dots, c_k$, la condition $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = 0_E$ entraîne $c_1 = c_2 = \cdots = c_k = 0$.*

Une autre façon de exprimer cette caractérisation d'une famille libre est que la combinaison linéaire triviale est *la seule* combinaison linéaire dont la valeur est le vecteur nul.

1.3.10. Proposition. *Si $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ est une famille libre, chaque $v \in \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ s'écrit comme combinaison linéaire $v = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ pour un unique k-uplet $(c_1, \dots, c_k)$.*

Si on avait deux telle expressions différentes, leur soustraction donnerait une relation entre les vecteurs $c_1, \dots, c_k$, en contradiction avec leur indépendance linéaire.

Plus une famille de vecteurs est nombreuse, plus il est difficile d'être libre. La famille vide est toujours libre ; une famille à un seul vecteur est libre sauf si c'est le vecteur nul ; une famille de deux vecteurs est libre sauf si l'un est un multiple scalaire de l'autre. Mais au delà de deux vecteurs, il ne suffit pas de faire une exception pour de tels cas très simples : il faut exclure *tous* les cas où l'un des vecteurs est combinaison linéaire des autres (1.3.7 (i)). On verra que dans un espace de dimension finie, au delà d'un certain nombre de vecteurs, *toute* famille devient liée.

L'idée qu'une famille liée $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ est redondante pour décrire $\mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ suggère qu'on peut rendre une telle famille libre par la suppression de certains vecteurs (redondants), et cela sans changer le s.e.v. engendré par la famille. C'est effectivement le cas.

1.3.11. Proposition. *Si $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ est une famille quelconque, alors il existe une famille libre extraite d'elle qui engendre le s.e.v. $V = \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k)$. Formellement, il existe $i_1, \dots, i_m$ avec $1 \leq i_1 < \cdots < i_m \leq k$ tels que $[\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_m}]$ est une famille libre et $\mathrm{Vect}(\mathbf{v}_{i_1}, \dots, \mathbf{v}_{i_m}) = V$.*

Cela se montre par récurrence sur k. L'idée est que pour chaque indice i on est dans l'un de deux cas de figure : soit (1) $\mathbf{v}_i \in \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_{i-1})$ et on peut se passer de $\mathbf{v}_i$, soit (2) $\mathbf{v}_i \notin \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_{i-1})$ et la famille libre extraite de $[\mathbf{v}_1, \dots, \mathbf{v}_{i-1}]$ restera libre si l'on y rajoute $\mathbf{v}_i$. Seulement dans le cas (2) on retient i dans la liste d'indices $i_1, \dots, i_m$. On a donc une *procédure* qui produit une suite extraite comme décrit dans la proposition. La suite produite a même une propriété supplémentaire, a savoir que chaque famille *initiale* $[\mathbf{v}_1, \dots, \mathbf{v}_i]$ de $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ partage la propriété qu'elle engendre le même s.e.v. que la famille libre qui est extraite d'elle. (Avec cette condition supplémentaire, la famille extraite devient unique.)

1.4. *Bases, coordonnées, dimension.*

• *Base d'un espace vectoriel.*

1.4.1. Définition. *Une base de E est une famille génératrice de E qui est aussi libre.*

La même définition peut être utilisée pour un s.e.v. V de E ; une base de V est donc une famille de vecteurs de V qui est à la fois génératrice de V et libre. Elle décrit V sans redondance. Comme pour les familles génératrices, une base n'est pas unique ; ne dites jamais *la* base de E.

1.4.2. Propriété. *Une famille $[\mathbf{b}_1, \ldots, \mathbf{b}_n]$ de vecteurs de E forme une base de E si et seulement si, d'une part tout vecteur $v \in E$ s'écrit comme une combinaison linéaire de cette famille (elle est génératrice de E), et si d'autre part cette écriture est unique (la famille est libre).*

Si E est de dimension finie, alors il existe par définition une famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ telle que $E = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$. En appliquant la proposition 1.3.11 pour ce cas, on obtient comme conclusion qu'il existe une *base* de E, qui est extraite de la famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$. On conclut :

1.4.3. Fait. *Tout K-espace de dimension finie possède (au moins) une base finie.*

• *Théorème de la base incomplète.*

Une partie d'une famille libre (et en particulier d'une base) est toujours libre. Si $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une famille libre de E, il est donc possible de poser la question analogue à celle dont la proposition 1.3.11 parle pour les familles génératrices, à savoir, s'il est possible de *rajouter* des vecteurs à cette famille pour obtenir une base de E. Puisque pour établir l'existence d'une base tout court de E on a du supposer que E est de dimension finie, on va maintenir cela comme hypothèse. Dans ce cas on montre effectivement le théorème suivant.

1.4.4. Théorème de la base incomplète. *Si dans un K-espace vectoriel E de dimension finie, $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une famille libre, alors on peut choisir dans E des vecteurs supplémentaires $\mathbf{v}_{k+1}, \ldots, \mathbf{v}_n$ de sorte que $[\mathbf{v}_1, \ldots, \mathbf{v}_n]$ soit une base de E.*

Voici une esquisse de démonstration. D'abord on choisit une famille génératrice $[\mathbf{w}_1, \ldots, \mathbf{w}_l]$ de E, qui existe car E est de dimension finie. Ensuite on forme la famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k, \mathbf{w}_1, \ldots, \mathbf{w}_l]$, qui est aussi génératrice (propriété 1.3.5). On peut alors appliquer à cette famille la même méthode utilisée dans la proposition 1.3.11, à savoir passer au crible ses vecteurs de gauche à droite, rejetant chaque vecteur qui se trouve dans le s.e.v. engendré par les vecteurs avant lui. Le fait que $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une famille libre implique qu'on ne rejette aucun de ses vecteurs. La base obtenue à la fin de la procédure est donc bien une extension de $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$, comme voulu.

• *Coordonnées.*

La notion de base est fondamentale en algèbre linéaire, à cause de la propriété 1.4.2.

1.4.5. Définition. *Si $\mathcal{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_n]$ est une base de E et $v \in E$ un vecteur, les coordonnées de v par rapport à $\mathcal{B}$ sont les $c_1, \ldots, c_n \in K$ tels que $v = c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$; le n-uplet $(c_1, \ldots, c_n)$ existe et est unique d'après la propriété 1.4.2. Pour $i \in \{1, \ldots, n\}$, la fonction $V \to K$ associant à v sa coordonnée c_i par rapport à $\mathcal{B}$, est appelé i-ème fonction coordonnée pour à la base $\mathcal{B}$.*

Dans cette définition le n-uplet $(c_1, \ldots, c_n)$ est un élément de K^n. La base $\mathcal{B}$ établit donc une correspondance entre le K-espace vectoriel E (espace dont on ignore la nature de

ses vecteurs) et l'espace particulier K^n de n-uplet, et qui peut être décrite ainsi. Il existe une application $K^n \to E$ définie par la formation de combinaisons linéaires par rapport à $\mathcal{B}$; c'est $(c_1, \ldots, c_n) \mapsto c_1\mathbf{b}_1 + \cdots + c_n\mathbf{b}_n$. Comme $\mathcal{B}$ est une base de E, cette application est *bijective*, et l'expression en coordonnées par rapport à $\mathcal{B}$ est son application réciproque $E \to K^n$.

La correspondance d'expression en coordonnées (et la fait que tout espace vectoriel de dimension finie possède une base, ce qui découle de la proposition 1.3.11) permet en quelque sorte de ramener tout K-espace de dimension finie à un espace particulier de la forme K^n. Mais on fera attention au fait que la correspondance dépend de la base $\mathcal{B}$ utilisée. Quand on passe d'une base à une autre (et il y aura des situations où cela facilite la compréhension d'un problème) les coordonnées pour un même vecteur changeront ; en verra plus tard comment.

• *Base canonique dans les espaces particuliers K^n et $K[X]_{<n}$.*

La plupart du temps, on ne saura conclure qu'une famille donnée dans E forme une base de E qu'après avoir vérifie (éventuellement par un calcul) soit qu'elle est à la fois génératrice et libre, soit (et au fond cela revient au même) qu'elle possède la propriété 1.4.2. Mais dans certains cas, le rapport entre les coefficients d'une combinaison linéaire et le vecteur qui en résulte est tellement simple que cette dernière propriété est évidente. Par exemple dans K^3 on a vu que pour des scalaires x, y, z on a $x(1,0,0) + y(0,1,0) + z(0,0,1) = (x,y,z)$. Donc avec $\mathbf{e}_1 = (1,0,0)$, $\mathbf{e}_2 = (0,1,0)$ et $\mathbf{e}_3 = (0,0,1)$ il est clair que tout $v \in K^3$ s'écrit d'une façon unique comme combinaison linéaire de $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$. Cela montre donc que $[\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3]$ est une base de K^3.

Pour généraliser cet exemple aux autres espaces K^n, il est utile d'introduire une notation connu comme le «symbole de Kronecker».

1.4.6. Définition. *Si $i, j \in X$ pour un certain ensemble X, on définit $\delta_{i,j} \in \{0,1\} \subseteq K$ par*

$$\delta_{i,j} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Ainsi les vecteurs $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ ci-dessus peuvent être décrit par une seule formule : ils sont donnés par $\mathbf{e}_i = (\delta_{i,1}, \delta_{i,2}, \delta_{i,3})$ pour $i = 1, 2, 3$. La propriété fondamentale de cette notation est

1.4.7. Propriété. *Pour toute fonction $f : X \to K$ et tout $i \in X$ on a : $\sum_{j \in X} \delta_{i,j} f(j) = f(i)$.*

1.4.8. Exemple. *Dans K^n les vecteurs $\mathbf{e}_1, \ldots, \mathbf{e}_n$ donnés par $\mathbf{e}_i = (\delta_{i,1}, \ldots, \delta_{i,n})$ ont la propriété que tout $v \in K^n$ s'écrit d'une façon unique comme combinaison linéaire de $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$. En effet $(c_1, \ldots, c_n) = c_1\mathbf{e}_1 + \cdots + c_n\mathbf{e}_n$. Donc $[\mathbf{e}_1, \ldots, \mathbf{e}_n]$ est une base de K^n, dite base canonique.*

La base canonique $\mathcal{E}$ de K^n a la propriété que l'expression en coordonnées par rapport à $\mathcal{E}$ établit une correspondance $K^n \to K^n$ qui est l'identité : pour tout vecteur $v \in K^n$, le n-uplet de ces coordonnées par rapport á E est égal à v lui-même (qui est un n-uplet).

1.4.9. Exemple. *Dans $K[X]_{<n}$, les vecteurs $[\mathbf{e}_0, \ldots, \mathbf{e}_{n-1}]$ donnés par $\mathbf{e}_i = X^i$ (avec $X^0 = 1 \in K[X]$ et $X^1 = X$) ont la propriété que tout $v \in K[X]_{<n}$ s'écrit d'une façon unique comme combinaison linéaire de $[\mathbf{e}_0, \ldots, \mathbf{e}_{n-1}]$. En effet $c_0X^0 + \cdots + c_{n-1}X^{n-1} = c_0\mathbf{e}_0 + \cdots + c_{n-1}\mathbf{e}_{n-1}$. Donc $[\mathbf{e}_0, \ldots, \mathbf{e}_{n-1}] = [X^0, \ldots, X^{n-1}]$ est une base de $K[X]_{<n}$ dite base canonique de $K[X]_{<n}$.*

On fera attention au fait que seulement dans ces espaces très particuliers (et dans encore quelques autres qui sont du même genre) on saura définir une base canonique ; dans d'autres K-espaces vectoriels le terme «base canonique» n'a aucun sens. On remarque d'ailleurs que même dans ces espaces qui possèdent une base canonique, leurs sous-espaces n'en ont pas.

• *Dimension.*

On avait remarqué que dans un espace vectoriel de dimension finie, au delà d'un certain nombre de vecteurs *toute* famille devient liée. En effet on peut énoncer de façon plus précise le suivant.

1.4.10. Lemme. *Si $V = \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, toute famille de plus de k vecteurs de V est liée.*

On note qu'aucun lien entre cette autre famille et $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ n'est supposé, juste qu'elle contient $l > k$ vecteurs. On prouvera ce lemme important en réduisant la question au fait qu'un système d'équations homogènes linéaires avec plus d'inconnues que d'équations admet toujours une solution non triviale, un fait qui découle directement de la méthode de Gauss pour résoudre des systèmes linéaires (car le nombre de pivots ne peut jamais dépasser le nombre initial d'équations, donc pour des systèmes de ce type il restera forcément des colonnes sans pivot).

Soit $[\mathbf{w}_1, \dots, \mathbf{w}_l]$ la famille de $l > k$ vecteurs de V. Puisque $w_j \in V = \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, il existe des scalaires $a_{1,j}, \dots, a_{k,j}$ (pas forcément uniques) tels que $w_j = a_{1,j}\mathbf{v}_1 + \cdots + a_{k,j}\mathbf{v}_k$. En faisant cela pour $j = 1, 2, \dots, l$, les coefficients $a_{i,j}$ forment une matrice A de taille $k \times l$. On peut maintenant réécrire toute combinaison linéaire $x_1\mathbf{w}_1 + \cdots + x_l\mathbf{w}_l$ des $\mathbf{w}_j$ comme une combinaison linéaire $y_1\mathbf{v}_1 + \cdots + y_k\mathbf{v}_k$ en prenant $\vec{y} = A \cdot \vec{x}$ (ici $\vec{x}$ et $\vec{y}$ sont les vecteurs colonnes avec $x_1, \dots, x_l$ respectivement $y_1, \dots, y_k$ comme coefficients) car le coefficient y_i de $\mathbf{v}_i$ dans $\sum_{j=1}^{l} x_j\mathbf{w}_j$ est $\sum_{j=1}^{l} a_{i,j}x_j$, qui est le coefficient i du produit matriciel $A \cdot \vec{x}$. Une condition suffisante pour que $x_1\mathbf{w}_1 + \cdots + x_l\mathbf{w}_l = 0$ est donc que $A \cdot \vec{x} = 0$, et l'existence d'une solution $\vec{x} \neq 0$ de ce système de k équations linéaires homogènes en l inconnues $x_1, \dots, x_l$ prouvera que $[\mathbf{w}_1, \dots, \mathbf{w}_l]$ est une famille liée. Mais puisque $l > k$, l'existence d'une telle solution est assurée.

1.4.11. Théorème. *Si $[\mathbf{b}_1, \dots, \mathbf{b}_n]$ est une base de E, alors toute autre base de E contient aussi précisément n vecteurs.*

C'est une conséquence immédiate du lemme. D'une part, puisque $E = \mathrm{Vect}(\mathbf{b}_1, \dots, \mathbf{b}_n)$, aucune famille avec plus de n vecteurs n'est libre, et donc pas une base non plus. D'autre part si E avait une base $[\mathbf{b}'_1, \dots, \mathbf{b}'_m]$ avec $m < n$, alors on raisonne en inversant les rôles : le fait $E = \mathrm{Vect}(\mathbf{b}'_1, \dots, \mathbf{b}'_m)$ forcerait les familles de plus de m vecteurs à être liées, mais cela contredit le fait que $[\mathbf{b}_1, \dots, \mathbf{b}_n]$ est libre. Donc une base de E ne peut avoir ni plus ni moins de n vecteurs.

Puisque tout espace de dimension finie possède au moins une base, on peut formuler la définition fondamentale suivante.

1.4.12. Définition. *Si E est de dimension finie, sa dimension $\dim E$ est définie comme le nombre n de vecteurs dans une base de E. Ce nombre ne dépend pas de la base choisie.*

Une fois la dimension $d = \dim V$ d'un (sous-)espace vectoriel V connue, il est plus facile de vérifier si une famille de vecteurs de V est une base de V. D'une part le nombre de vecteurs doit être d pour pouvoir être une base. D'autre part, si la famille contient d vecteurs, il n'est plus nécessaire de vérifier séparément que la famille soit libre *et* génératrice ; l'un des deux suffit :

1.4.13. Proposition. *Si E est une espace vectoriel, et $\dim E = n$, alors*
(1) Toute famille libre de n vecteurs est aussi génératrice (et donc une base).
(2) Toute famille génératrice de n vecteurs est aussi libre (et donc une base).

Pour (1), la famille libre s'étend à une base d'après le théorème 1.4.4, mais celle-ci ne pouvant pas contenir plus de n vecteurs, la famille de départ doit déjà être cette base. Pour (2), on peut extraire de la famille génératrice une base d'après la proposition 1.3.11, mais celle-ci ne pouvant pas contenir moins de n vecteurs, la famille de départ doit déjà être cette base.

On peut combiner cette proposition 1.4.13 avec et la définition 1.4.12 de la dimension, sous la forme d'un énoncé qui prend la forme de «trois conditions pour le prix de deux».

1.4.14. Proposition. *Soit E est un espace vectoriel, et $\mathbf{v}_1, \dots, \mathbf{v}_n \in E$. Alors pour chaque choix de deux parmi les trois condition suivantes, ces deux entraînent la troisième condition.*
(1) $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ *est une famille génératrice de E, c'est-à-dire* $E = \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_n)$.
(2) $[\mathbf{v}_1, \dots, \mathbf{v}_n]$ *est une famille libre.*
(3) $\dim(E) = n$.

1.5. *Applications linéaires.*

Pour une grande partie l'algèbre linéaire est concernée avec des applications entre des espaces vectoriels. Parmi les très nombreuses applications possibles, seulement les applications *linéaires* sont compatibles avec la structure d'un espace vectoriel, et on se limite donc à l'étude de ces applications. Les applications linéaires sont suffisamment limitées pour qu'on puisse (du moins dans le cas d'espaces de dimension finie) complètement décrire une application linéaire par une quantité finie de données (une matrice), ce qui n'est pas le cas pour les applications plus générales. En même temps les applications linéaires peuvent manifester une grande variété de comportements. (Et entre espaces réels, elles permettent du moins d'*approcher localement* toutes les fonctions qui sont "lisses", mais cela est plus une sujet pour l'analyse que pour l'algèbre.)

• *Définition et constructions.*

1.5.1. Définition. *Si E, F sont deux K-espaces vectoriels (qui peuvent être confondus), une application $f : E \to F$ est dite K-linéaire (le plus souvent abrégé «linéaire») si*
1. *$f(x + y) = f(x) + f(y)$ pour tout $v, w \in E$, et*
2. *$f(\lambda v) = \lambda f(v)$ pour tout $\lambda \in K$ et $v \in E$.*

En itérant la définition on voit qu'une application linéaire est compatible avec des combinaisons linéaires

$$f(\lambda_1 \mathbf{v}_1 + \cdots + \lambda_k \mathbf{v}_k) = \lambda_1 f(\mathbf{v}_1) + \cdots + \lambda_k f(\mathbf{v}_k), \tag{1}$$

et cette égalité caractérise les applications linéaires. Pour une application linéaire $f : E \to F$ on a toujours $f(0_E) = 0_F$ (prendre $\lambda = 0$ dans 1.5.1:2, ou considérer la famille vide dans (1)).

1.5.2. Proposition. *Soit $[\mathbf{b}_1, \dots, \mathbf{b}_n]$ une base de E, et $[\mathbf{w}_1, \dots, \mathbf{w}_n]$ une famille quelconque du même nombre de vecteurs de F. Alors il existe une et une seule application linéaire $f : E \to F$ telle que $f(\mathbf{b}_i) = \mathbf{w}_i$ pour $i = 1, \dots, n$.*

Le fait que $[\mathbf{b}_1, \dots, \mathbf{b}_n]$ est une base veut dire que tout $v \in E$ s'écrit de façon unique $v = \lambda_1 \mathbf{b}_1 + \cdots + \lambda_n \mathbf{b}_n$ pour $\lambda_1, \dots, \lambda_n \in K$, et alors la linéarité de f exige que

$$f(v) = \lambda_1 f(\mathbf{b}_1) + \cdots + \lambda_n f(\mathbf{b}_n) = \lambda_1 \mathbf{w}_1 + \cdots + \lambda_n \mathbf{w}_n$$

Ainsi f est entièrement déterminé par les valeurs $f(\mathbf{b}_i) = \mathbf{w}_i$, et on vérifie facilement que la fonction f ainsi définie est en effet linéaire.

Les applications linéaires sont ainsi facile à définir. En plus il y a diverses constructions qui permettent de construire de nouveaux applications linéaires à partir d'autres. Si $f, g : E \to F$ sont des applications linéaires et $\lambda \in K$, alors on peut définir des applications $f + g, \lambda f : E \to F$ par, pour tout $v \in E$:

$$(f + g)(v) = f(v) + g(v),$$
$$(\lambda f)(v) = \lambda . f(v),$$

qui sont aussi linéaires. Ainsi l'ensemble noté $\mathcal{L}(E, F)$ des applications linéaires $E \to F$ devient lui-même un K-espace vectoriel, car les conditions de la définition 1.1.3 sont vérifiées.

L'autre construction important pour fabriquer des applications linéaires est la composition. Si $f : E \to F$ et $g : F \to G$ sont des applications linéaires entre K-espaces vectoriels E, F, G, alors leur composée $g \circ f$ (prononcé « g rond f », ou plus parlant « g après f ») est définie par

$$(g \circ f)(v) = g(f(v)) \qquad \text{pour } v \in E$$

et c'est une application linéaire $E \to G$.

1.5.3. Exemple. *Pour tout K-espace vectoriel, application identique $\mathbf{I}_E : E \to E$ est linéaire.*

1.5.4. Exemple. *Si $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une famille de vecteurs de E, l'application $l : K^k \to E$ qui forme des combinaisons linéaires, c'est-à-dire $l : (c_1, \ldots, c_k) \mapsto c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$, est linéaire.*

1.5.5. Exemple. *Si $V \subseteq E$ est un sous-espace, l'application $g : V \to E$ vérifiant $g(v) = v$ pour tout $v \in V$ est linéaire. On l'appelle l'inclusion de V dans E, et on écrit parfois $g : V \hookrightarrow E$.*

1.5.6. Définition. *Si $f \in \mathcal{L}(E, F)$ et V est un s.e.v. de E, la restriction $f|_V$ de f à V est l'application linéaire $f \circ g \in \mathcal{L}(V, F)$ où $g : V \hookrightarrow E$ est l'inclusion de V dans E.*

• *Image d'une application linéaire.*

En étudiant les applications linéaires $E \to F$, on est souvent amené à considérer certains sous-espaces de E ou de F. C'est en fait la raison principale pour laquelle on s'est tellement intéressé aux sous-espaces vectoriels.

1.5.7. Définition. *L'image d'une application linéaire $f : E \to F$, notée $\mathrm{Im}(f)$ ou $f[E]$, est le sous-ensemble $\{\, f(v) \mid v \in E \,\}$ de F. Il s'agit dans tous les cas d'un sous-espace vectoriel de F.*

Pour que $w \in F$ soit dans l'image de f, il faut qu'il existe $v \in E$ avec $f(v) = w$, mais ce vecteur v n'est pas forcément unique.

1.5.8. Exemple. *Pour $\mathbf{v}_1, \ldots, \mathbf{v}_k \in E$, le sous-espace $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ de E n'est autre que l'image $\mathrm{Im}(l)$ de l'application linéaire $l : K^k \to E$ de l'exemple 1.5.4.*

On rappelle qu'en général une application $f : X \to Y$ entre ensembles est appelé surjective si pour tout $y \in Y$ il existe au moins un x (dit antécédent de y) tel que $f(x) = y$. On a alors :

1.5.9. Fait. *Une application linéaire $f : E \to F$ est surjective si et seulement si $\mathrm{Im}(f) = F$.*

On fera attention au fait que la notion de surjectivité de f utilise explicitement l'espace F vers lequel f est défini (son codomaine). Cela peut poser problème, en vue du fait qu'on se permet parfois d'utiliser le même nom pour des fonctions avec des codomaines différentes ; par exemple si on sait que $\mathrm{Im}(f)$ est contenu dans un sous-espace W de f, on dira parfois de considérer f comme un application linéaire $E \to W$, ce qui change son codomaine. Pour être précis, il faudra avoir une notation de "restriction à l'arrivée" pour distinguer ce nouveau f de l'ancienne application, mais une telle notation n'existe pas vraiment (ou n'est pas standardisée). En tout cas le codomaine doit être clair chaque fois qu'on parle d'une application surjective.

• *Noyau d'une application linéaire.*

1.5.10. Définition. *Le noyau d'une application linéaire $f : E \to F$, notée* $\mathrm{Ker}(f)$*, est le sous-ensemble* $\{ v \in E \mid f(v) = 0_F \}$ *de E. Il s'agit dans tous les cas d'un sous-espace vectoriel de E.*

La notation pour les noyaux vient du terme anglais "kernel" pour noyau. Le noyau permet de reconnaître si une application linéaire est injective ou non. On rappelle qu'une application $f : X \to Y$ est injective si tout $y \in Y$ possède *au plus* un antécédent $x \in X$ (donc avec $f(x) = y$).

1.5.11. Proposition. *Une application linéaire $f : E \to F$ est injective si et seulement si* $\mathrm{Ker}(f) = \{0_E\}$ *(c'est-à-dire si 0_E est le seul antécédent de 0_F).*

Si $v, v' \in E$ sont des antécédents d'un même $w \in F$, il s'agit de montrer que $v = v'$. Par linéarité $f(v - v') = w - w = 0_F$, et l'hypothèse $\mathrm{Ker}(f) = \{0_E\}$ permet de conclure $v - v' = 0_E$.

1.5.12. Exemple. *L'application $l : K^k \to E$ de l'exemple 1.5.4 est injective si et seulement si la famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est libre, et elle est surjective si et seulement si elle est génératrice de E.*

• *Isomorphisme de K-espaces.*

Si $f \in \mathcal{L}(E, F)$ est à la fois injectif et surjectif, alors f est bijectif et admet donc une application réciproque $F \to E$ qui comme d'habitude sera notée f^{-1}. Cette réciproque est encore linéaire :

1.5.13. Proposition. *Si $f \in \mathcal{L}(E, F)$ est une application bijective, alors $f^{-1} \in \mathcal{L}(F, E)$.*

Si $w, w' \in F$ et $\lambda, \mu \in K$ on pose $v = f^{-1}(w)$ et $v' = f^{-1}(w')$, donc $w = f(v)$ et $w' = f(v')$, et $f^{-1}(\lambda w + \mu w') = f^{-1}(\lambda f(v) + \mu f(v')) = f^{-1}(f(\lambda v + \mu v')) = \lambda v + \mu v' = \lambda f^{-1}(w) + \mu f^{-1}(w')$ en utilisant la linéarité de f, ce qui prouve que f^{-1} est linéaire $F \to E$.

1.5.14. Définition/Proposition. *Si E, F sont des K-espaces vectoriels, un isomorphisme de K-espaces $E \to F$ est un $f \in \mathcal{L}(E, F)$ qui admet une réciproque $g \in \mathcal{L}(F, E)$, c'est-à-dire telle que $g \circ f = \mathbf{I}_E$ et $f \circ g = \mathbf{I}_F$. Toute application linéaire qui est bijective est un isomorphisme.*

Cette dernière caractérisation est une conséquence de la proposition 1.5.13, car les conditions $g \circ f = \mathbf{I}_E$ et $f \circ g = \mathbf{I}_F$ veulent dire qu'on ne peut que prendre $g = f^{-1}$. Mais la définition est donnée sous la forme initiale car il est fondamental pour toute notion d'isomorphisme (et il en existe dans beaucoup de domaines autres que l'algèbre linéaire) d'admettre un isomorphisme inverse. Aussi il peut arriver d'avoir $f \in \mathcal{L}(E, F)$ et $g \in \mathcal{L}(F, E)$ pour lesquels la vérification directe de $g \circ f = \mathbf{I}_E$ et $f \circ g = \mathbf{I}_F$ est plus simple que d'établir autrement la bijectivité de f.

1.5.15. Proposition. *La composée de deux isomorphismes est un isomorphisme.*

1.5.16. Définition. *Deux K-espaces vectoriels E, F sont isomorphes s'il existe (au moins) un isomorphisme $E \to F$.*

On voit à l'aide de la proposition que être isomorphe est un relation d'équivalence. Un isomorphisme $E \to F$ permet de traduire toute question d'algèbre linéaire dans E en une question correspondante dans F, et vice versa (par l'isomorphisme inverse). Pour cette raison on considère deux espace isomorphes comme équivalents du point de vue de l'algèbre linéaire.

1.5.17. Exemple. *L'application linéaire $l : K^k \to E$ de l'exemple 1.5.4 est bijective, et donc un isomorphisme, si et seulement si la famille $[\mathbf{v}_1, \ldots, \mathbf{v}_k]$ est une base de E. Dans ce cas l^{-1} est l'application $E \to K^k$ qui exprime un vecteur en coordonnées sur cette base. Cette application est donc linéaire, comme le sont aussi les fonctions coordonnées individuelles $v \mapsto c_i : E \to K$.*

Un isomorphisme entre deux espace permet de traduire toute question d'algèbre linéaire qui concerne l'un des espaces en une question équivalente qui concerne l'autre. Comme illustration de ce principe on peut énoncer.

1.5.18. Fait. *Si l'on sait que $f \in \mathcal{L}(E, F)$ est injectif ou surjectif, il en est de même pour sa composée à gauche ou à droite avec un isomorphisme de K-espaces vectoriels.*

En appliquant cela a ce qui est dit dans l'exemple 1.5.12, on peut donc généraliser :

1.5.19. Conséquence. *Soit $f \in \mathcal{L}(E, F)$ et $\mathcal{B}$ une base de E. La famille $[\, f(b) \mid b \in \mathcal{B}\,]$ est une famille libre si et seulement si f est injectif, et elle est une famille génératrice de F si et seulement si f est surjectif. Elle est donc une base de F si et seulement si f est un isomorphisme.*

L'exemple 1.5.17 montre qu'un K-espace vectoriel E de dimension finie est isomorphe à K^n pour $n = \dim E$, et 1.5.19 montre qu'un isomorphe $E \to F$ ne peut exister que si $\dim(F) = n$.

1.5.20. Conclusion. *Deux espaces E, F de dimension finie sont isomorphes si et seulement si* $\dim(E) = \dim(F)$. *Pour un isomorphisme $f : E \to F$ et s.e.v. V de E on a* $\dim(f[V]) = \dim(V)$.

1.6. *Le calcul matriciel.*

Le calcul matriciel est souvent utile pour trouver une réponse à certains questions concrètes d'algèbre linéaire. Une matrice est simplement un objet qui (comme un n-uplet dans K^n) regroupe plusieurs scalaires, qu'on organise en une grille rectangulaire formée de lignes et de colonnes. Diverses opérations et méthodes algorithmiques sont définies sur les matrices.

1.6.1. Définition. *Une matrice M à coefficients dans K consiste en la donnée d'un nombre $n \in \mathbf{N}$ dit de lignes, un nombre $m \in \mathbf{N}$ dit de colonnes, et pour chaque couple d'entiers (i, j) avec $0 < i \leq n$ et $0 < j \leq m$ un scalaire $M_{i,j} \in K$. L'ensemble de telles matrices pour lesquelles les nombres n, m sont fixés est noté* $\mathrm{Mat}_{n,m}(K)$, *et ces matrices sont appelées des matrices (de taille) $n \times m$ à coefficients dans K. L'ensemble* $\mathrm{Mat}_{n,n}(K)$ *des matrices carrés de taille n est d'un intérêt particulier, et est pour cette raison souvent abrégé* $\mathrm{Mat}_n(K)$.

La ligne i de M est l'élément $(M_{i,1}, \ldots, M_{i,m}) \in K^m$ (pour un $i \in \{1, \ldots, n\}$ fixé) et la colonne j de M est l'élément $(M_{1,j}, \ldots, M_{n,j}) \in K^n$ (pour un $j \in \{1, \ldots, m\}$ fixé).

Un matrice dont tous les coefficients sont nuls est appelée matrice nulle, et l'unique matrice nulle dans $\mathrm{Mat}_{n,m}(K)$ sera notée $\mathbf{0}_{n,m}$, ou simplement $\mathbf{0}$ si sa taille est claire dans le contexte.

1.6.2. Définition. *Pour $n \in \mathbf{N}$, la matrice identité notée $\mathbf{I}_n$ est la matrice $M \in \mathrm{Mat}_{n,n}(K)$ telle que $M_{i,j} = \delta_{i,j}$ pour tout $i, j \in \{1, \ldots, n\}$.*

1.6.3. Définition. *La transposée de la matrice $A \in \mathrm{Mat}_{n,m}(K)$ est la matrice ${}^tA \in \mathrm{Mat}_{m,n}(K)$ avec ${}^tA_{j,i} = A_{i,j}$ pour $0 < i \leq n$ et $0 < j \leq m$ (réflection par rapport à la diagonale principale).*

• *Opérations matricielles.*

∘ *Opérations d'un espace vectoriel.*

On peut interpréter $\mathrm{Mat}_{n,m}$ comme l'espace de fonctions K^I où $I = \{1,\dots,n\} \times \{1,\dots,m\}$ est l'ensemble rectangulaire de positions dans ces matrices. Ce point de vue, bien que limité, permet déjà de munir $\mathrm{Mat}_{n,m}$ d'une structure de K-espace vectoriel, c'est-à-dire d'une addition et une multiplication scalaire (et le vecteur nul sera $\mathbf{0}_{n,m}$).

1.6.4. Définition. *La somme de matrices $A, B \in \mathrm{Mat}_{n,m}$ est la matrice $C \in \mathrm{Mat}_{n,m}$ dont le coefficients sont donnés par $C_{i,j} = A_{i,j} + B_{i,j}$ pour tout i, j. Pour $\lambda \in K$ le multiple scalaire λA de A est la matrice $M \in \mathrm{Mat}_{n,m}$ dont le coefficients sont $M_{i,j} = \lambda A_{i,j}$ pour tout i, j.*

Ainsi chaque $\mathrm{Mat}_{n,m}$ devient un K-espace vectoriel, pour lequel $\dim(\mathrm{Mat}_{n,m}) = nm$.

∘ *Produit matriciel.*

Pour les matrices divers produits sont définis (mais l'opération qui consiste à multiplier en chaque position séparément comme pour l'addition de matrices ne s'avère pas très utile). On saura multiplier un matrice $A \in \mathrm{Mat}_{n,m}(K)$ à gauche avec un élément de K^n, ou à droite avec un élément de K^m. Mais les deux seront aussi obtenus comme des cas particuliers d'un produit de deux matrices, donc on définira toute de suite cette opération plus générale.

1.6.5. Définition. *Le produit $A \cdot B$ de deux matrices est défini seulement si le nombre de colonnes de A est égal au nombre de lignes de B, donc quand $A \in \mathrm{Mat}_{n,m}(K)$ et $B \in \mathrm{Mat}_{m,l}(K)$ pour certains $n, m, l \in \mathbf{N}$. Dans ce cas le produit est la matrice $C \in \mathrm{Mat}_{n,l}(K)$ telle que $C_{i,j} = \sum_{t=1}^{m} A_{i,t}B_{t,j} = A_{i,1}B_{1,j} + A_{i,2}B_{2,j} + \cdots + A_{i,m}B_{m,j}$ pour tout $0 < i \leq n$ et $0 < j \leq l$.*

Chaque coefficient du produit $A \cdot B$ est donc le produit scalaire d'une ligne de A et d'une colonne de B, où «produit scalaire» de deux m-uplets $(x_1,\dots,x_m), (y_1,\dots,y_m) \in K^m$ veut dire le scalaire $x_1y_1 + x_2y_2 + \cdots + x_my_m$. Plus exactement, le coefficient en position i, j est le produit scalaire de la ligne i de A et la colonne j de B. Plusieurs coefficients du résultat utilisent la même ligne de A, ou la même colonne B, ce qui nous mène au suivant

1.6.6. Constat.

(1) La ligne i d'un produit matriciel $A \cdot B$ ne dépend de A que par sa ligne i : c'est la combinaison linéaire des lignes de B formée en utilisant les coefficients de la ligne i de A.

(2) La colonne j d'un produit matriciel $A \cdot B$ ne dépend de B que par sa colonne j de la matrice B : c'est la combinaison linéaire des colonnes de A formée en utilisant les coefficients de la colonne j de B.

Le point (2) donne une indication qu'on pourra définir la multiplication d'une matrice $A \in \mathrm{Mat}_{n,m}(K)$ avec un vecteur $v \in K^m$ en considérant ce dernier comme une colonne de B ; la colonne correspondante du produit $A \cdot B$ est un élément de K^n qui donnera notre produit $A \cdot v$. C'est en particulier ce qu'on obtient pour le produit matriciel dans le cas $l = 1$ où B n'a qu'une colonne, et on pourrait définir $A \cdot v$ en identifiant le façon évidente l'espace K^m avec celui $\mathrm{Mat}_{m,1}$ des matrices à une seule colonne. Mais dans ce cours on veut éviter de faire une telle identification juste pour faire l'économie d'une définition, qu'on donne donc explicitement :

1.6.7. Définition. *Le produit d'une matrice A et un vecteur $v \in K^m$ est défini si le nombre de colonnes de A est m. Pour $A \in \mathrm{Mat}_{n,m}(K)$ et $v = (v_1,\dots,v_m)$, ce produit $A \cdot v$ est l'élément $w = (w_1,\dots,w_n) \in K^n$ dont les coefficients sont $w_i = \sum_{j=1}^{m} A_{i,j}v_j$ pour $i = 1,\dots,n$. De façon équivalente w est la combinaison linéaire des colonnes de A avec comme coefficients $v_1,\dots,v_m$.*

On peut également définir le produit de (dans cet ordre) $u \in K^n$ et $B \in \mathrm{Mat}_{n,m}(K)$, en utilisant 1.6.6:(1). Ceci fait interpréter u comme un matrice à une *ligne*, et le résultat sera un

élément de K^m (le rôles de n, m sont inversés) qui est la combinaison linéaire des lignes de B avec comme coefficients ceux de u. Bien que cette méthode de définition est symétrique avec celle de la définition 1.6.7, les deux produits sont assez différents, et dans la pratique on est obligé d'en privilégier un. La tradition majoritaire est d'utiliser le produit de la définition 1.6.7, et c'est qu'on fera dans ce cours. Une justification pour cette préférence pourrait être que dans un produit $A \cdot v$ on associera A à une application qui opère sur v, comme dans la notation $f(x)$.

En fait, si on fixe $A \in \mathrm{Mat}_{n,m}(K)$, ce produit détermine une application $K^m \to K^n$ donnée par $v \mapsto A \cdot v$. Puisque cette application f_A utilise (les coefficients de) son argument v pour former une combinaison linéaire, comme dans l'exemple 1.5.4, il s'agit d'une application linéaire : $f_A \in \mathcal{L}(K^m, K^n)$. Prenant pour v de vecteurs de la base canonique de K^m, on voit que $f_A(\mathbf{e}_j)$ est égal à la colonne j de A. D'après la proposition 1.5.2, ces m valeurs particulières de f_A (qui sont les m colonnes de A) déterminent f totalement en tant qu'application linéaire. Puisque réciproquement tout m-uplet de vecteurs de K^n forme le jeu de colonnes d'une unique matrice de $\mathrm{Mat}_{n,m}(K)$, on a ainsi associé chaque application linéaire $K^m \to K^n$ à une et une seule matrice de $\mathrm{Mat}_{n,m}(K)$. C'est le fondement de la représentation matricielle des applications linéaires.

1.6.8. Définition. *Pour $n, m \in \mathbf{N}$, l'application $\mathrm{Mat}_{n,m}(K) \to \mathcal{L}(K^m, K^n)$ qui associe à la matrice $A \in \mathrm{Mat}_{n,m}(K)$ l'application $v \mapsto A \cdot v$ est appelée l'isomorphisme canonique de $\mathrm{Mat}_{n,m}(K)$ avec $\mathcal{L}(K^m, K^n)$. Les colonnes de A sont les images de la base canonique de K^m.*

Le fait que la correspondance entre $\mathrm{Mat}_{n,m}(K)$ et $\mathcal{L}(K^m, K^n)$ est linéaire (compatible avec leurs structures d'espace vectoriel) est peu surprenant. Par contre le fait suivant n'est pas anodin, et fournit une justification pour la définition un peu étrange du produit matriciel.

1.6.9. Proposition. *Sous les isomorphismes canoniques, le produit matriciel correspond à la composition d'applications linéaires. Explicitement, si $A \in \mathrm{Mat}_{n,m}(K)$ et $B \in \mathrm{Mat}_{m,l}(K)$ sont des matrices pour lesquelles le produit $A \cdot B$ est défini, et $f_A \in \mathcal{L}(K^m, K^n)$ et $f_B \in \mathcal{L}(K^l, K^m)$ sont les application linéaires y associées sous les isomorphismes canoniques respectifs, alors $f_A \circ f_B$ est l'application linéaire $f_{A \cdot B} : K^l \to K^n$ associée à $A \cdot B$ par l'isomorphisme canonique.*

Pour montrer la proposition, il suffit (d'après la proposition 1.5.2) de vérifier que $f_{A \cdot B}$ coïncide avec $f_A \circ f_B$ quand elle est appliquée à un vecteur $\mathbf{e}_j$ de la base canonique de K^l. Or $f_{A \cdot B}(\mathbf{e}_j)$ est la colonne j de $A \cdot B$, qui d'après le constat 1.6.6:(2) et la définition 1.6.7 est égale à $A \cdot b_j$ où b_j désigne la colonne j de B. Mais cette dernière est égale à $f_B(\mathbf{e}_j)$, donc on trouve $f_{A \cdot B}(\mathbf{e}_j) = A \cdot b_j = A \cdot f_B(\mathbf{e}_j) = f_A(f_B(\mathbf{e}_j)) = (f_A \circ f_B)(\mathbf{e}_j)$, comme voulu.

1.6.10. Corollaire. *Le produit matriciel est associatif, c'est-à-dire on a $A \cdot (B \cdot C) = (A \cdot B) \cdot C$ dès que l'un ou l'autre des deux membres est défini.*

La composition de fonctions est toujours associative, car quand elles sont définies, les composées $f \circ (g \circ h)$ et $(f \circ g) \circ h$ font toutes deux $x \mapsto f(g(h(x)))$. En appliquant la proposition 1.6.9 à ce fait pour la composition d'applications linéaires, on voit que le produit matriciel lui aussi doit être associatif. Bien sûr, cela peut être vérifié par un calcul explicite (mais un peu fastidieux).

∘ *Matrices inversibles.*

1.6.11. Proposition. *La matrice $\mathbf{I}_n \in \mathrm{Mat}_{n,n}(K)$ correspond à l'identité $\mathbf{I}_{K^n} \in \mathcal{L}(K^n, K^n)$ par l'isomorphisme canonique.*

Il suffit d'observer que pour tout $j \in \{1, \ldots, n\}$ la colonne j de $\mathbf{I}_n$ est le vecteur $\mathbf{e}_j \in K^n$. Par conséquent, $I_n \cdot v = v$ pour tout $v \in K^n$, et on a $\mathbf{I}_n.A = A$ ainsi que $A \cdot \mathbf{I}_n = A$ dès que le produit matriciel en question est bien défini (car $\mathbf{I}_{K^n}$ a cette propriété pour la composition).

1.6.12. Définition. *Une matrice A est inversible s'il existe une matrice B telle que les produits $A \cdot B$ et $B \cdot A$ sont chacun une matrice identité. Dans ce cas B est unique, et est appelé l'inverse de A, noté A^{-1}. Par symétrie B est aussi inversible, et $A = B^{-1}$.*

Cette définition est formulée de telle sorte que, comparant avec la définition 1.5.14, il soit immédiat que A est inversible si et seulement si il correspond à un isomorphisme de K-espaces vectoriels. Puisque K^m et K^n ne sont pas isomorphes si $n \neq m$, on conclut qu'une matrice ne peut être inversible que si elle est carrée (le même nombre de lignes et colonnes), et donc les deux matrices identité $A \cdot B$ et $B \cdot A$ de la définition doivent être en fait la même matrice $\mathbf{I}_n$.

1.6.13. Proposition. *Seulement les matrices carrées peuvent être inversibles. Or, pour $A, B \in \mathrm{Mat}_n(K)$, on aura $B = A^{-1}$ dès que l'une des conditions $A \cdot B = \mathbf{I}_n$ ou $B \cdot A = \mathbf{I}_n$ est vérifiée.*

Supposons par exemple $B \cdot A = \mathbf{I}_n$ (l'autre cas est symétrique car $B = A^{-1}$ et $A = B^{-1}$ sont équivalents). Cela veut dire $f_B \circ f_A = \mathbf{I}_{K^n}$, et implique que f_A est injectif (si $f_A(v) = f_A(w)$ alors $v = f_B(f_A(v)) = f_B(f_A(w)) = w$). L'exemple 1.5.12 et la proposition 1.4.13 (qu'on peut appliquer grâce au fait que A est carrée) montrent que f_A est aussi surjectif, donc inversible, et A l'est aussi. Une fois que ceci est établi, on peut écrire $B = B \cdot A \cdot A^{-1} = \mathbf{I}_n \cdot A^{-1} = A^{-1}$.

1.6.14. Proposition. *Si $A, B \in \mathrm{Mat}_n(K)$ sont tous les deux inversibles, alors $A \cdot B$ est aussi inversible, et $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$.*

• *Matrice d'une application linéaire.*

Seulement les applications linéaires $K^m \to K^n$ correspondent *canoniquement* (c'est-à-dire d'une façon qui ne dépend d'aucun choix) à une matrice (de taille $n \times m$). Pour d'autres K-espaces E, F de dimension finie on peut aussi représenter chaque application linéaire $f \in \mathcal{L}(E, F)$ par une matrice, mais à condition de fixer au préalable des bases $\mathcal{E}$ de E, et $\mathcal{F}$ de F. On l'appellera la matrice de f par rapport à ces bases, et dans ce cours on la notera ${}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ (le placement des bases dans cette notation est un peu spécial mais voulu suggestif : la base $\mathcal{E}$ au départ est à droite où viendra le vecteur argument, et la base $\mathcal{F}$ à l'arrivée est du côté gauche « du résultat »).

Le procédé pour associer cette matrice à f est simple. Les bases $\mathcal{E}$ et $\mathcal{F}$ déterminent des isomorphismes $E \leftrightarrow K^m$ et $F \leftrightarrow K^n$. En composant f avec ces isomorphismes dans le sens convenable on obtient un application $K^m \to E \xrightarrow{f} F \to K^n$, et ${}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ est la matrice canoniquement associée à cette application linéaire composée. Concrètement, la composée appliquée à $x \in K^m$ forme d'abord la combinaison linéaire des vecteurs de $\mathcal{E}$ à coefficients donnés par x, puis applique f a ce vecteur, et finalement trouve les coordonnées par rapport à $\mathcal{F}$ du résultat.

1.6.15. Fait. *La matrice ${}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ de $f \in \mathcal{L}(E, F)$ par rapport aux bases $\mathcal{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m]$ de E et $\mathcal{F} = [\mathbf{f}_1, \ldots, \mathbf{f}_n]$ de F est la matrice de taille $n \times m$ dont pour $j \in \{1, \ldots, m\}$ la colonne j est formée des coordonnées de $f(\mathbf{e}_j)$ par rapport à la base $\mathcal{F}$.*

• *Systèmes d'équations en leur utilisation en algèbre linéaire.*

La solution concrète de beaucoup de problèmes en algèbre linéaire repose sur la résolution d'un système d'équations linéaires. Pour cela on connaît la méthode (du pivot) de Gauss, dont ici on ne fera que résumer la conclusion.

Pour un système de n équations en m inconnues, on peut renommer (si besoin) les inconnues $x_1, \ldots, x_m$, et les interpréter comme les composantes d'un seul vecteur inconnu $x \in K^m$. Alors le système s'écrit comme une seule équation vectorielle de la forme $Ax = b$, où $A \in \mathrm{Mat}_{n,m}(K)$

regroupe les coefficients des inconnues, et $b \in K^n$ regroupe les seconds membres des équations ; c'est la forme matricielle du système d'équations. La méthode de Gauss consiste á transformer le système en un système équivalent et *échelonné*. Dans sa forme matricielle $A'x = b'$, la propriété d'être échelonné veut dire la chose suivante de la matrice A', où dans chaque ligne on appelle « pivot » la position du premier coefficient non nul (s'il existe) : (1) Tous les pivots sont dans des colonnes différentes, et chacun se trouve à droite des pivots (éventuels) des lignes précédentes ; (2) les éventuelles lignes sans pivot (donc entièrement nulles) viennent après toutes les lignes avec pivot. Le fait que le système échelonné est équivalent au système initial est assuré par le fait que l'effet de la méthode de Gauss est de multiplier les deux membres de l'équation par une même matrice inversible M (c'est-à-dire $A' = M \cdot A$ et $b' = M \cdot b$), d'où on pourra retrouver l'équation originale de la nouvelle en multipliant par M^{-1}.

Si dans l'équation échelonnée il existe une ligne de A' sans pivot pour laquelle le coefficient correspondant b'_i de b' n'est pas nul, alors cela correspond à une équation $0 = b'_i$ qui est impossible satisfaire, et le système n'aura pas de solution (il est contradictoire). On exclut désormais ce cas. Alors dans les cas restants, si A' contient des lignes sans pivot, elles correspondent à des équations triviales $0 = 0$ qu'on peut enlever du système sans changer ses solutions. Ce qui reste est une matrice $A'' \in \mathrm{Mat}_{n',m}(K)$ avec $n' \leq n$, dont chaque ligne contient un pivot, et ces pivots sont dans des colonnes distinctes. Soit $P \subseteq \{1, \ldots, m\}$ l'ensemble des indices de colonnes contenant un pivot (les inconnues x_j pour $j \in P$ sont appelées les « variables primaires » du système), et $S = \{1, \ldots, m\} \setminus P$ les indices de colonnes sans pivot (les inconnues x_j pour $j \in S$ sont les « variables secondaires » du système). Les valeurs des variables secondaires ne sont pas contraintes par le système, et peuvent donc être choisies librement. En fait les solutions du système sont paramétrées par ces valeurs, et une fois choisies les valeurs de variables primaires peuvent être déduites . Chaque ligne donne une équation qui permet de résoudre l'inconnue correspondant à la colonne de son pivot, une fois les valeurs des inconnues après elle sont connues ; les variables primaires peuvent donc être résolues de la dernière à la première.

Un système homogène ne peut être contradictoire. S'il a plus d'inconnues que d'équations, on a $S \neq \emptyset$ ci-dessus, et il admet des solutions non nulles, complétant la preuve du lemme 1.4.10.

La méthode de Gauss permet donc de résoudre complètement une équation $A \cdot x = b$, autrement dit de trouver tous les antécédents $x \in K^n$ de $b \in K^m$ pour l'application linéaire $f_A : K^m \to K^n$ correspondant à A. Pour trouver plus généralement pour $f \in \mathcal{L}(E, F)$ les antécédents $v \in E$ d'un vecteur $w \in F$ donné, il suffit de choisir des bases $\mathcal{E}, \mathcal{F}$ dans E, F, de trouver $A = {}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ et les coordonnées $b \in K^m$ de w par rapport à $\mathcal{F}$, de résoudre $A \cdot x = b$, et finalement d'interpréter ces solutions x comme coordonnées par rapport à $\mathcal{E}$ de vecteurs $v \in E$.

∘ *Inverser une matrice.*

Le problème de trouver l'inverse d'une matrice carrée $A \in \mathrm{Mat}_n(K)$ (si elle existe) peut être considéré comme celui de trouver pour chaque vecteur $\mathbf{e}_j$ de la base canonique de K^n son antécédent par f_A, c'est-à-dire la solution (supposée unique, sinon A n'est pas inversible) de $A \cdot x = \mathbf{e}_j$, car cet x forme la colonne j de A^{-1}. Cette idée donne la méthode suivante.

1.6.16. Méthode pour trouver l'inverse (éventuelle) d'une matrice $A \in \mathrm{Mat}_n(K)$. *On forme l'« équation à second membre multiple » $A \cdot x = \mathbf{I}_n$ et y applique la méthode de Gauss. En fait c'est chaque fois une colonne individuelle $\mathbf{e}_j$ de $\mathbf{I}_n$ qui est le vrai second membre, mais comme le second membre joue un rôle passif dans la phase d'échelonnement de la méthode de Gauss (aucun pivot n'y est choisi) on peut traiter les n cas au même temps par le maintien de n colonnes dans la partie à droite. Si jamais dans une colonne à gauche on ne peut pas trouver de pivot, cela montre que A n'était pas inversible et on peut s'arrêter. Sinon, on trouvera à*

gauche une forme échelonnée triangulaire sas coefficients diagonaux nuls. On continue ensuite la réduction par opérations sur les lignes, pour rendre les coefficients des pivots à la valeur 1*, et en les utilisant pour rendre* 0 *les autres coefficients dans leur colonne. Ainsi la matrice à gauche est devenu* $\mathbf{I}_n$*, à quel moment la matrice à droite (celle des* n *«seconds membres») est devenu* A^{-1}*.*

Cette méthode maintient donc à chaque instant un couple de matrices carrées $(M \mid N)$, qui est initialement $(A \mid \mathbf{I}_n)$, et procède en multipliant chaque fois les deux à gauche par une même matrices inversible. Comme le produit de matrices inversibles est inversible, le transformation entière du couple $(A \mid \mathbf{I}_n)$ en $(M \mid N)$ est elle aussi celle d'une multiplication à gauche par une même matrice inversible, et visiblement cette matrice n'est autre que N. Quand finalement cela a rendu $M = N \cdot A$ égale à $\mathbf{I}_n$, on aura donc $N = A^{-1}$. Pour le calcul manuel, il est utile de noter la relation $M = N \cdot A$ qui doit rester valable à tout moment ; elle est relativement facile à tester, et permet ainsi une détection précoce d'éventuelles erreurs de calcul.

∘ *Noyau et image.*

Mis à part le calcul d'antécédents et d'inverses, les applications de la méthode de Gauss en algèbre linéaire vont toujours concerner exclusivement le cas d'équations homogènes, donc dans le reste de cette section les seconds membres des équations sont toujours nuls, et on les mentionnera plus.

1.6.17. Méthode pour trouver une base du noyau de $f \in \mathcal{L}(E,F)$**.** *Soit* $\mathcal{E} = [\mathbf{e}_1, \dots, \mathbf{e}_m]$ *et* $\mathcal{F} = [\mathbf{f}_1, \dots, \mathbf{f}_n]$ *des bases de* E *respectivement de* F *(les choisir si besoin), et* $A = {}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$*. On résout l'équation vectorielle* $A \cdot x = 0$ *pour* $x \in K^m$*. Si* $S \subseteq \{1, \dots, m\}$ *est l'ensemble d'indices des colonnes sans pivot, chaque* $j \in S$ *contribue un vecteur* $v^{(j)} \in E$ *à la base cherchée, comme suit. On trouve la solution particulière* $x^{(j)}$ *dans laquelle on prend* $x^{(j)}_k = \delta_{j,k}$ *pour* $k \in S$ *; les valeurs des* $x^{(j)}_i$ *restants, les variables premières, sont déduites comme d'habitude des valeurs de ces variables secondaires. Alors* $v^{(j)}$ *est le vecteur* $x^{(j)}_1\mathbf{e}_1 + \cdots + x^{(j)}_m\mathbf{e}_m \in E$ *dont les coordonnées par rapport à* $\mathcal{E}$ *sont données par* $x^{(j)}$*. Finalement* $[v^{(j)}]_{j\in S}$ *est une base de* $\mathrm{Ker}(f)$*.*

On remarque que si $A \cdot x = 0$ a un solution unique ($P = \emptyset$), alors c'est forcément la solution $x = 0$ (l'équation étant homogène), et $\ker(f) = \{0\}$; la méthode trouve alors la base vide.

Il n'est pas difficile de voir que ces vecteurs $v^{(j)}$ vérifient $f(v^{(j)}) = 0_F$, et forment une base de $\mathrm{Ker}(f)$, car c'est essentiellement le sens d'avoir trouvé une solution complète de $A \cdot x = 0$. Il est un peu plus surprenant que le même travail permet de trouver aussi une base de $\mathrm{Im}(f)$. La famille $[f(\mathbf{e}_j), \dots, f(\mathbf{e}_j)]$ de vecteurs de F est certainement génératrice de $\mathrm{Im}(f)$, mais elle peut être liée. En fait chaque solution $x^{(j)} \in K^m$ ci-dessus vérifie $A \cdot x^{(j)} = 0_{K^n}$, et les scalaires $x^{(j)}_1, \dots, x^{(j)}_m$ sont donc les coefficients d'une relation entre les colonnes de A, une combinaison linéaire dont la valeur est nulle. Puisque $x^{(j)}_j = 1$ et $x^{(j)}_k = 0$ pour tout autre $k \in S$, cette relation permet d'exprimer la colonne j de A en terme de ces colonnes indicées par le complémentaire P de S. Ainsi les colonnes indicées par S sont redondantes, et $[f(\mathbf{e}_j)]_{j\in P}$ est encore une famille génératrice de $\mathrm{Im}(f)$. Le calcul de $\mathrm{Ker}(f)$ montre qu'une relation entre les colonnes de A est triviale dès que ses coefficients indicés par S sont nuls, donc cette famille est une base de $\mathrm{Im}(f)$.

1.6.18. Méthode pour trouver une base de l'image de $f \in \mathcal{L}(E,F)$**.** *Comme dans la méthode 1.6.17 on fixe des bases* $\mathcal{E} = [\mathbf{e}_1, \dots, \mathbf{e}_m]$ *de* E *et* $\mathcal{F}$ *de* F*, et pour* $A = {}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ *on applique la méthode de Gauss à l'équation* $A \cdot x = 0$ *pour trouver l'ensemble* $P \subseteq \{1, \dots, m\}$ *d'indices des colonnes qui contiennent un pivot. Alors* $[f(\mathbf{e}_j)]_{j\in P}$ *est une base de* $\mathrm{Im}(f)$*.*

On remarque que ce n'est pas la seule méthode qui existe pour trouver une base de $\mathrm{Im}(f)$. Une autre méthode, inspirée par celle utilisée pour la proposition 1.3.11, est de échelonner les

colonnes de la matrice A, en inversant les rôles des lignes et des colonnes dans la méthode de Gauss, ce qui permet de façon relativement simple de décider pour chaque nouvelle colonne si elle est dans le sous-espace engendré par les colonnes précédentes (et donc redondante) ou non. En fait l'échelonnement peut être vu comme une méthode de remplacer une famille de vecteurs dans K^n par une famille plus commode mais qui engendre le même sous-espace. L'échelonnement des colonnes de A ne fournit donc pas seulement (après suppression les colonnes nulle produites) une base de $\mathrm{Im}(f)$, mais une base simplifiée qui simplifie le test si $v \in \mathrm{Im}(f)$ pour $v \in K^n$.

∘ *Changement de présentation d'un sous-espace.*

Un sous-espace vectoriel peut être présenté essentiellement de deux façons : soit sous la forme $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ par une famille génératrice, soit comme solution d'un système d'équations linéaires homogènes. Dans les deux cas des modifications aux données (vecteurs ou équations) sont possibles qui ne changent pas le s.e.v. désigné. Par exemple l'échelonnement d'un système d'équations le simplifie, et on vient de voir qu'on peut aussi échelonner une famille génératrice.

Pour diverses raisons on peut être amené a passer d'une présentation à une autre., Le sous-espace défini par un système d'équations sous forme matricielle $A \cdot x = 0$ est simplement $\mathrm{Ker}(f_A)$, donc la méthode 1.6.17 effectue la transition d'un système d'équations vers une famille génératrice (qui sera en fait une base). Le passage réciproque, trouver un système d'équations pour décrire un sous-espace donné par des vecteurs générateurs, est moins souvent employé, mais peut être utile (par exemple parce que dans K^n, les s.e.v. de dimension $n-1$ ont besoin de $n-1$ vecteurs générateurs, mais sont aussi définis par une seule équation). Mais la méthode pour le faire est quasiment le même, par l'observation suivante. Une équation homogène $c_1x_1 + \cdots + c_nx_n = 0$ peut être représentée la matrice à une ligne $C = (c_1 \ \ c_2 \ \cdots \ c_n)$ de ces coefficients (c'est comme cela que les équations d'un système apparaissent dan sa forme matricielle). Si on prend les vecteurs générateurs du s.e.v. V de K^n comme colonnes d'un matrice $A \in \mathrm{Mat}_{n,m}$ (c'est-à-dire on écrit $V = \mathrm{Im}(f_A)$), alors la condition que l'équation est vérifiée pour tous ces vecteurs s'écrit $C \cdot A = \mathbf{0} \in \mathrm{Mat}_{1,m}(K)$. On peut interpréter C comme ligne d'inconnues, et trouver tous les valeurs C pour qui $C \cdot A = \mathbf{0}$ est vrai, se fait par la même méthode (de Gauss) utilisée pour résoudre $A \cdot x = 0_{K^m}$, sauf que les rôles des lignes et des colonnes sont inversés.

∘ *Trouver une base pour l'intersection de sous-espaces.*

Un intersection de deux s.e.v. de E est toujours un s.e.v., mais sa description est assez implicite ; on voudrait souvent en trouver une base. Trois cas de figure se présentent.

(1) Les deux s.e.v. sont chacun donné par un système d'équations. Dans ce cas on peut simplement combiner tous les équations, ce qui donne une système plus grand, quel système définit l'intersection des sous-espaces. On peut résoudre ce nouveau système par la méthode 1.6.17.

(2) L'un des deux espaces est donné par un système d'équations $A \cdot x = 0$, et l'autre comme $V = \mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$, par une famille génératrice. Dans ce cas on peut former un combinaison linéaire $v = c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k$ avec des inconnues pour les coefficients c_i, et considérer l'équation $A \cdot v = 0$ qui décrit l'intersection comme une nouvelle équation en ces inconnues. Si on résout cette équation, alors pour $(c_1, \ldots, c_k)$ parcourant une famille génératrice de l'espace de solutions, les vecteurs v correspondants forment une famille génératrice de l'intersection cherchée. Il est souhaitable de s'assurer au préalable (en supprimant des vecteurs si besoin, ce qui ne fait que simplifier le reste du travail) que la famille $\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_k)$ soit une *base* d V, car dans ce cas on est sûr d'obtenir une base de l'intersection. Concrètement, si $B \in \mathrm{Mat}_{n,k}$ est la matrice dont les $\mathbf{v}_j$ sont les colonnes, le système à résoudre est $(A \cdot B) \cdot c = 0$, et pour chaque solution $c \in K^k$ dans une base de l'espace $\mathrm{Ker}(A \cdot B)$ des solutions, le vecteur $B \cdot c$ est un vecteur de la base de l'intersection.

(3) Les deux s.e.v. sont chacun donné par une famille génératrice. Dans ce cas la meilleure approche semble de trouver un système d'équations pour l'un des deux (de préférence celui avec le plus grand nombre de vecteurs générateurs, car cela donne moins d'équations), et ainsi de se ramener au cas (2). C'est notamment plus facile que l'approche directe de chercher des vecteurs qui sont à la fois des combinaisons linéaires de l'une et de l'autre famille, car cela introduit beaucoup d'inconnues, et un grand système d'équations à résoudre.

• *Changement de bases.*

Parfois il est souhaitable de changer d'une représentation de vecteurs en coordonnées par rapport à une base $\mathcal{E}$ vers une représentation en coordonnées par rapport à une autre base $\mathcal{E}'$, notamment quand cela rend un problème plus transparent (cela se produira notamment quand on parlera plus tard de la diagonalisation d'une endomorphisme). Déjà le problème d'exprimer dans K^n des vecteurs dans une autre base que la base canonique est de cette nature, car par nature les vecteurs de K^n sont exprimé par rapport à la base canonique.

L'idée pour le faire est assez simple : pour exprimer $f \in \mathcal{L}(E,F)$ sous forme matricielle, on a besoin indépendemment d'une base de E et d'une base de F ; cela rend possible d'utiliser deux bases distinctes même dans le cas où $E = F$. Pour isoler l'effet du changement de base, on le fait pour $f = \mathbf{I}_E$ en sorte qu'on obtienne les nouvelles coordonnées du *même* vecteur qu'au départ. Si $x \in K^n$ donne les coordonnées de $v \in E$ par rapport à l'ancienne base $\mathcal{E}$, les coordonnées de v par rapport à $\mathcal{E}'$ sont données par ${}_{\mathcal{E}'}\operatorname{Mat}_{\mathcal{E}}(\mathbf{I}_E) \cdot x$. Comme toujours, chaque colonne de ${}_{\mathcal{E}'}\operatorname{Mat}_{\mathcal{E}}(\mathbf{I}_E)$ contient les coordonnées par rapport de la base $\mathcal{E}'$ à l'arrivée de l'image d'un vecteur de la base $\mathcal{E}$ au départ. Comme $f = \mathbf{I}_E$, cette image est le vecteur lui-même, donc la matrice donne les coordonnées des vecteurs de $\mathcal{E}$ par rapport à la base $\mathcal{E}'$. Mais dans la pratique on aura souvent décrit la nouvelle base $\mathcal{E}'$ par les coordonnées de ces vecteurs par rapport à $\mathcal{E}$, et ces coordonnées se trouvent dans la matrice *inverse* ${}_{\mathcal{E}}\operatorname{Mat}_{\mathcal{E}'}(\mathbf{I}_E)$. La convention veut qu'on appelle *matrice de passage* de $\mathcal{E}$ vers $\mathcal{E}'$ cette dernière matrice $P = {}_{\mathcal{E}}\operatorname{Mat}_{\mathcal{E}'}(\mathbf{I}_E)$. On en conclut que la conversion de coordonnées pour la base $\mathcal{E}$ vers celles pour la base $\mathcal{E}'$ se fait en multipliant par l'inverse P^{-1} de la matrice de passage de $\mathcal{E}$ vers $\mathcal{E}'$ (aussi étrange que cela peut paraître).

Le changement de base affecte aussi les matrices d'applications linéaires qui se servent de cette base au départ ou à l'arrivée. La formule découle de l'informelle «règle de dominos», qui dit que dans un produit de matrices, la bases doivent être les même au point de contact. Donc si pour $f \in \mathcal{L}(E,F)$ on change à l'arrivée (dans F) on change de la base $\mathcal{B}$ vers $\mathcal{B}'$, la nouvelle matrice est donnée par ${}_{\mathcal{B}'}\operatorname{Mat}_{\mathcal{E}}(f) = {}_{\mathcal{B}'}\operatorname{Mat}_{\mathcal{B}}(\mathbf{I}_F) \cdot {}_{\mathcal{B}}\operatorname{Mat}_{\mathcal{E}}(f)$ et si on change au départ (dans E) de la base $\mathcal{E}$ vers $\mathcal{E}'$, la nouvelle matrice est donnée par ${}_{\mathcal{B}}\operatorname{Mat}_{\mathcal{E}'}(f) = {}_{\mathcal{B}}\operatorname{Mat}_{\mathcal{E}}(f) \cdot {}_{\mathcal{E}}\operatorname{Mat}_{\mathcal{E}'}(\mathbf{I}_E)$.

Si l'on reconnaît le rôle des matrices dans cette formule, on peut résumer ainsi, en n'oubliant pas le cas important de $f \in \mathcal{L}(E,E)$, avec le même changement de base au départ et à l'arrivée.

1.6.19. Conclusion. *Pour $f \in \mathcal{L}(E,F)$ donné par $A = {}_{\mathcal{B}}\operatorname{Mat}_{\mathcal{E}}(f)$, un changement de base avec matrice de passage P au départ donne comme nouvelle matrice ${}_{\mathcal{B}}\operatorname{Mat}_{\mathcal{E}'}(f) = A \cdot P$, pendant qu'un changement de base avec matrice de passage Q à l'arrivée donne comme nouvelle matrice ${}_{\mathcal{B}'}\operatorname{Mat}_{\mathcal{E}}(f) = Q^{-1} \cdot A$. Pour $f \in \mathcal{L}(E,E)$ avec $A = {}_{\mathcal{E}}\operatorname{Mat}_{\mathcal{E}}(f)$, un changement de base avec matrice de passage P donne comme nouvelle matrice ${}_{\mathcal{E}'}\operatorname{Mat}_{\mathcal{E}'}(f) = P^{-1} \cdot A \cdot P$.*

• *Le déterminant d'une matrice carrée.*

Comme ce qui précède montre, les déterminants ne sont pas indispensables pour le calcul en algèbre linéaire. Néanmoins il est important de connaître leur existence, leur propriétés fondamentales, et de savoir les calculer au moins dans des cas relativement simples (sachant que la

formule du déterminant se complique très rapidement avec la taille de la matrice, et déjà pour des matrices 4×4 elle est très longue, car elle a 24 termes).

Voici une liste de propriétés fondamentales du déterminant.

(1) Le déterminant d'une matrice n'est défini que si elle est carrée. Sa valeur est un scalaire.
(2) Le déterminant de $A \in \mathrm{Mat}_n(K)$ est donné par une *formule* en les coefficients de A, formée par addition ou soustraction d'un certain nombre ($n!$, en fait) de produits de n coefficients.
(3) Les matrices carrées inversibles sont précisément celles dont le déterminant n'est pas nul.
(4) Le déterminant est multiplicatif : $\det(A \cdot B) = \det(A)\det(B)$.
(5) Pour tout $n \in \mathbf{N}$ on a $\det(\mathbf{I}_n) = 1$.
(6) Pour les cas $n = 1, n = 2$, les formules sont $\det(a) = a$ respectivement $\det(\begin{smallmatrix} a & c \\ b & d \end{smallmatrix}) = ad - bc$.
(7) Le déterminant est linéaire en chacune des colonnes. Concrètement, si dans $A \in \mathrm{Mat}_n(K)$ on choisit une colonne, et on remplace les coefficients de cette colonne par n variables, la fonction $K^n \to K$ définie par le déterminant de cette matrice est K-*linéaire*. En particulier multiplication d'une seule colonne par un scalaire $\lambda \in K$ multiplie le déterminant par λ.
(8) Si une matrice carrée A contient (au moins) deux colonnes identiques alors $\det(A) = 0$.
(9) L'acte d'additionner à une colonne d'une matrice carrée A un multiple scalaire d'une *autre* colonne de A ne change pas son déterminant (conséquence des deux points précédents).
(10) L'acte d'intervertir deux colonnes d'une matrice carrée multiplie son déterminant par -1.
(11) Le déterminant de la matrice transposée de A est égal à $\det(A)$. Par conséquent la propriétés précédentes restent valables si l'on remplace « colonne(s) » partout par « ligne(s) ».
(12) Le déterminant d'une matrice triangulaire est égal au produit de ces coefficients diagonaux. (Plus généralement le déterminant d'un matrice qui est « diagonal en blocs » est le produit des déterminants de ses blocs diagonaux.)
(13) Pour une matrice $A \in \mathrm{Mat}_n(K)$ pour laquelle soit la colonne j est égale au vecteur $\mathbf{e}_i$ de la base canonique de K^n, soit la ligne i est égale au vecteur $\mathbf{e}_j$ (pour certains indices i, j), on a $\det(A) = (-1)^{i-j}\det(A')$ où A' est obtenu de A par suppression de la ligne i et de la colonne j. Par conséquent, on peut dans le cas général calculer $\det(A)$ par développement pour une ligne ou colonne choisie : cela revient à écrire cette ligne ou colonne comme combinaison linéaire des vecteurs de la base canonique de K^n, et d'appliquer la linéarité (7).

1.7. *Rang.*

Pour les (sous) espaces vectoriels V de dimension finie, la dimension est une quantité caractéristique de V, et c'est essentiellement la seule (car deux espaces de la même dimension sont isomorphes). Pour les applications linéaires il existe une quantité similaire, son *rang*.

1.7.1. Définition. *Si E, F sont des espaces vectoriels de dimension finies, et $f \in \mathcal{L}(E, F)$, le rang de f est défini comme* $\mathrm{rg}(f) = \dim(\mathrm{Im}(f))$.

1.7.2. Fait. *Pour $f \in \mathcal{L}(E, F)$, $\mathrm{rg}(f)$ et un entier entre 0 et* $\min(\dim(E), \dim(F))$.

Que $\mathrm{rg}(f)$ ne peut pas dépasser $\dim(F)$ découle de $\mathrm{Im}(f) \subseteq F$; il ne peut pas dépasser $\dim(E)$ non plus car $\mathrm{Im}(f) = \mathrm{Vect}(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_m))$ quand $[\mathbf{e}_1, \ldots, \mathbf{e}_k]$ est une base de E.

Puisque sa définition n'en fait pas mention, le rang de f ne dépend d'aucun choix de base. Mais pour calculer concrètement $\mathrm{rg}(f)$, on se servira en général d'une matrice $A = {}_{\mathcal{F}}\mathrm{Mat}_{\mathcal{E}}(f)$ par rapport à certaines bases $\mathcal{E}$ de E et $\mathcal{F}$ de F. Si $\mathcal{E} = [\mathbf{e}_1, \ldots, \mathbf{e}_m]$, alors on a par définition $\mathrm{rg}(f) = \dim\big(\mathrm{Vect}(f(\mathbf{e}_1), \ldots, f(\mathbf{e}_m))\big)$, et les colonnes de A contiennent les coordonnées de ces $f(\mathbf{e}_j)$ par rapport à $\mathcal{F}$. Comme l'expression en coordonnées est un isomorphisme (exemple 1.5.17), on conclut que $\mathrm{rg}(f)$ est la dimension du sous-espace de K^n engendré par les colonnes de A.

1.7.3. Définition/Proposition. *Pour $A \in \mathrm{Mat}_{n,m}(K)$ on définit son rang comme $\mathrm{rg}(A) = \mathrm{rg}(f_A)$, où $f_A \in \mathcal{L}(K^m, K^n)$ est l'application linéaire associée à A par l'isomorphisme canonique. Ce rang est égal à $\dim\big(\mathrm{Vect}(\mathbf{v}_1, \ldots, \mathbf{v}_m)\big)$ où $\mathbf{v}_j$ est la colonne j de A. Si $Q \in \mathrm{Mat}_n(K)$ et $P \in \mathrm{Mat}_m(K)$ sont inversibles, on a $\mathrm{rg}(Q^{-1} \cdot A \cdot P) = \mathrm{rg}(A)$.*

La dernière partie est la conséquence du fait qu'un changement du base ne change pas une application linéaire, et donc pas non plus son rang, bien que sa matrice change par multiplication à droite ou/et à gauche par une matrice inversible.

1.7.4. Définition/Proposition. *Le rang d'un système d'équations est défini comme le rang de la matrice A dans sa forme vectorielle $A \cdot x = b$. C'est aussi le rang de la matrice A' dans la forme échelonnée $A' \cdot x = b'$ du système, et donc aussi (d'après la méthode 1.6.18) le nombre $\#P$ de colonnes avec pivots dans A', ou encore le nombre de lignes non nulles de A'.*

Le fait que $\mathrm{rg}(A) = \mathrm{rg}(A')$ découle de $A' = Q \cdot A$ pour une certain matrice inversible Q ; la remarque finale découle du fait que chaque ligne non nulle de A' contient un unique pivot.

Comparaison des méthodes 1.6.17 et 1.6.18 montre le résultat important suivant.

1.7.5. Théorème du rang. *Pour $f \in \mathcal{L}(E, F)$, on a $\dim \mathrm{Ker}(f) + \mathrm{rg}(f) = \dim(E)$.*

De façon informelle ce résultat dit : pour une application linéaire définie sur un espace donné E, plus son noyau est grand, plus son image est petite (en dimension) par rapport à E.

Le résultat suivant montre une symétrie supplémentaire du rang.

1.7.6. Proposition. *Pour toute matrice A on a $\mathrm{rg}(A) = \mathrm{rg}({}^tA)$.*

On ne se sert relativement peu de cette proposition, mais elle est équivalente au théorème du rang. En fait dans la réduction à la forme échelonnée, les pivots marquent les *lignes* qui se sont avérées linéairement indépendantes des lignes précédentes, donc le nombre de pivots s'interprète comme $\mathrm{rg}({}^tA)$. Dans le théorème du noyau ce nombre s'identifie avec le nombre $\dim(E)$ de colonnes moins le nombre $\dim \mathrm{Ker}(f)$ de colonnes sans pivot (car chaque telle colonne donne un élément de base de $\mathrm{Ker}(f)$).

Puisque l'égalité $\mathrm{rg}(Q^{-1} \cdot A \cdot P) = \mathrm{rg}(A)$ de la proposition 1.7.3 permet d'appliquer des opérations sur les lignes comme sur les colonnes, on peut réduire la question au cas où A est remplacée par sa matrice échelonnée A'. Dans ce cas $\mathrm{rg}(A')$ est le nombre de (colonnes avec) pivots, et $\mathrm{rg}({}^tA')$ est le nombre de lignes indépendantes ; puisque les lignes non nulles de A' forment une famille libre (les colonnes avec pivot le montrent), c'est aussi le nombre de pivots.

1.8. *Somme de sous-espaces ; décomposition de l'espace.*

1.8.1. Définition. *Pour des s.e.v. $V_1, \ldots, V_k$ de E, leur somme est un s.e.v. définie comme $V_1 + \cdots + V_k = \{\, v_1 + \cdots + v_k \mid v_1 \in V_1, \ldots, v_k \in V_k \,\}$.*

1.8.2. Définition. *Une somme $V_1 + \cdots + V_k$ de s.e.v. est directe si l'équation $v_1 + \cdots + v_k = 0_E$ pour $v_1 \in V_1, \ldots, v_k \in V_k$ n'a qu'une seule solution, à savoir la triviale : $v_1 = \cdots = v_k = 0_E$. Pour marquer que la somme $V_1 + \cdots + V_k$ est directe, elle peut être écrite $V_1 \oplus \cdots \oplus V_k$.*

De façon équivalente, la somme $V_1 + \cdots + V_k$ est directe si tout vecteur $v \in V_1 + \cdots + V_k$ s'écrit de façon unique sous la forme $v = v_1 + \cdots + v_k$ avec $v_i \in V_i$ pour tout i. (L'existence d'une telle écriture découle de la définition même de $V_1 + \cdots + V_k$ donc c'est seulement l'unicité

qui est en jeu ici ; en plus l'exigence de cette unicité pour $v = 0_E$ suffit, car si elle fait défaut pour $v \neq 0_E$, la soustraction des deux écritures donne une écriture non triviale pour 0_E.) La notion d'une somme directe de s.e.v. est donc semblable à celle d'une famille libre de vecteurs : elle exige l'unicité d'une écriture dans le cas où elle existe. La proposition suivante fait un lien.

1.8.3. Proposition. *Pour deux familles* $[\mathbf{v}_1, \dots, \mathbf{v}_k]$ *et* $[\mathbf{w}_1, \dots, \mathbf{w}_l]$ *dans* E *on a toujours* $\mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k) + \mathrm{Vect}(\mathbf{w}_1, \dots, \mathbf{w}_l) = \mathrm{Vect}(\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}_1, \dots, \mathbf{w}_l)$. *Si en plus les deux familles sont libres, la somme est directe si et seulement si la famille* $[\mathbf{v}_1, \dots, \mathbf{v}_l, \mathbf{w}_1, \dots, \mathbf{w}_l]$ *est libre. Plus généralement, on a les mêmes énoncés pour des sommes de plusieurs sous-espaces.*

Souvent une écriture d'un espace vectoriel comme somme de certains sous-espaces n'est vraiment utile que si on peut montrer que cette somme est directe. En voici deux illustrations.

1.8.4. Définition. *Une écriture* $E = V_1 \oplus \cdots \oplus V_k$ *s'appelle décomposition de l'espace* E *en somme directe de sous-espaces. On peut, grâce à l'écriture unique de tout* $v \in E$ *comme* $v = v_1 + \cdots + v_k$ *avec* $v_1 \in V_1, \dots, v_k \in V_k$, *définir pour* $1 \leq i \leq k$ *une application* $\pi_i : E \to V_i$, *dite projection sur* V_i *selon cette décomposition, par* $\pi_i : (v_1 + \cdots + v_k) \mapsto v_i$ *; elle est linéaire.*

1.8.5. Proposition. *Une somme* $V_1 + \cdots + V_k$ *de sous-espaces vectoriels de dimension finie est directe si et seulement si* $\dim(V_1 + \cdots + V_k) = \dim(V_1) + \cdots + \dim(V_k)$.

La condition d'une solution unique dans le définition d'une somme directe peut être lourde à vérifier. Dans le cas de deux sous-espace, le critère suivant est souvent une alternative utile.

1.8.6. Proposition. *Un somme* $V + W$ *est directe si et seulement si* $V \cap W = \{0_E\}$.

Attention : la généralisation «évidente» de cette proposition à plusieurs s.e.v. est *fausse*. Néanmoins cette caractérisation peut servir pour montrer par récurrence que certaines sommes de plusieurs sous-espaces sont directes, grâce au résultat suivant.

1.8.7. Proposition. *Si une somme* $(V_1 \oplus \cdots \oplus V_k) + (W_1 \oplus \cdots \oplus W_l)$ *de deux somme directes est directe, alors la somme de l'ensemble de s.e.v. est directe, et on a le droit de l'écrire comme* $V_1 \oplus \cdots \oplus V_k \oplus W_1 \oplus \cdots \oplus W_l$. *En particulier* $(V_1 \oplus \cdots \oplus V_k) \oplus V_{k+1} = V_1 \oplus \cdots \oplus V_{k+1}$.

1.8.8. Définition. *Un supplémentaire d'un s.e.v.* V *de* E *est un s.e.v.* W *tel que* $E = V \oplus W$.

1.8.9. Proposition. *Tout sous-espace de* E *admet au moins un s.e.v. supplémentaire.*

C'est une conséquence du théorème de la base incomplète (1.4.4), et la proposition 1.8.3. Le supplémentaire n'est en général pas unique (ce qui n'est pas trop surprenant, en vue du fait qu'on a utilisé le théorème de la base incomplète pour prouver son existence).

Comme illustration de la notion de supplémentaire, on mentionne le résultat suivant, dont la démonstration est laissée comme exercice.

1.8.10. Proposition. *Si* $f \in \mathcal{L}(E, F)$ *et* $V \subseteq E$ *est un sous-espace, alors la restriction* $f|_V : V \to \mathrm{Im}(f)$ *est un isomorphisme si et seulement si* V *est un supplémentaire de* $\mathrm{Ker}(f)$.

Grâce aux propositions 1.8.5 et 1.8.9, cette dernière proposition fournit une preuve alternative du théorème du rang, en choisissant un supplémentaire quelconque du s.e.v. $\mathrm{Ker}(f)$ de E.

Chapitre 2. Endomorphismes.

Dans ce cours on s'intéressera particulièrement aux applications linéaires d'un espace vectoriel E vers lui-même, qui sont appelées des *endomorphismes* de E.

2.0.1. Définition. *Un endomorphisme d'un K-espace vectoriel E est une application $E \to E$ qui est K-linéaire. On note* $\mathrm{End}(E) = \mathcal{L}(E, E)$ *l'espace des endomorphismes de E.*

2.1. *Notions nouvelles pour les endomorphismes.*

La particularité d'endomorphismes, par rapport aux applications linéaires plus générales, est que dans le cas d'un endomorphisme on peut comparer des vecteurs au départ (avant avoir utilisé l'application) avec ceux obtenus à l'arrivée comme images par l'application. Cela permet d'introduire un certain nombre de notions spécifiquement pour les endomorphismes.

• *Similitude de matrices carrées.*

Quand on veut exprimer un endomorphisme par une matrice, les bases dont on à besoin au départ et à l'arrivée sont des base d'un même espace. Il est alors habituel d'une seule base pour les deux objectifs. En fait on fera cela systématiquement (à l'exception du cas déjà traité d'une matrice de passage, dont le but même est la comparaison de deux bases différentes, et pour lequel l'endomorphisme est simplement l'identité de E). Pour un endomorphisme ϕ de E on écrit alors $\mathrm{Mat}_{\mathcal{E}}(\phi)$ au lieu de ${}_{\mathcal{E}}\,\mathrm{Mat}_{\mathcal{E}}(\phi)$.

Une conséquence de l'utilisation d'une même base au départ et à l'arrivée est que pour un endomorphisme, un changement de base veut dire automatiquement des changements simultané au départ et à l'arrivée, comme on l'a évoqué dans la conclusion 1.6.19.

2.1.1. Définition. *Deux matrices carrées $A, B \in \mathrm{Mat}_n(K)$ sont dites semblable s'il existe (au moins) une matrice inversible $P \in \mathrm{Mat}_n(K)$ telle que $B = P^{-1} \cdot A \cdot P$.*

La condition d'être semblable est une relation d'équivalence sur $\mathrm{Mat}_n(K)$. Les classes pour cette relation sont appelés classe de *similitude*.

2.1.2. Proposition. *Si $A = \mathrm{Mat}_{\mathcal{E}}(f)$ et $B = \mathrm{Mat}_{\mathcal{E}'}(f)$, alors A et B sont semblables.*

Dans ce cours on s'intéressera à des situations où, pour mieux comprendre un endomorphisme donné ϕ, on cherche une base spéciale $\mathcal{B}$ telle que $\mathrm{Mat}_{\mathcal{B}}(\phi)$ est d'une forme particulièrement simple (souvent on pourra atteindre une forme dite diagonale, où tous les coefficients hors la diagonale principale sont nuls). Si ϕ est donné par $A = \mathrm{Mat}_{\mathcal{E}}(\phi)$ pour une certaine base $\mathcal{E}$, cela revient donc à chercher un matrice de cette forme dans la classe de similitude de A.

• *Composition, et polynômes d'un endomorphisme.*

Puisque l'espace d'arrivée est celui de départ, les endomorphismes peuvent être composés librement entre eux. En général l'ordre des endomorphismes dans une composition est importante ; l'opération de composition *n'est pas commutative*. En composant ensemble k copies de $\phi \in \mathrm{End}(E)$ on obtient sa *puissance* ϕ^k. Par convention $\phi^0 = \mathbf{I}_E$, quel que soit $\phi \in \mathrm{End}(E)$. En formant des combinaisons linéaires de puissances de ϕ, on construit des *polynômes en ϕ*.

2.1.3. Définition. *Si $P = c_0 + c_1X^1 + \cdots + c_dX^d \in K[X]$, et $\phi \in \mathrm{End}(E)$, on définit le polynôme P en ϕ comme l'endomorphisme $c_0\mathbf{I}_E + c_1\phi^1 + \cdots + c_d\phi^d$, noté $P[X := \phi]$ ou $P[\phi]$.*

L'opération de substitution de ϕ pour X est compatible avec les opérations arithmétiques sur les polynômes. Pour l'addition c'est facile à voir : $(P+Q)[\phi] = P[\phi]+Q[\phi]$. La multiplication

de polynômes correspond à la composition d'endomorphismes : $(PQ)[\phi] = P[\phi] \circ Q[\phi]$. Pour voir pourquoi, les lois distributives pour la multiplication de polynômes et pour la composition de polynômes permettent de réduire cet énoncé au cas où P, Q n'ont qu'un seul terme chacun : $P = cX^k$ et $Q = dX^l$; alors on a $P[\phi] \circ Q[\phi] = c\phi^k \circ d\phi^l = cd\phi^{k+l} = (cdX^{k+l})[\phi] = (PQ)[\phi]$.

La relation pour la multiplication est remarquable, car la multiplication de polynômes est commutative, mais la composition d'endomorphismes ne l'est pas en général. La conséquence suivante est donc utile à souligner ; elle sera souvent utilisée dans les raisonnements, soit explicitement, soit implicitement (en choisissant judicieusement l'ordre de composition quand on écrit une expression de la forme $(PQ)[\phi]$ comme composée de $P[\phi]$ et $Q[\phi]$).

2.1.4. Corollaire. *Deux polynômes $P[\phi], Q[\phi]$ en ϕ commutent : $P[\phi] \circ Q[\phi] = Q[\phi] \circ P[\phi]$.*

• *Sous-espaces stables.*

Pour un endomorphisme donné $\phi \in \mathrm{End}(E)$, certains sous-espaces de E sont spéciaux dans la mesure ou ils sont fermés pour l'opération d'appliquer ϕ.

2.1.5. Définition. *Pour $\phi \in \mathrm{End}(E)$, un sous-espace F de E est ϕ-stable si $\phi(F) \subseteq F$.*

L'intérêt de ses sous-espaces est qu'ils apportent une vue « localisée » et donc simplifiée de l'action de l'endomorphisme.

2.1.6. Définition. *Si $\phi \in \mathrm{End}(E)$, et F est un sous-espace ϕ-stable de E, on peut former la restriction $\phi|_F \in \mathrm{End}(F)$, qui envoie $v \in F \mapsto \phi(v) \in F$ (restriction au départ et à l'arrivée).*

Un seul sous-espace ϕ-stable n'apporte pas une vue sur *tout* l'endomorphisme, mais cela sera possible si on peut trouver une *décomposition de E en somme directe de sous-espaces ϕ-stables.*

2.2. *Vecteurs propres, valeurs propres, diagonalisation.*

Les sous-espaces ϕ-stables V les plus simples, après le trivial $\{0_E\}$, sont ceux avec $\dim(V) = 1$. Pour un tel V on a $\phi|_V = \lambda \mathbf{I}_V$ pour un scalaire λ, donc si $V = \mathrm{Vect}(v)$ avec $v \neq 0$ on a $\phi(v) = \lambda v$.

2.2.1. Définition/Proposition. *Un vecteur propre de ϕ est un vecteur $v \in E$ tel que $\mathrm{Vect}(v)$ soit un sous-espace ϕ-stable de dimension 1. Un vecteur $v \in V$ est vecteur propre de ϕ si $v \neq 0_E$ et s'il existe $\lambda \in K$ tel que $\phi(v) = \lambda v$. Dans ce cas le scalaire λ est unique ; on appelle λ la valeur propre associée au vecteur propre v, et v un vecteur propre (de ϕ) pour λ.*

Les vecteurs propres avec une même valeur propre λ sont souvent considérés ensemble.

2.2.2. Définition/Proposition. *Ayant fixé $\phi \in \mathrm{End}(E)$, le sous-espace propre pour $\lambda \in K$ est $\mathrm{Ker}(\phi - \lambda\mathbf{I})$, noté (dans ce cours) E_λ. L'ensemble des vecteurs propres de ϕ pour λ est celui des vecteurs non nuls de ce sous-espace propre, et λ est une valeur propre de ϕ si et seulement si $\dim\big(\mathrm{Ker}(\phi - \lambda\mathbf{I})\big) > 0$, c'est-à-dire si cet ensemble de vecteurs propres n'est pas vide.*

Le multiple $\lambda\mathbf{I}$ de l'identité est appelé l'homothétie de facteur λ, et E_λ est formé des vecteurs de E pour lesquels l'action de ϕ coïncide avec cette homothétie. Chaque espace propre de ϕ est clairement ϕ-stable, et la restriction de ϕ à E_λ est $\lambda\mathbf{I}_{E_\lambda}$. Plus généralement tout polynôme en ϕ agit par une homothétie sur chaque sous-espace propre de ϕ, comme le dit l'énoncé suivant.

2.2.3. Proposition. *Si E_λ est le sous-espace propre de $\phi \in \mathrm{End}(E)$ pour $\lambda \in K$, et $P \in K[X]$, alors $P[\phi]|_{E_\lambda} = P[\lambda]\mathbf{I}_{E_\lambda}$: sur E_λ l'endomorphisme $P[\phi]$ agit comme la multiplication par $P[\lambda]$.*

La raison est que chaque puissance $(\lambda\mathbf{I}_{E_\lambda})^k$ est égale à l'homothétie $\lambda^k\mathbf{I}_{E_\lambda}$, et la restriction $P[\phi]|_{E_\lambda}$ est une combinaison linéaire de telles puissances avec comme coefficients ceux de P.

Un résultat important est que la somme de différents espaces propres est toujours directe :

2.2.4. Théorème. *Si $\lambda_1, \ldots, \lambda_k \in K$ sont des valeurs distinctes, la somme des espaces propres correspondants $\mathrm{Ker}(\phi - \lambda_1\mathbf{I}_E) + \cdots + \mathrm{Ker}(\phi - \lambda_k\mathbf{I}_E)$ est une somme directe.*

Sur une telle somme de sous-espaces propres E_λ, l'action de ϕ est facile à comprendre : chaque vecteur est somme de ses projections (selon la somme directe) sur les E_λ, et la projection sur E_λ de $\phi(v)$ et obtenu en multipliant celle de v par λ. Ceci exclut clairement la possibilité que v soit un vecteur propre de ϕ pour une valeur propre μ qui ne se trouve pas parmi les valeurs λ dans la somme (car dans ce cas la projection de v sur E_λ serait aussi multipliée par $\mu \neq \lambda$, ce qui est impossible si cette projection n'est pas nulle). Ce constat donne en fait une méthode pour démontrer le théorème par récurrence sur le nombre de valeurs propres, car il dit qu'en considérant les valeurs propres successivement, chaque nouvel espace propre E_μ coupe la somme des espaces propres précédents en $\{0_E\}$ seulement ; la somme reste directe après rajout de E_μ.

2.2.5. Définition/Proposition. *On appelle $\phi \in \mathrm{End}(E)$ diagonalisable si E s'écrit comme une somme (toujours directe) d'espaces propres pour ϕ, donc $E = \bigoplus_{\lambda\in\Lambda} E_\lambda$ avec $\Lambda \subseteq K$. C'est le cas si et seulement si E possède une base constituée entièrement de vecteurs propres de ϕ.*

Pour trouver une base de E de vecteurs propres, on choisit une base dans chaque espace propre E_λ : la réunion de ces bases sera une base de la somme directe $\bigoplus_{\lambda\in\Lambda} E_\lambda = E$. Réciproquement, si on a une base de E constituée entièrement de vecteurs propres de ϕ, on trouve pour chaque λ, en regroupant les vecteurs propres pour λ dans la base, une base d'un sous-espace $\tilde{E}_\lambda$ de E_λ. Comme par construction $E = \bigoplus_{\lambda\in\Lambda} \tilde{E}_\lambda$, et d'autre part la somme des espaces propres E_λ reste directe (théorème 2.2.4), on a nécessairement $\tilde{E}_\lambda = E_\lambda$ pour chaque λ.

2.2.6. Exemple. *L'exemple le plus simple est celui d'une homothétie $\phi = \lambda\mathbf{I}_E$ pour un certain $\lambda \in K$. Dans ce cas E est égal à (la somme direct de) l'unique sous-espace propre E_λ et donc diagonalisable ; tout vecteur non nul est valeur propre de ϕ (pour la valeur propre λ), et toute base de E est une base de vecteur propres.*

2.2.7. Exemple. *Soit $E = V \oplus W$, et ϕ la projection sur V selon cette somme directe (donc parallèle à W) ; on aura donc $\phi(v + w) = v$ pour tout $v \in V$ et $w \in W$. Alors V est E_1, le sous-espace propre de ϕ pour $\lambda = 1$, et W est E_0, le sous-espace propre pour $\lambda = 0$. Comme $E = E_1 \oplus E_0$, on voit que ϕ est diagonalisable. On peut former une base de E formée de vecteurs propres en combinant une base du sous-espace V et avec une base du sous-espace W.*

On généralise facilement ce dernier exemple en associant d'autres scalaires que 1 et 0 aux sous-espaces V, W de la somme directe, ou à une somme de plusieurs sous-espaces.

2.2.8. Définition. *On appelle une matrice $A \in \mathrm{Mat}_n(K)$ diagonale si $A_{i,j} = 0$ dès que $i \neq j$. On l'appelle diagonalisable si elle est semblable à une matrice diagonale.*

Une matrice A est diagonale si et seulement si pour tout indice j, la colonne j de A est est multiple de $\mathbf{e}_j$, ce qui veut dire que $\mathbf{e}_j$ est vecteur propre pour l'application linéaire $K^n \to K^n$ correspondant à A. Si on avait $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ pour $\phi \in \mathrm{End}(E)$, et $\mathcal{B}$ une base de E, cela donne : A est diagonale si et seulement si tous les vecteurs de $\mathcal{B}$ sont des vecteurs propres de ϕ.

2.2.9. Proposition. *Une matrice $A \in \mathrm{Mat}_n(K)$ diagonale si et seulement si chaque vecteur de la base canonique de K^n est vecteur propre de l'endomorphisme de K^n associée à A. Pour $\phi \in \mathrm{End}(E)$ et une base $\mathcal{B}$ de E, la matrice $\mathrm{Mat}_{\mathcal{B}}(\phi)$ est diagonale si et seulement si $\mathcal{B}$ est entièrement constituée de vecteurs propres de ϕ.*

Le terme "endomorphisme diagonalisable" de la définition 2.2.5 pour un endomorphisme qui admet une base constituée de vecteurs propres est justifié par la caractérisation suivante.

2.2.10. Corollaire. *Si $\phi \in \mathrm{End}(E)$ et $\mathcal{B}$ est une base de E, alors ϕ est diagonalisable si et seulement si sa matrice $\mathrm{Mat}_{\mathcal{B}}(\phi)$ est diagonalisable.*

Par définition $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ est diagonalisable si un changement de base, disons de $\mathcal{B}$ vers $\mathcal{C}$, transforme A en $\mathrm{Mat}_{\mathcal{C}}(\phi)$ qui soit diagonale. Mais cela veut dire que $\mathcal{C}$ est constituée de vecteurs propres de ϕ, ce qui est possible si et seulement si ϕ est diagonalisable (proposition 2.2.5).

2.2.11. Exemple. *La matrice $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$ n'est pas diagonale, mais l'endomorphisme ϕ de $\mathbf{R}^2$ associé à cette matrice possède deux sous-espaces propres $E_1 = \mathrm{Vect}(\binom{1}{1})$ et $E_{-1} = \mathrm{Vect}(\binom{1}{-1})$ dont la somme est $E = \mathbf{R}^2$, donc ϕ est diagonalisable. En effet, pour la base $\mathcal{C} = [\binom{1}{1}, \binom{1}{-1}]$ on a la matrice de passage P et matrice $\mathrm{Mat}_{\mathcal{C}}(\phi) = P^{-1}AP$ après changement de bases comme suit :*

$$P = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}, \quad P^{-1} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{pmatrix}, \quad P^{-1}AP = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

quelle matrice est effectivement diagonale, avec les valeurs propres $1, -1$ sur la diagonale.

Le produit matriciel restreint à l'ensemble matrices diagonales est assez facile à comprendre :

2.2.12. Proposition. *Si $A, B \in \mathrm{Mat}_n(K)$ sont des matrices diagonales, le produit matriciel de A et B multiplie simplement les coefficients diagonaux : on a $(A \cdot B)_{i,i} = A_{i,i}B_{i,i}$ pour tout i. Par conséquent A et B commutent : $A \cdot B = B \cdot A$. Si $P \in K[X]$ est un polynôme, la matrice $P[X := A]$ est diagonale, et son coefficient diagonal à la position (i, i) est égal à $P[X := A_{i,i}]$.*

La forme diagonale est surtout utile pour calculer les puissances d'un endomorphisme. Si D est une matrice diagonale avec coefficients diagonaux $c_1, \ldots, c_n$, alors sa puissance D^k est aussi diagonale, avec coefficients diagonaux $c_1^k, \ldots, c_n^k$. Par conséquent, pour un endomorphisme diagonalisable ϕ, l'action de la puissance ϕ^k sur un vecteur quelconque v est facile à décrire *en termes de ses coordonnées par rapport à une base $\mathcal{B}$ de vecteurs propres* : chaque coordonnée de v pour un vecteur b (propre) de $\mathcal{B}$ est multipliée par λ^k, où λ est la valeur propre de b.

Les deux exemples suivants illustrent le calcul des puissances d'un endomorphisme diagonalisable, mais donné par une matrice non diagonale. Le premier est un peu artificiel dans la mesure où on part d'un base de vecteurs propres, avec vecteurs propres associés (ce qui évite d'avoir à chercher une telle base). Dans le second l'endomorphisme est introduit de façon plus naturelle ; il s'agit d'une application classique de la diagonalisation d'une endomorphisme.

2.2.13. Exemple. *Il existe un endomorphisme ϕ de $\mathbf{Q}^2$ unique pour lequel $\mathbf{b}_1 = \binom{1}{1}$ est vecteur propre avec valeur propre -3, et $\mathbf{b}_2 = \binom{2}{1}$ est vecteur propre avec valeur propre 2. Sa matrice A par rapport à la base canonique $[\mathbf{e}_1, \mathbf{e}_2]$ a pour colonnes $\phi(\mathbf{e}_1)$ et $\phi(\mathbf{e}_2)$ qu'on trouve facilement en écrivant $\mathbf{e}_1$ et $\mathbf{e}_2$ comme combinaisons linéaires de $\mathbf{b}_1, \mathbf{b}_2$; le résultat est*

$$A = \begin{pmatrix} 7 & -10 \\ 5 & -8 \end{pmatrix}.$$

Pour mieux comprendre ses puissances A^k, dont les premières sont ($k = 0, 1, \ldots, 5$)

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 7 & -10 \\ 5 & -8 \end{pmatrix}, \quad \begin{pmatrix} -1 & 10 \\ -5 & 14 \end{pmatrix}, \quad \begin{pmatrix} 43 & -70 \\ 35 & -62 \end{pmatrix}, \quad \begin{pmatrix} -49 & 130 \\ -65 & 146 \end{pmatrix}, \quad \begin{pmatrix} 307 & -550 \\ 275 & -518 \end{pmatrix},$$

on établit une similitude de A avec une matrice diagonale D. On sait que $\mathcal{B} = [\mathbf{b}_1, \mathbf{b}_2]$ est une base de vecteurs propres de (l'endomorphisme de matrice) A. La matrice de passage P de $[\mathbf{e}_1, \mathbf{e}_2]$ vers $\mathcal{B}$, et son inverse P^{-1}, sont respectivement

$$P = \begin{pmatrix} 1 & 2 \\ 1 & 1 \end{pmatrix} \qquad \text{et} \qquad P^{-1} = \begin{pmatrix} -1 & 2 \\ 1 & -1 \end{pmatrix},$$

et effectivement $D = P^{-1} \cdot A \cdot P$ est diagonale avec coefficients diagonaux -3 et 2. On a alors

$$D^n = \begin{pmatrix} (-3)^n & 0 \\ 0 & 2^n \end{pmatrix} \quad \text{et} \quad A^n = P \cdot D^n \cdot P^{-1} = \begin{pmatrix} -(-3)^n + 2 \times 2^n & 2 \times (-3)^n - 2 \times 2^n \\ -(-3)^n + 2^n & 2 \times (-3)^n - 2^n \end{pmatrix}.$$

Les coefficients de A sont forcément un peu compliqués à cause de changement de la base $\mathcal{B}$ (par rapport à laquelle la matrice est simplement D^n) vers la base canonique, mais on reconnaît que chaque coefficient est une combinaison linéaire des puissances $(-3)^k$ et 2^k des valeurs propres. Si l'on s'intéresse de façon approximative aux puissances A^k, on peut remarquer que la puissance $(-3)^k$ domine de plus en plus la puissance 2^k quand k devient grand ; en général c'est la puissance de la valeur propre avec la plus grande valeur absolue (ici $|-3| = 3$) qui va dominer les autres.

2.2.14. Exemple. *La suite de Fibonacci $0, 1, 1, 2, 3, 5, 8, 13, 21, \ldots$ se prolonge en rajoutant comme nouveau terme chaque fois la somme de deux termes précédents : $F_0 = 0$, $F_1 = 1$, et ensuite $F_{i+2} = F_i + F_{i+1}$ pour tout $i \in \mathbf{N}$. Trouver une formule pour F_k en termes de k.*

On appelle $F_{i+2} = F_i + F_{i+1}$ la relation de récurrence pour cette suite, et $F_0 = 0$, $F_1 = 1$ les valeurs initiales. Il est utile de considérer l'ensemble de toutes les suites vérifiant cette relation de récurrence, mais avec un couple de valeurs initiales quelconque (dans K^2). Alors on peut remarquer que si (v_i, v_{i+1}) sont deux termes consécutifs d'une telle suite, le passage à (v_{i+1}, v_{i+2}) est décrit par

$$\begin{pmatrix} v_{i+1} \\ v_{i+2} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} v_i \\ v_{i+1} \end{pmatrix}.$$

En itérant cette relation à partir d'un couple (v_0, v_1) de valeurs initiales, on trouve

$$\begin{pmatrix} v_k \\ v_{k+1} \end{pmatrix} = A^k \cdot \begin{pmatrix} v_0 \\ v_1 \end{pmatrix} \quad \text{où} \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}.$$

Donc une fois trouvé une expression explicite pour A^k en fonction de k, on saura exprimer v_k par l'équation $(v_k) = (1 \quad 0) \cdot A^k \cdot \binom{v_0}{v_1}$. Si $\binom{v_0}{v_1}$ est un vecteur propre de A avec valeur propre λ, alors on aura $A^k \cdot \binom{v_0}{v_1} = \lambda^k \binom{v_0}{v_1}$ et donc $v_k = \lambda^k v_0$, ce qui décrit une *suite géométrique* de raison λ et terme initial v_0. La condition qu'une telle suite (non nulle, donc avec $v_0 \neq 0$) vérifie la relation de récurrence est $\lambda^{i+2} = \lambda^i + \lambda^{i+1}$ pour tout $i \in \mathbf{N}$. Il suffit d'avoir l'instance $i = 0$:

$$\lambda^2 = 1 + \lambda,$$

car elle entraîne toutes les autres instances (qui sont ses multiples par λ^i). Or, si $K = \mathbf{R}$ deux telles valeurs λ existent, à savoir $\lambda_+ = \frac{1+\sqrt{5}}{2} \approx 1{,}618034$ (valeur connue comme le nombre d'or) et $\lambda_- = \frac{1-\sqrt{5}}{2} \approx -0{,}618034$. On peut fixer des vecteurs propres concrets en choisissant $v_0 = 1$, ce qui donne $\mathbf{b}_1 = \binom{1}{\lambda_+}$ et $\mathbf{b}_2 = \binom{1}{\lambda_-}$.

Le reste du calcul de A^k est un peu calculatoire, mais sans surprise. Avec deux vecteurs propres indépendants en dimension 2, l'endomorphisme est diagonalisable, avec $\mathcal{B} = [\mathbf{b}_1, \mathbf{b}_2]$ comme une base de vecteurs propres. La matrice de passage P de la base canonique vers $\mathcal{B}$ est celle avec comme colonnes les coordonnées de $\mathbf{b}_1, \mathbf{b}_2$, et P^{-1} est trouvé après un petit calcul

$$P = \begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} \quad \text{et} \quad P^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} -\lambda_- & 1 \\ \lambda_+ & -1 \end{pmatrix}.$$

Alors changement de base vers la base $\mathcal{B}$ donne, comme ce sont des vecteurs propres de valeurs propres λ_+ respectivement λ_-, la matrice diagonale D avec coefficients diagonaux λ_+, λ_- :

$$P^{-1} \cdot A \cdot P = D = \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix} \quad \text{et donc} \quad A = P \cdot D \cdot P^{-1}.$$

La puissance D^k de la matrice diagonale est diagonale avec coefficients diagonaux λ_+^k et λ_-^k. De là on pourra trouver $A^k = P \cdot D^k \cdot P^{-1}$ en effectuant deux multiplications matricielles. Mais on a vu qu'une expression pour F_k est ultimement donnée par le produit matriciel $(1\ \ 0) \cdot A^k \cdot \binom{0}{1}$, et on peut obtenir cette expression par un raccourci en simplifiant $(1\ \ 0) \cdot P = (1\ \ 1)$ et $P^{-1} \cdot \binom{0}{1} = \binom{1}{-1}$:

$$(F_k) = (\,1 \quad 0\,) \cdot P \cdot D^k \cdot P^{-1} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \frac{1}{\sqrt{5}} (\,1 \quad 1\,) \cdot \begin{pmatrix} \lambda_+^k & 0 \\ 0 & \lambda_-^k \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} = \frac{1}{\sqrt{5}} (\lambda_+^k - \lambda_-^k).$$

Cette expression est très utile pour comprendre approximativement le comportement de F_k quand k devient grand. Comme $|\lambda_+| > |\lambda_-|$, sa valeur est alors dominée par la contribution du terme λ_+^k. En fait comme $|\lambda_-| < 1 < |\lambda_+|$, la contribution du terme λ_-^k tendra rapidement vers 0, et F_k est tout simplement le nombre entier le plus proche de $\lambda_+^k/\sqrt{5}$ pour tout $k \geq 0$. On voit aussi que cet entier est plus petit ou plus grand que $\lambda_+^k/\sqrt{5}$ selon la parité de k. Mais si l'on étend la suite de Fibonacci aux indices $k < 0$ (en utilisant la relation de récurrence à rebours $F_{k-1} = F_{k+1} - F_k$; l'expression trouvée ci-dessus pour F_k reste alors valable), on voit que c'est le terme λ_-^k qui va dominer la valeur de F_k pour $k < 0$.

On remarque que la méthode utilisée pour trouver une expression pour F_k est applicable plus généralement, pour les suites récurrentes linéaires d'ordre d (celle de Fibonacci est d'ordre 2), définies par d valeurs initiales et une relation de récurrence qui exprime v_{i+d} comme combinaison linéaire (avec des coefficients fixes) de $v_i, \ldots, v_{i+d-1}$. Les vecteurs propres correspondront toujours aux suites géométriques qui vérifient la relation de récurrence, dont la raison λ est égal à la valeur propre. Les possibilités pour λ sont données par une équation polynomiale de degré d (qui dépend de la relation de récurrence), qui prendra la place de l'équation $\lambda^2 = \lambda + 1$ ci-dessus.

2.3. *Détermination des valeurs propres de* $\phi \in \mathrm{End}(E)$.

Il est facile de trouver l'espace propre d'une valeur λ donnée, mais parmi les scalaires de K il n'y a que peu de valeurs propres (pas plus que la dimension de l'espace). On a donc intérêt à pouvoir limiter les possibilités des valeurs propres d'un endomorphisme donné à un ensemble fini. Cela se fera en trouvant un polynôme dont les valeurs propres doivent être des racines.

• *Polynômes annulateurs.*

2.3.1. Proposition. *Si $\phi \in \mathrm{End}(E)$ admet $P \in K[X]$ comme polynôme annulateur ($P[\phi] = 0$), alors toutes les valeurs propres λ de ϕ sont parmi les racines de P : on a $P[\lambda] = 0$.*

C'est une conséquence direct du fait que $P[\phi]$ agit sur chaque sous-espace propre E_λ par multiplication par le scalaire $P[\lambda]$ (proposition 2.2.3) ; comme $P[\phi] = 0$, il faut que pour tout $\lambda \in K$ avec $\dim(E_\lambda) > 0$ (c'est-à-dire pour chaque valeur propre) on ait $P[\lambda] = 0$.

2.3.2. Exemple. *Un projecteur de E est défini comme un endomorphisme ϕ qui vérifie $\phi^2 = \phi$; autrement dit c'est un endomorphisme pour lequel $X^2 - X$ est un polynôme annulateur (on dit aussi : un endomorphisme annulé par $X^2 - X$). La proposition dit donc que les valeurs propres possibles d'un projecteur sont restreintes à l'ensemble $\{0, 1\}$ des racines de $X^2 - X$.*

2.3.3. Exemple. *Une involution de E est définie comme un endomorphisme ϕ qui est son propre inverse, autrement dit qui vérifie $\phi^2 = \mathbf{I}_E$. C'est un endomorphisme pour lequel $X^2 - 1$ est un polynôme annulateur. La proposition dit donc que les valeurs propres possibles d'une involution sont restreintes à l'ensemble $\{-1, 1\}$ des racines de $X^2 - 1$.*

Les endomorphismes dans ces exemples sont bien particuliers, même s'il existe beaucoup de projecteurs et d'involutions différentes. Mais si on choisit un endomorphisme ϕ (ou une matrice carrée) quelconque, on peut toujours trouver un polynôme annulateur non nul pour ϕ.

2.3.4. Proposition/Définition. *Soit E un K-espace vectoriel de dimension finie n. Pour tout $\phi \in \mathrm{End}(E)$, ou de façon équivalente pour sa matrice $A = \mathrm{Mat}_{\mathcal{B}}(E)$ dans une base donnée $\mathcal{B}$ de E, il existe un polynôme annulateur unitaire $P \in K[X]$ de degré minimal. Ce polynôme, appelé polynôme minimal de ϕ, est déterminé par la première relation de dépendance linéaire dans la suite d'endomorphismes $\phi^0 = \mathbf{I}_E, \phi^1 = \phi, \phi^2, \phi^3, \ldots$: si $[\phi^0, \ldots, \phi^{d-1}]$ est une famille libre dans $\mathrm{End}(E)$, mais ϕ^d vérifie une relation $\phi^d = c_0\phi^0 + \cdots + c_{d-1}\phi^{d-1}$ (respectivement si $[A^0, \ldots, A^{d-1}]$ est une famille libre dans $\mathrm{Mat}_n(K)$, mais $A^d = c_0A^0 + \cdots + c_{d-1}A^{d-1}$), alors le polynôme minimal de ϕ (et de A) est $X^d - c_{d-1}X^{d-1} - \cdots - c_0X^0$.*

• *Polynôme caractéristique.*

Les valeurs propres λ sont les valeurs pour lesquelles l'équation vectorielle $(\phi - \lambda\mathbf{I}_E)(v) = 0_E$ en $v \in E$ (qui décrit le sous-espace propre pour λ) possède des solutions non nulles. Si $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ pour une certaine base $\mathcal{B}$, cela est équivalent à la condition que la matrice carrée $A - \lambda\mathbf{I}_n$ de $\phi - \lambda\mathbf{I}_E$ *ne soit pas* inversible, et d'après une propriété fondamentale du déterminant, cette condition est équivalente à $\det(A - \lambda\mathbf{I}_n) = 0$. Puisque le déterminant est défini en termes de additions, soustraction et multiplications seulement, on peut définir ce déterminant même si on remplace le scalaire λ par une indéterminée X, et c'est l'idée fondamentale qui mène au polynôme caractéristique. Seulement pour des raisons expliquées ci-dessous on préfère utiliser non pas $A - X\mathbf{I}_n$ mais sa matrice opposée $X\mathbf{I}_n - A$, ce qui ne fait que peu de différence, car leurs déterminants sont au signe près égaux (le signe est $(-1)^n$) et on s'intéresse en général qu'aux valeurs à substituer pour X pour lesquelles le déterminant s'annule, qui seront les mêmes.

2.3.5. Définition. *Si $A \in \mathrm{Mat}_n(K)$, le polynôme caractéristique χ_A de A est un polynôme dans $K[X]$ obtenu comme $\det(X\mathbf{I}_n - A)$, le déterminant de la matrice à coefficients dans $K[X]$ donnés par $X - a_{i,i}$ sur la diagonale principale et $-a_{i,j}$ hors de la diagonale, où $A = (a_{i,j})_{i,j=1}^n$.*

La proposition suivante donne quelques propriétés de la forme du polynôme caractéristique. Elle parle de la trace $\mathrm{tr}(A)$ d'une matrice, qui est par définition simplement la somme de ses coefficients diagonaux $a_{i,i}$.

2.3.6. Proposition. *Le polynôme caractéristique χ_A de $A \in \mathrm{Mat}_n(K)$ est unitaire de degré n (son terme dominant est X^n), son coefficient de X^{n-1} est $\mathrm{tr}(-A) = -\mathrm{tr}(A)$, et son coefficient de X^0 (terme constant) est $\det(-A) = (-1)^n \det(A)$.*

Le terme dominant de χ_A vient du produit des coefficients diagonaux $X - a_{i,i}$ de $X\mathbf{I}_n - A$, d'où il est X^n. Si on avait pris $A - X\mathbf{I}_n$ au lieu de $X\mathbf{I}_n - A$, le produit des coefficients diagonaux

$a_{i,i} - X$ aurait eu pour terme dominant $(-X)^n = (-1)^n X^n$, et le polynôme n'aurait pas été unitaire pour n impair. C'est la raison (et la seule) pour laquelle on a choisi d'utiliser $X\mathbf{I}_n - A$ dans la définition de χ_A. Les deux autres coefficients mentionnés dans la proposition sont également facilement identifiés dans l'expansion du déterminant.

2.3.7. Théorème. *Soit $\phi \in \mathrm{End}(E)$, et $A = \mathrm{Mat}_{\mathcal{B}}(E)$ sa matrice dans une base donnée $\mathcal{B}$ de E. Alors les valeurs propres de A sont précisément les racines du polynôme caractéristique χ_A.*

Ce théorème est la principale motivation pour la définition du polynôme caractéristique, et s'explique par le fait que $\chi_A[\lambda] = 0$ veut dire que $\det(\lambda\mathbf{I}_n - A) = 0$, et donc que l'équation $(\lambda\mathbf{I}_n - A)x = 0$ pour $x \in K^n$ possède des solutions non nulles (les vecteurs propres pour λ).

2.3.8. Proposition/Définition. *Deux matrices semblables A et $A' = C^{-1}AC$ ont le même polynôme caractéristique $\chi_A = \chi_{A'}$. Par conséquent, pour $\phi \in \mathrm{End}(E)$ donné, le polynôme caractéristique de $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ est indépendant de la base $\mathcal{B}$ utilisée, et on peut définir le polynôme caractéristique χ_ϕ de ϕ comme ce polynôme.*

Si l'on connaît une sous-espace ϕ-stable V de dimension d, on peut choisir une base $\mathcal{B}$ de E dont les d premiers vecteurs forment une base de V. Dans ce cas les d premières colonnes de $\mathrm{Mat}_{\mathcal{B}}(\phi)$ auront des coefficients nuls après les d premières lignes. On dit qu'une telle matrice est "triangulaire en blocs" $\left(\begin{smallmatrix} A & B \\ 0 & D \end{smallmatrix}\right)$ selon le groupement des d premiers vecteurs de la base et les $n-d$ derniers. Le déterminant d'une telle matrice est $\det(A)\det(D)$, le produit des déterminants de ses blocs (carrés) sur la diagonale principale. D'après la proposition précédente donne alors :

2.3.9. Proposition. *Si $\phi \in \mathrm{End}(E)$ et $V \subseteq E$ est un sous-espace ϕ-stable, alors le polynôme caractéristique $\chi_{\phi|_V}$ de la restriction de ϕ à V divise le polynôme caractéristique χ_ϕ de ϕ.*

Le résultat suivant est célèbre et facile à retenir, bien que sa démonstration soit plus compliquée (voir à la fin du cours). Il est assez étonnant si on le voit comme une identité formelle en les coefficients de A (écrire les cas 2×2 ou 3×3 pour s'en convaincre), un peu moins étonnant si on connaît la relation des polynômes annulateurs et du polynôme caractéristique aux valeurs propres. En tout cas il n'est pas d'une importance absolue pour le développement de la théorie.

2.3.10. Théorème de Cayley-Hamilton. *Pour toute matrice $A \in \mathrm{Mat}_n(K)$ le polynôme caractéristique χ_A de A est un polynôme annulateur de A, c'est-à-dire $\chi_A[A] = 0 \in \mathrm{Mat}_n(K)$.*

∘ *Calcul pratique du polynôme caractéristique.*

Comme un déterminant, différentes méthodes existent pour calculer un polynôme caractéristique à partir de la matrice $X\mathbf{I}_n - A$. On peut calculer de déterminant de manière directe en utilisant la formule générale (surtout pour les cas $n = 2, 3$, car au delà la formule générale pour le déterminant devient très grosse) ou par une expansion par une ligne ou colonne (bien choisie si possible). On peut aussi essayer de simplifier d'abord la matrice $X\mathbf{I}_n - A$ (pas la matrice A!) par des opérations sur les lignes et les colonnes, en «sortant» certains facteurs selon les propriétés connues des déterminants (voir la fin de la section 1.6). Si on arrive à réduire ainsi la matrice en une forme triangulaire, le polynôme caractéristique est simplement le produit des coefficients sur la diagonale ; ceci est l'idéal, car non seulement on aura trouve χ_A, on l'aura en plus sous forme factorisée, et c'est en général juste ce qu'on veut. Mais contrairement à la situation pour une matrice à coefficients dans K, il est souvent impossible de transformer la matrice en une forme triangulaire, car pour utiliser un coefficient "pivot" c de la matrice pour ramener à 0 d'autres coefficients dans sa ligne ou colonne, il faut que c divise ces coefficients, ce qui est une

circonstance exceptionnelle si c n'est pas un polynôme constant. Pour cette raison l'expansion du déterminant reste la seule méthode qui marche dans tous les cas.

• *Recherche des racines d'un polynôme.*

Après avoir calculé le polynôme caractéristique ou avoir trouvé un (autre) polynôme annulateur d'un endomorphisme ϕ, il faudra encore trouver ses racines pour connaître les valeurs propres de ϕ. Cela se fait facilement si le degré du polynôme est (au plus) 2 en utilisant la formule bien connue, mais pour les polynômes de degré 3 ou plus, il n'y a pas de méthode pratique et générale pour trouver ces racines. On ne saura donc trouver les racines de telles polynômes que si on "reconnaît" au moins une racine "évidente" (qui n'a aucune autre raison d'exister que le fait qu'on a tendance à vous proposer des exercices faisables). Un polynôme a une racine 0 si et seulement si son coefficient constant est nul, et les valeurs du polynôme en 1 ou en -1 sont aussi faciles à calculer ; on reconnaîtra donc le telles racines sans problèmes se elles existent. Si cela n'est pas le cas, mais votre polynôme est à coefficients entiers, comme il est souvent le cas dans les exercices, on peut toujours trouver toutes ses racines rationnelles grâce au fait suivant.

2.3.11. Fait. *Si un polynôme unitaire à coefficients entiers P possède une racine rationnelle λ, alors celle-ci est entière ($\lambda \in \mathbf{Z}$) et elle divise le coefficient constant de P.*

On n'oubliera pas d'essayer les diviseurs négatifs du coefficient constant comme ses diviseurs positifs. (Le résultat mentionné se généralise aux polynômes à coefficients rationnelles, qu'on peut rendre à coefficients entiers en multipliant par un scalaire mais qui ne seront alors plus unitaires ; dans ce cas des racines rationnelles non entières peuvent exister, mais seulement avec un dénominateur qui divise le coefficient dominant du polynôme et un numérateur qui divise son coefficient constant. Mais cette généralisation n'est rarement nécessaire dans la pratique.)

Une fois une racine λ de P trouvée, la tâche de trouver les racines restantes se simplifie, car on peut diviser P par $X - \lambda$, c'est-à-dire écrire $P = (X - \lambda)Q$, et ces racines restantes de P sont juste les racines de Q. (Il est possible, mais rare, que λ soit encore une racine de Q ; dans ce cas λ est une racine multiple de P, ce qui est de toute façon une circonstance importante à savoir.) On trouve Q par un procédé appelé division euclidienne de polynômes, qui est semblable à la division d'entiers. On cherche Q tel que le reste $R = P - (X - \lambda)Q$ soit zéro ; on commence avec $Q = 0$ puis on rajoute des termes à Q, à commencer par le plus haut degré, pour faire baisser le degré du reste R. Il est clair à chaque étape il y a un terme (et un seul) qui fera baisser ce degré, jusqu'à ce que R soit devenu constant, et à ce moment $R = P[\lambda]$; puisque on a supposé que λ est racine de P, on aura $R = 0$ comme voulu.

2.3.12. Fait. *Un polynôme $P \in K[X]$ possède $\lambda \in K$ comme racine si et seulement si $X - \lambda$ divise P. Dans ce cas on peut trouver le quotient $Q = P/(X - \lambda)$ par division euclidienne.*

2.4. *Conditions pour que ϕ soit diagonalisable.*

Pour décider si un endomorphisme ϕ de E est diagonalisable, il est en général nécessaire de trouver toutes les racines du polynôme caractéristique de ϕ, ou celles d'un (autre) polynôme annulateur de ϕ, ce qui donne (une liste finie qui contient) les valeurs propres de ϕ. Une fois c'est fait, on peut pour chaque λ dans la liste déterminer le sous-espace propre E_λ, et ϕ sera, d'après les propositions 2.2.4 et 1.8.5, diagonalisable si et seulement si la somme des dimensions de ces sous-espaces est égal à $n = \dim(E)$ (sachant que la somme des dimensions ne peut pas dépasser n, la seule question est de savoir si elle atteint n). Mais dans certains cas il n'est pas nécessaire de faire ce dernier calcul pour pouvoir décider si ϕ est diagonalisable ou non.

• *En utilisant le polynôme caractéristique.*

D'après la proposition 2.3.8, le polynôme caractéristique d'un endomorphisme diagonalisable est déterminé par ses valeurs propres et les dimensions de leurs sous-espaces propres. En faisant un changement de base vers une base de vecteurs propres, on voit que $\chi_A = \chi_D = \prod_{i=1}^{n}(X - \lambda_i)$ si D est une matrice diagonale semblable à A, et λ_i son i-ème coefficient diagonal (qui est une valeur propre de A). Une même valeur propre λ peut apparaître plusieurs fois comme coefficient λ_i ; en fait elle apparaît un nombre de fois égale à la dimension de l'espace propre E_λ. Cela donne :

2.4.1. Proposition. *Si ϕ est diagonalisable, alors $\chi_\phi = \prod_\lambda (X - \lambda)^{\dim(E_\lambda)}$ ou λ parcourt l'ensemble des valeurs propres de ϕ. Par conséquent ϕ est diagonalisable si et seulement si le polynôme caractéristique χ_ϕ se décompose en facteurs de la forme $X - \lambda$, et si pour chaque λ qui apparaît ainsi, $\dim(E_\lambda)$ est égal au nombre d_λ de facteurs $X - \lambda$ dans la décomposition (la multiplicité de λ comme racine de χ_ϕ). En particulier si χ_ϕ se décompose en facteurs $X - \lambda$ qui sont tous distincts (donc χ_ϕ possède n racines simples), alors ϕ est toujours diagonalisable.*

Dans une telle décomposition le nombre total de facteurs est $\sum_\lambda d_\lambda = \deg(\chi_\phi) = \dim(E)$, d'où avoir $\dim(E_\lambda) = d_\lambda$ pour tout λ garantit que $\dim(\bigoplus_\lambda E_\lambda) = \dim(E)$, c'est-à-dire que ϕ est diagonalisable. Remarquons que $\dim(E_\lambda) \leq d_\lambda$ est assuré par la proposition 2.3.9, que ϕ soit diagonalisable ou non. Pour les racines simples λ de χ_ϕ on a certainement $\dim(E_\lambda) \geq 1 = d_\lambda$ car toute racine de χ_ϕ est une valeur propre, ce qui donne la dernière partie de la proposition.

Si $d_\lambda > 1$ et ϕ n'est pas diagonalisable, l'inégalité $\dim(E_\lambda) \leq d_\lambda$ peut très bien être stricte. Par exemple, une matrice triangulaire mais pas diagonale $T \in \mathrm{Mat}_n(K)$ avec des coefficients diagonaux tous égal à λ aura $\chi_T = (X - \lambda)^n$, avec donc $d_\lambda = n$, mais on aura $\dim(E_\lambda) < n$ (car $\dim(E_\lambda) = n$ veut dire $E_\lambda = K^n$, donc $T = \lambda \mathbf{I}_n$, ce qui est faux). Le plus souvent (quand aucun coefficient de la diagonale au-dessus de la principale n'est nul) on aura même $\dim(E_\lambda) = 1$.

La proposition 2.4.1 indique deux types de raisons pour lesquelles un endomorphisme peut ne pas être diagonalisable. Si E est de dimension finie, et $\phi \in \mathrm{End}(E)$ *n'est pas diagonalisable*, alors cela est due à l'une ou l'autre des deux raisons suivantes.

- Le polynôme caractéristique χ_ϕ ne se décompose pas en facteurs de la forme $X - \lambda$ avec $\lambda \in K$; on dit « χ_ϕ n'a pas toutes ces racines dans K » ou « χ_ϕ n'est pas scindé sur K ».
- Pour au moins une racine λ de χ_ϕ dans K avec multiplicité d_λ, on a $\dim(E_\lambda) < d_\lambda$.

Dans le premier cas de figure il reste, après avoir isolé un maximum de facteurs de la forme $X - \lambda$ de χ_λ, comme quotient un polynôme Q non constant sans aucune racine dans K. Des racines de Q existent toujours dans un corps plus grand que K, d'où ce premier type d'obstruction à être diagonalisable dépend fondamentalement du corps K. Dans certaines applications il peut être intéressant d'« élargir le corps K » à un corps K' tel que χ_ϕ se décompose complètement en facteurs de la forme $X - \lambda$ avec $\lambda \in K'$ (il « se scinde sur K' »).

Par exemple, la récurrence définissant la suite de Fibonacci n'utilise que des coefficients entiers, donc on aurait pu aborder notre analyse avec un $\mathbf{Q}$-espace vectoriel de suites à coefficients rationnels. Mais au cours de cette analyse on avait besoin des racines λ_+, λ_- du polynôme $X^2 - X - 1$; ces racines sont réelles, mais pas rationnels. Donc pour pouvoir formuler une formule explicite pour le nombres de Fibonacci, on avait intérêt à élargir le corps $K = \mathbf{Q}$ au corps $K' = \mathbf{R}$ qui contient les racines de $X^2 - X - 1$ (qui est en effet le polynôme caractéristique de la matrice $A = \left(\begin{smallmatrix} 0 & 1 \\ 1 & 1 \end{smallmatrix}\right)$ utilisée). De façon similaire, un endomorphisme d'un $\mathbf{R}$-espace vectoriel peut avoir un polynôme caractéristique qui manque de racines réelles pour de décomposer entièrement en facteurs de la forme $X - \lambda$ avec $\lambda \in \mathbf{R}$, mais qui se décompose ainsi en admettant certains facteurs $X - \lambda$ avec $\lambda \in \mathbf{C} - \mathbf{R}$ (qui viennent en paires avec le facteur conjugué complexe $X - \overline{\lambda}$).

Contrairement à $\mathbf{Q}$ ou $\mathbf{R}$, le corps $\mathbf{C}$ a la propriété que *tout* polynôme non constant (dans $\mathbf{C}[X]$) possède une racine dans $\mathbf{C}$; c'est le théorème d'Alembert–Gauss (prouvé en 1814 par le mathématicien suisse Jean-Robert Argand). Par conséquent, il est garanti avec $K = \mathbf{C}$ que tout polynôme unitaire se décompose en facteurs de la forme $X - \lambda$ pour divers valeurs de $\lambda \in \mathbf{C}$. Donc pour un endomorphisme d'un $\mathbf{C}$-espace vectoriel, on ne peut pas avoir le premier type d'obstruction à être diagonalisable. Dans ce cas, le seul type d'obstruction possible est donc la second, qui n'entre en considération que si le polynôme caractéristique possède une racine multiple. Comme c'est une condition assez rare, on peut dire que pour les endomorphismes d'espaces vectoriels complexes, le cas non diagonalisable est l'exception plutôt que la règle.

• *En utilisant un polynôme annulateur.*

Si λ est une valeur propre de ϕ, tout polynôme annulateur P de ϕ doit avoir λ comme racine (proposition 2.3.1). Autrement dit, "$P[\lambda] = 0$ pour toute valeur propre λ de ϕ" est nécessaire pour que P soit annulateur de ϕ. Mais si $\dim(E_\lambda) > 1$, cela *n'entraîne pas* que λ est racine multiple de P, comme on l'a vu pour le polynôme caractéristique. En fait, si ϕ est diagonalisable, cette condition seule "$P[\lambda] = 0$ pour toute valeur propre λ de ϕ"est déjà suffisante pour que P soit annulateur de ϕ, car $P[\phi]$ agit par $P[\lambda] = 0$ sur chaque E_λ (d'après la proposition 2.2.3), et leur somme remplit l'espace entier dans le cas diagonalisable. Dans ce cas le polynôme *minimal* aura donc juste les valeurs propres comme racines, et chacune juste une seul fois :

2.4.2. Proposition. *Si ϕ est diagonalisable, et si $\{\lambda_1, \ldots, \lambda_k\}$ est l'ensemble de ses valeurs propres, sans répétitions parmi les λ_i, alors le polynôme minimal de ϕ est $(X - \lambda_1) \ldots (X - \lambda_k)$.*

2.4.3. Corollaire. *Le polynôme minimal d'un endomorphisme diagonalisable se décompose en facteurs de la forme $X - \lambda$ pour de valeurs $\lambda \in K$ toutes distinctes ; ses racines sont simples.*

En reformulant cela par contraposée, si le polynôme minimal de ϕ admet soit un diviseur non constant sans racines (qui forme une obstruction à la décomposition en facteurs de la forme $X - \lambda$), soit une racine multiple, alors ϕ ne peut pas être diagonalisable. La seconde clause est en contraste avec le polynôme caractéristique, pour lequel l'existence d'une racine multiple λ n'exclut pas à elle seule la possibilité que ϕ puisse être diagonalisable.

La forme du polynôme minimal décrit dans le corollaire *caractérise* les endomorphismes diagonalisables : dès qu'une endomorphisme ϕ a un tel polynôme minimal, ϕ est forcément diagonalisable. Un énoncé dans cette direction est même vrai avec un polynôme annulateur quelconque au lieu du polynôme minimal, donc c'est comme cela qu'on le formulera :

2.4.4. Théorème. *Si $P \in K[X]$ se décompose comme produit $P = \prod_{i=1}^{k}(X - \lambda_i)$ pour des scalaires $\lambda_1, \ldots, \lambda_k \in K$ toutes distinctes (on dit que P est scindé et à racines simples), alors tout endomorphisme ϕ pour lequel P est polynôme annulateur est diagonalisable.*

On peut donc par exemple affirmer que tout projecteur (exemple 2.3.2) et que toute involution (exemple 2.3.3) est diagonalisable : $X^2 - X$ et $X^2 - 1$ sont scindés et à racines simples.

On verra plus loin (4.2.2) le «théorème de décomposition des noyaux» qui fournit une explication fondamentale pourquoi ce théorème est valable : les différents facteurs $(X - \lambda_i)$ dans la décomposition de P sont premiers entre eux deux à deux, et dans ce cas le noyau de $P[\phi]$ (c'est-à-dire ici E tout entier) se décompose en somme directe des noyaux des $(X - \lambda_i)[\phi] = \phi - \lambda_i\mathbf{I}$ individuels, c'est-à-dire des sous-espaces propres E_{λ_i}. Mais sa démonstration nécessitera de nouvelles idées pas encore abordées. Entre temps, le théorème actuel 2.4.4 étant moins général que le théorème 4.2.2, on saura déjà le montrer avec un argument plus basique. On donne cet argument pas pour le retenir, mais juste à titre d'exemple d'application des résultats précédents.

Preuve. Si P et ϕ sont comme dans l'énoncé, alors on sait d'un côté que $P[\phi] = 0$, et d'autre côté que $P[\phi]$ est la composée $f_1 \circ \cdots \circ f_k$ d'endomorphismes $f_i = (X - \lambda_i)[\phi] = \phi - \lambda_i \mathbf{I}$ pour $i = 1, 2, \ldots, k$. On veut montrer que cela entraîne que la somme des sous-espaces E_{λ_i} remplit E. Sachant que cette somme est certainement directe (théorème 2.2.4), il suffira de montrer l'inégalité

$$\dim(E) \leq \dim\left(\sum_{i=1}^{k} E_{\lambda_i}\right) = \sum_{i=1}^{k} \dim(E_{\lambda_i})$$

(cette inégalité sera alors une égalité, car on a évidemment $E \supseteq \sum_{i=1}^{k} E_{\lambda_i}$). Comme on a $E = \ker(P[\phi]) = \ker(f_1 \circ \cdots \circ f_k)$, et $E_{\lambda_i} = \ker(f_i)$ par définition, cette inégalité est une instance du fait plus général que la dimension du noyau d'une composée d'applications linéaires ne peut pas dépasser la somme des dimensions des noyaux des applications linéaires individuelles. Pour démontrer ce fait, on peut se focaliser sur le cas d'une composée de *deux* applications linéaires, car le cas d'une composée de $k > 2$ applications linéaires en découlera facilement par récurrence :

$$\begin{aligned} \dim(\ker(f_1 \circ \cdots \circ f_k)) &\leq \dim(\ker(f_1)) + \dim(\ker(f_2 \circ \cdots \circ f_k)) \\ \text{(par hypothèse de récurrence :)} \quad &\leq \dim(\ker(f_1)) + \dim(\ker(f_2)) + \cdots + \dim(\ker(f_k)) \end{aligned}$$

Nous avons donc réduit la démonstration du théorème à celle du lemme suivant.

2.4.5. Lemme. *Si $f \in \mathcal{L}(E, F)$ et $g \in \mathcal{L}(F, G)$ pour certains K-espaces vectoriels E, F, G, alors* $\dim(\ker(g \circ f)) \leq \dim(\ker(f)) + \dim(\ker(g))$.

Preuve. La partie de E en dehors de $\ker(g \circ f)$ n'a pas d'importance pour ce lemme, donc on pose $V = \ker(g \circ f) \subseteq E$. Alors par définition $f(V) \subseteq \ker(g))$ et donc $\dim(f(V)) \leq \dim(\ker(g))$. En appliquant le théorème du rang à la restriction $f|_V$ on obtient, puisque $\ker(f|_V) = \ker(f)$:

$$\begin{aligned} \dim(\ker(g \circ f)) = \dim(V) = \dim(\ker(f|_V)) + \dim(f|_V(V)) &= \dim(\ker(f)) + \dim(f(V)) \\ &\leq \dim(\ker(f)) + \dim(\ker(g)). \end{aligned}$$

Dans la démonstration ci-dessus, un point essentiel utilisé est le fait que la somme des sous-espaces E_{λ_i} est directe, sans lequel on n'aurait pas pu utiliser la valeur de $\sum_{i=1}^{k} \dim(E_{\lambda_i})$. C'est pour obtenir ce point qu'a servi l'hypothèse essentiel que les λ_i sont des racines *simples* de P.

Puisque le polynôme minimal μ_ϕ de ϕ est par définition un polynôme annulateur de ϕ, le théorème 2.4.4 et le corollaire qui le précède donnent la condition suivante qui caractérise quand un endomorphisme d'un espace vectoriel de dimension finie est diagonalisable.

2.4.6. Corollaire. *Un endomorphisme ϕ est diagonalisable si et seulement si son polynôme minimal μ_ϕ est scindé et à racines simples.*

Chapitre 3. Polynômes à coefficients dans K.

On s'est déjà servi de polynômes en une indéterminée X, par exemple comme polynômes annulateur ou caractéristique d'un endomorphisme. Dans ce chapitre on étudie certains propriétés de l'ensemble $K[X]$ de tous ces polynômes.

3.1. *Structure d'anneau de* $K[X]$.

Les polynômes modélisent des expressions en une seule inconnue représentée par X, en utilisant des valeurs connues (constantes), et les opérations d'addition, soustraction, et multiplication. Par conséquent, l'ensemble $K[X]$ des polynômes en X est muni de ces opérations (et contient aussi toutes les constantes de K). Mais dans $K[X]$, le symbole X ne cache plus une valeur inconnue, mais est une valeur en soi, distincte de toute constante $c \in K$.

Une structure algébrique qui est muni des opérations d'addition, soustraction, et multiplication, avec les propriétés habituelles, est appelé un *anneau*. Un corps est un cas particulier d'un anneau, mais avec la propriété supplémentaire que la multiplication par tout élément non nul est inversible (ce qui rend possible la division par un tel élément) ; dans un anneau ce n'est pas nécessairement le cas (et ce n'est pas le cas dans $K[X]$). L'exemple le plus connu d'un anneau qui n'est pas un corps est l'anneau $\mathbf{Z}$ des entiers relatifs. On verra que, bien que les polynômes sont plus compliqués que les entiers, les anneaux $\mathbf{Z}$ et $K[X]$ ont bien des propriétés en commun.

Toute expression formée partir de X et de constantes dans K en utilisant addition, soustraction, et multiplication désigne donc un polynôme en X. Deux expressions désignent le même élément de $K[X]$ si l'on peut transformer l'une en l'autre par les règles telles que les lois associative, distributive et commutative, ainsi qu'en appliquant l'arithmétique dans K pour les constantes. Par ces moyens toute expression polynomiale se transforme en une somme de termes, chacun le produit d'une constante de K et d'une puissance de X (aussi appelé *monôme* ; on y inclut le cas $X^0 = 1$). En regroupant les termes avec le même monôme, on obtient une combinaison K-linéaire de monômes, et c'est sous cette forme qu'on peut définir formellement l'égalité de polynômes : deux expressions polynomiales désignent le même polynôme si et seulement si, après écriture de chacune sous forme de combinaison linéaire de monômes, on trouve pour chaque monôme que ses coefficients dans les deux combinaisons sont égaux (en convenant qu'en l'absence du monôme dans la combinaison linéaire, le coefficient est défini comme 0).

3.1.1. Caractérisation. *$K[X]$ est un K-espace vectoriel, et la famille $[X^0 = 1, X, X^2, X^3, \ldots]$ des monômes en forme une base. Cet espace est donc de dimension infinie. La multiplication $K[X] \times K[X] \to K[X]$ est bilinéaire (c'est-à-dire, quand on fixe l'un des deux arguments à un polynôme quelconque elle donne une application linéaire $K[X] \to K[X]$) et commutative, et elle est donnée sur la base des monômes par $X^i X^j = X^{i+j}$.*

Bien que le symbole X ne cache pas une valeur inconnue de K, le fait de pouvoir remplacer X par une valeur concrète, tout en respectant les opérations de combinaisons K-linéaires et de multiplication, reste un aspect fondamental de polynômes. On remarque que l'absence de divisions est essentiel ici, car le remplacement de X par $a \in K$ peut rendre une expression nulle.

3.1.2. Définition/Proposition. *Pour toute constante $a \in K$ une application $K[X] \to K$ est définie par la substitution de a pour X. Cette application, notée $P \mapsto P[X := a]$ (souvent abrégé $P[a]$) est K-linéaire, et compatible avec multiplication : $(PQ)[X := a] = P[X := a]\, Q[X := a]$.*

Le fait qu'un polynôme n'a qu'un nombre fini de termes (non nuls) permet de définir son degré comme le plus grand exposant de X dans un terme non nul. Ceci est donc bien défini, sauf pour le polynôme nul pour lequel un tel terme n'existe pas.

3.1.3. Définition. *Pour un polynôme non nul* $P = \sum_{i=0}^{d} p_i X^i \in K[X]$ *on définit son degré comme* $\deg(P) = \max\{\, i \mid p_i \neq 0 \,\}$. *Pour le polynôme nul on convient que* $\deg(0) = -\infty$.

Cette dernière convention est liée aux deux choses pour lesquelles on utilise le degré des polynômes : pour les comparer (et dans ce cas $-\infty$ est considéré plus petit que tout autre degré), et les additionner (et dans ce cas la somme de $-\infty$ et n'importe quel degré donne $-\infty$). Grâce à cette convention la proposition suivante est valable même si P, Q, $P+Q$ ou PQ est nul.

3.1.4. Proposition. *Pour tout* $P, Q \in K[X]$ *on a*
(1) $\deg(P+Q) \leq \max(\deg(P), \deg(Q))$, *avec égalité si* $\deg(P) \neq \deg(Q)$, *et*
(2) $\deg(PQ) = \deg(P) + \deg(Q)$.

La terminologie suivante rappelle les origines de X comme « variable » (même si dans $K[X]$ c'est une valeur fixe). C'est pour cette raison qu'on appelle « constant » ce qui ne contient pas X.

3.1.5. Définition. *Un polynôme* $P \in K[X]$ *est constant si* $\deg(P) \leq 0$, *donc* $P = cX^0$ *pour un certain* $c \in K$. *On confond souvent ce polynôme* $P \in K[X]$ *et* $c \in K$ *(donc on considère* K *comme une partie de* $K[X]$, *formée des polynômes constants). En général le terme de la forme* cX^0 *dans un polynôme est appelé son terme constant, et* c *son coefficient constant.*

Puisque $0^i = 0$ si $i > 0$ mais $0^0 = 1$, on a la caractérisation suivante du coefficient constant.

3.1.6. Fait. *Le coefficient constant d'un polynôme* P *est égal à* $P[X := 0]$.

3.1.7. Définition. *Si* $P \in K[X]$ *est non nul, son terme* $p_d X^d$ *avec* $d = \deg(P)$ *est le terme dominant de* P, *et* $p_d \in K$ *est le coefficient dominant de* P *(les deux sont non nuls par définition de* deg*). Un polynôme non nul dont le coefficient dominant est* 1 *est appelé unitaire. L'opération de « rendre* P *unitaire » consiste à le diviser par son coefficient dominant, le résultat étant unitaire.*

3.2. *Divisibilité, polynômes irréductibles, comparaison avec* **Z**.

L'anneau $K[X]$ n'est pas un corps. En fait les seuls polynômes inversibles, ceux par lesquels on peut toujours diviser, sont les polynômes constants et non nuls. Par comparaison, les éléments inversibles dans l'anneau **Z** sont -1 et 1. La division (exacte) par un polynôme non constant est parfois possible, mais souvent pas ; ceci donne naissance à la relation de divisibilité.

3.2.1. Définition. *Un multiple de* $A \in K[X]$ *est un* $B \in K[X]$ *tel que* $B = AQ$ *pour un certain* $Q \in K[X]$. *Le quotient* Q *est alors unique si* $A \neq 0$, *et on écrit* $Q = B/A$. *Dans ce cas on dit également que* A *divise* B, *que* B *est divisible par* A, *ou que* A *est un diviseur de* B.

Pour la relation de divisibilité les polynômes inversibles ne sont pas très intéressants, et leur existence est plutôt encombrante : tout diviseur A de B s'accompagne automatiquement de tous ses multiples scalaires cA comme autres diviseurs. Dans **Z**, pour ignorer des décompositions multiplicatives comme $7 = -1 \times -7$, on se limite pour la divisibilité qu'aux nombres positifs : ainsi les différents diviseurs d'un nombre ne seront pas équivalents. Pareillement, on se restreint pour la divisibilité dans $K[X]$ souvent aux polynômes unitaires.

3.2.2. Définition. *Un polynôme non constant* $P \in K[X]$ *est réductible s'il s'écrit comme produit* $P = AB$ *de polynômes non constants* $A, B \in K[X]$ *; si une telle décomposition n'existe pas, on dit que* P *est irréductible.*

Ces notions sont les analogues des nombres composés respectivement premiers dans **Z**. Il est à noter qu'un polynôme constant n'est considéré comme ni réductible, ni irréductible, tout comme dans **Z** on ne considère les nombres 0 et 1 (et -1) comme ni composés, ni premiers.

Dans $\mathbf{Z}$ on peut décomposer chaque nombre strictement positif comme produit de nombres premiers (en convenant que 1 est le produit d'une famille vide). Pour les polynômes on a une décomposition similaire en facteurs irréductibles. Pour des raisons indiquées ci-dessus on préfère d'imposer à ces facteurs irréductibles d'être unitaires, ce qui nous oblige d'admettre un facteur scalaire dans la décomposition (qui sera égal au coefficient dominant du polynôme).

3.2.3. Proposition. *Tout polynôme non nul s'écrit comme le produit d'un scalaire non nul (son coefficient dominant) et d'un certain nombre de facteurs unitaires irréductibles.*

Cette proposition assez facile à démontrer par récurrence sur le degré (ce qui ne veut pas dire qu'une telle *factorisation* est facile à trouver effectivement). Si un polynôme P est constant on prendra aucun facteur irréductible dans sa factorisation. Sinon, soit P admet une décomposition $P = AB$ avec A, B non constant, et alors on combine les factorisations de A et B qui existent par hypothèse de récurrence, soit P n'admet pas une décomposition et est lui-même irréductible.

Ce qui est aussi vrai, mais moins facile à démontrer (on l'admettra), est que cette factorisation d'un polynôme P est *unique*, à l'ordre des facteurs unitaires irréductibles près. Ce fait est analogue à ce qu'on sait pour la factorisation en nombres premiers dans $\mathbf{Z}_{>0}$.

D'après la proposition 3.1.4, le produit de deux polynômes non constants est toujours de degré 2 au moins, donc tout polynôme de degré 1 est automatiquement irréductible. Les polynômes unitaires de degré 1 ont la forme $X - a$ avec $a \in K$. On l'écrit comme cela à cause du fait 2.3.12, qui dit que pour un polynôme P, avoir un tel facteur correspond au fait que a est une racine de P. Par conséquent, un polynôme irréductible de degré > 1 de $K[X]$ est *sans racines* dans K (mais la réciproque est fausse). Puisque cela est impossible pour $K = \mathbf{C}$, on a :

3.2.4. Fait. *Dans $\mathbf{C}[X]$ les polynômes unitaires irréductibles sont les polynômes de la forme $X - a$ pour $a \in \mathbf{C}$; il n'existe pas dans $\mathbf{C}[X]$ d'autres polynômes unitaires irréductibles.*

Dans $\mathbf{R}[X]$ il existe des polynômes irréductibles de degré 2, ceux de discriminant < 0, car il n'ont pas de racines (dans $\mathbf{R}$), donc pas de diviseurs de degré 1, et sans cela un polynôme de degré 2 (ou 3) ne peut être réductible. On voit bien que la notion de réductibilité dépend de l'anneau de polynômes considéré, car ces polynômes se décomposent bien en deux facteurs dans $\mathbf{C}[X]$. Dans $\mathbf{Q}[X]$ des polynômes irréductibles existent même en tout degré $d > 0$. Dans ce cours on ne fera rien avec des facteurs irréductibles de degré > 1 ; il suffit de savoir que si le polynôme caractéristique ou minimal d'un endomorphisme ϕ a un tel facteur, cela empêche ϕ d'être diagonalisable. Mais pour $K = \mathbf{C}$ de tels facteurs n'existent pas. Et en l'absence de tels facteurs, factoriser un polynôme revient essentiellement à la recherche de ses racines.

3.3. *Division euclidienne, diviseurs communs, relations de Bézout.*

On a vu que si une racine a de $P \in K[X]$ est connue, on peut trouver le quotient $Q = P/(X-a)$ par un procédé de division qui ressemble à l'algorithme de division des nombres entiers. Ce procédé, qui accumule des contributions au quotient en descendant du plus haut degré, avec pour but de rendre le reste nul, peut aussi être appliqué pour trouver le quotient de P après division par un diviseur d'une autre forme que $X - a$. En fait c'est une méthode pratique pour savoir si un polynôme D est diviseur de P : c'est le cas si, après avoir retiré de P un multiple convenable de D et au moment où on ne peut plus continuer car le degré du reste est devenu plus petit que $\deg(D)$, le reste est en fait nul. Si au contraire on trouve une reste $R = P - QD$ non nul avec $\deg(R) < \deg(D)$, cela prouve que D ne divise pas P, car un tel R ne peut pas être multiple de D d'après la proposition 3.1.4 (2). Cela donne un algorithme dont l'existence est

d'une importance fondamentale dans l'étude de $K[X]$, et qui s'appelle la division euclidienne (à l'instar de l'algorithme pour division de nombres entiers avec reste, qui a la même structure).

3.3.1. Proposition [division euclidienne dans $K[X]$]. *Soit $A, B \in K[X]$ avec $B \neq 0$. Alors il existe $Q, R \in K[X]$ avec $\deg(R) < \deg(B)$ tels que $A = QB + R$. Le couple (Q, R) est unique.*

Plus important que la condition pour un polynôme seul d'être irréductible sera pour nous la relation entre deux polynômes de ne pas avoir des diviseurs communs (sauf les inversibles).

3.3.2. Définition. *Deux polynômes $A, B \in K[X]$ sont premiers entre eux si aucun polynôme non constant divise à la fois A et B.*

Une condition équivalente est que A et B n'ont pas de diviseur irréductible en commun. Il est donc facile à reconnaître que A, B sont premiers entre eux si on connaît les factorisations de A et de B. Mais cela n'est pas nécessaire, grâce à la notion et le résultat suivants.

3.3.3. Définition/Théorème. *Une combinaison polynomiale de $A, B \in K[X]$ est un polynôme obtenu comme $SA + TB$ pour certains $S, T \in K[X]$. Si A, B ne sont pas tous deux nuls, il existe parmi les combinaisons polynomiales unitaires de A, B un polynôme unique D de plus petit degré. Ce D est diviseur commun de A et B, et tout diviseur commun de A et B divise aussi D. On appelle D le plus grand diviseur commun de A et de B, et on écrit $D = \operatorname{pgcd}(A, B)$.*

Une combinaison polynomiale non nulle rendu unitaire reste une combinaison polynomiale, d'où l'existence de D est claire. Son unicité est aussi facile à voir : si on avait deux candidats différents, leur différence serait une combinaison polynomiale de plus petit degré, qui rendu unitaire contredirait leur définition. Si D ne divisait pas A, alors le reste $R = A - QD$ de la division euclidienne de A par D serait une combinaison polynomiale non nulle de A, B de plus petit degré que D ; là encore, c'est absurde. Pareillement D divise B. Un diviseur commun C de A et de B divise aussi toutes les combinaisons polynomiales de A, B, donc en particulier D.

3.3.4. Corollaire. *$A, B \in K[X]$ sont premiers entre eux si et seulement si $\operatorname{pgcd}(A, B) = 1$.*

Contrairement à la factorisation de la proposition 3.2.3, la détermination de $\operatorname{pgcd}(A, B)$ est possible par un algorithme, qui consiste à trouver des combinaisons polynomiales de degré de plus en plus petit, jusqu'à l'obtention du pgcd. On ne donne pas ici les détails (pourtant simples) de cet « algorithme d'Euclide dans $K[X]$ », car il est difficile à mettre en œuvre manuellement, sauf si $\deg(A)$ ou $\deg(B)$ est assez petit (disons ≤ 2). Et dans ce dernier cas, on peut aussi appliquer la méthode suivante, encore plus simple. On initialise un polynôme C par la valeur de A ou B (ce C sera toujours une combinaison polynomiale de A, B), et tant que C ne divise pas l'un des polynômes A et B, on remplace C par son reste dans la division euclidienne correspondante, rendu unitaire ; finalement C divisera A et B, et on aura alors $\operatorname{pgcd}(A, B) = C$.

3.3.5. Proposition. *$A, B \in K[X]$ sont premiers entre eux si et seulement s'il existe $S, T \in K[X]$ tels que $SA + TB = 1$. On appelle (S, T) un couple de coefficients de Bézout pour A, B.*

Si A, B sont premiers entre eux, la définition 3.3.3 et le corollaire 3.3.4 donnent l'existence de S, T. Réciproquement $SA + TB = 1$ exclut tout diviseur commun de A, B non constants.

Finalement, on aura besoin du résultat suivant, qu'on admet. Une démonstration possible est basée sur le "lemme d'Euclide" qui dit que si un polynôme irréductible divise un produit, alors il doit diviser l'un au moins de ses facteurs. Il est aussi la clé pour l'unicité des factorisations.

3.3.6. Proposition. *Deux produits $P_1 \cdots P_k$ et $Q_1 \cdots Q_l$ (avec $P_i \in K[X]$, $Q_j \in K[X]$) sont premiers entre eux si et seulement si P_i et Q_j sont premiers entre eux pour toute paire (i, j).*

Chapitre 4. Réduction d'endomorphismes.

Dans ce dernier chapitre on étudie comment on peut décomposer E en somme directe de sous-espaces ϕ stables, sans faire l'hypothèse que l'endomorphisme ϕ soit diagonalisable (car dans ce cas la décomposition en sous-espaces propres convient). Il sera souvent utile de considérer la restriction $\phi|_V$ de ϕ à un certain sous-espace V (comme endomorphisme de V). C'est possible seulement si le sous-espace V est ϕ-stable, ce qu'il faudra donc vérifier à chaque occasion. Mais on ne mentionnera pas chaque fois cette vérification, car elle découlera toujours du fait suivant.

4.0.1. Lemme. *Pour $\phi \in \mathrm{End}(E)$ et $P \in K[X]$ quelconques, les sous-espaces $\mathrm{Ker}(P[\phi])$ et $\mathrm{Im}(P[\phi])$ sont ϕ-stables.*

La preuve est un exercice facile ; dans les deux cas le point essentiel est que ϕ et $P[\phi]$ commutent. (En fait on peut remplacer $P[\phi]$ par tout endomorphisme qui commute avec ϕ.)

Une autre observation qui va être utilisée plusieurs fois est la suivante :

4.0.2. Lemme. *Si $P, Q \in K[X]$, alors PQ est polynôme annulateur de $\mathrm{End}(E)$ si et seulement si $\mathrm{Ker}(P[\phi]) \supseteq \mathrm{Im}(Q[\phi])$.*

La condition que PQ est polynôme annulateur veut dire que $(PQ)[\phi] = P[\phi] \circ Q[\phi] = 0$, ou encore $\forall v \in E : P[\phi](Q[\phi(v)) = 0$, est puisque $\mathrm{Im}(Q[\phi]) = \{\, Q[\phi](v) \mid v \in E \,\}$ c'est équivalent à $\forall w \in \mathrm{Im}(Q[\phi]) : w \in \mathrm{Ker}(P[\phi])$, ou simplement à $\mathrm{Im}(Q[\phi]) \subseteq \mathrm{Ker}(P[\phi])$. (Là encore, plus généralement une composée $g \circ f$ d'applications linéaires est nulle précisément si $\mathrm{Ker}(g) \supseteq \mathrm{Im}(f)$.)

4.1. *Compléments d'information concernant le polynôme minimal.*

Concernant le polynôme minimal μ_ϕ d'un endomorphisme ϕ on a vu jusqu'ici sa définition (2.3.4), le fait qu'il compte chaque valeur propre de ϕ parmi ses racines car c'est un polynôme annulateur (proposition 2.3.1), que ϕ est diagonalisable si et seulement si μ_ϕ est scindé et à racines simples (corollaire 2.4.6), quelles racines sont alors précisément les valeurs propres de ϕ (proposition 2.4.2). Dans cette section on apportera quelques faits complémentaires.

4.1.1. Proposition. *Tout polynôme annulateur $P \in K[X]$ de ϕ est multiple polynomial du polynôme minimal μ_ϕ de ϕ.*

Comme souvent, la division euclidienne est à la clé de ce résultat : comme P et μ_ϕ sont polynômes annulateurs de ϕ, c'est aussi le cas du reste $R = P - Q\mu_\phi$ de la division euclidienne de P par μ_ϕ (car $R[\phi] = P[\phi] - Q[\phi] \circ \mu_\phi[\phi] = 0$), et en vue de $\deg(R) < \deg(\mu_\phi)$ et la définition du polynôme minimal, ceci n'est possible que si $R = 0$. On remarque que cette démonstration est analogue à celle dans la définition 3.3.3 du fait que $D = \mathrm{pgcd}(A, B)$ divise A, la seule différence étant que la notion "polynôme annulateur de ϕ" remplace "combinaison polynomiale de A, B".

4.1.2. Théorème. *Les racines du polynôme minimal μ_ϕ forment l'ensemble des valeurs propres de ϕ. Plus généralement, on a $\dim(\ker(F[\phi])) > 0$ pour tout diviseur non constant F de μ_ϕ.*

Le point nouveau dans la première partie est que toute racine de μ_ϕ est forcément une valeur propre de ϕ (ce qui n'est pas toujours le cas pour une racine d'un polynôme annulateur). La second partie est une généralisation de la première, car en prenant $F = X - a$ pour une telle racine a, la conclusion $\dim(\ker(\phi - a\mathbf{I}_E)) > 0$ dit que a est valeur propre de ϕ. On raisonne par la minimalité de μ_ϕ : si l'on écrit $\mu_\phi = FQ$, le quotient Q ne peut pas être annulateur de ϕ car $Q \neq 0$ et $\deg(Q) < \deg(\mu_\phi)$. Mais bien que $Q[\phi] \neq 0$ on a $0 = \mu_\phi[\phi] = F[\phi] \circ Q[\phi]$, ce qui veut dire $\ker(F[\phi]) \supseteq \mathrm{Im}(Q[\phi])$. Alors $\dim(\ker(F[\phi])) \geq \dim(\mathrm{Im}(Q[\phi])) > 0$, comme voulu.

Si V est un sous-espace ϕ-stable, on peut former la restriction $\phi|_V$ de ϕ à ce sous-espace, qui est un endomorphisme de V et qui a donc son propre polynôme minimal $\mu_{\phi|_V}$. Une première question qu'on peut se poser est s'il y a une relation avec le polynôme minimal μ_ϕ. On voit facilement que $\mu_{\phi|_V}$ divise toujours μ_ϕ, car $\mu_\phi[\phi|_V]) = \mu_\phi[\phi]|_V = 0|_V = 0$ dit que μ_ϕ est multiple du polynôme minimal $\mu_{\phi|_V}$ de $\phi|_V$. Cette relation est comparable à la proposition 2.3.9 pour le polynôme caractéristique ; une différence pour le polynôme minimal est que égalité entre $\mu_{\phi|_V}$ et μ_ϕ est possible, même si V est un sous-espace ϕ-stable *strict* de E.

Une seconde question est si tous les diviseurs unitaires de μ_ϕ sont polynôme minimal d'une certaine restriction de ϕ à un sous-espace ϕ-stable. Le lemme suivant donne une réponse positive.

4.1.3. Lemme. *Si $\mu_\phi = QP$ est une décomposition avec P et Q unitaires, alors Q est égal au polynôme minimal $\mu_{\phi|_V}$ de la restriction $\phi|_V$ de ϕ au sous-espace ϕ-stable $V = \mathrm{Im}(P[\phi])$.*

Preuve. Un polynôme R est annulateur de $\phi|_V$ si et seulement si $\mathrm{Ker}(R[\phi]) \supseteq V = \mathrm{Im}(P[\phi])$, ce qui est équivalent (lemme 4.0.2) à RP est annulateur de ϕ, ou encore à RP est multiple de μ_ϕ. Clairement $R = \mu_\phi/P = Q$, est le polynôme unitaire du plus petit degré avec cette propriété.

Le sous-espace ϕ-stable $V' = \mathrm{Ker}(Q[\phi])$ parait un candidat plus naturel que $V = \mathrm{Im}(P[\phi])$ pour être sous-espace dont Q est polynôme minimal de la restriction de ϕ. Et en effet, V' a cette propriété, tout comme V. D'une part, l'argument ci-dessus montre que $V' \supseteq V$, ce qui implique que le polynôme minimal de $\phi|_{V'}$ est un multiple de $\mu_{\phi|_V} = Q$ (car $\phi|_V$ est restriction de $\phi|_{V'}$). D'autre part, Q est (par définition de V') déjà un polynôme annulateur de $\phi|_{V'}$. On a donc :

4.1.4. Proposition. *Si Q est diviseur unitaire de μ_ϕ, alors Q est le polynôme minimal de la restriction de ϕ au sous-espace ϕ-stable* $\mathrm{Ker}(Q[\phi])$.

On note que, même si cette proposition semble plus directe que le lemme 4.1.3, c'est ce dernier qui lui fournit une démonstration simple. Le lemme est aussi utile pour calculer μ_ϕ : si on trouve d'abord un polynôme unitaire P dont on sait qu'il divise μ_ϕ (par exemple le polynôme minimal P d'une restriction de ϕ), alors le facteur manquant $Q = \mu_\phi/P$ peut être trouvé comme le polynôme minimal de la restriction de ϕ au sous-espace $\mathrm{Im}(P[\phi])$. Si par chance on a déjà $P[\phi] = 0$, alors $Q = 1$ et donc $\mu_\phi = P$. Pour amorcer ce procédé, on peut choisir un vecteur $v \neq 0$ et à l'aide de ces images $\phi^k(v)$ chercher P unitaire minimal tel que $P[\phi](v) = 0$, qui divise forcément μ_ϕ (car P est polynôme minimal de $\phi|_V$ où $V = \mathrm{Vect}(v, \phi(v), \phi^2(v)\ldots)$).

4.2. *Décomposition des noyaux.*

4.2.1. Lemme. *Si $P, Q \in K[X]$ sont premiers entre eux, alors le sous-espace $V = \mathrm{Ker}((PQ)[\phi])$ se décompose en somme directe : $V = \mathrm{Ker}(P[\phi]) \oplus \mathrm{Ker}(Q[\phi])$. Les projecteurs de V sur les facteurs de la somme directe peuvent être écrits comme des polynômes en $\phi|_V$.*

Preuve. On voit facilement que $V = \mathrm{Ker}((PQ)[\phi])$ contient $\mathrm{Ker}(P[\phi])$ et $\mathrm{Ker}(Q[\phi])$. Posons donc $\phi' = \phi|_V$, la restriction de ϕ à V. On veut montrer que V est la somme directe de ses sous-espaces $V_1 = \mathrm{Ker}(P[\phi]) = \mathrm{Ker}(P[\phi'])$ et $V_2 = \mathrm{Ker}(Q[\phi]) = \mathrm{Ker}(Q[\phi'])$. C'est ici que des coefficients de Bézout pour P, Q seront utilisés ; soit donc $S, T \in K[X]$ tels que $SP + TQ = 1$. Puisque PQ est annulateur de ϕ', le lemme 4.0.2 dit que l'image $Q[\phi']$ est incluse dans V_1 et celle de $P[\phi']$ dans V_2. Pour tout $v \in V$ on peut écrire $v = (SP+TQ)[\phi](v) = P[\phi'](S[\phi'](v)) + Q[\phi'](T[\phi'](v)) \in \mathrm{Im}(P[\phi']) + \mathrm{Im}(Q[\phi']) \subseteq V_2 + V_1$, ce qui montre que $V_1 + V_2 = V$. On pose $\pi_1 = (TQ)[\phi']$ et $\pi_2 = (SP)[\phi']$. Alors π_2 s'annule sur V_1 et π_1 s'annule sur V_2 et à l'aide de $\pi_1 + \pi_2 = \mathbf{I}_V$ on en déduit que π_1 fixe les vecteurs de V_1 et π_2 fixe les vecteurs de V_2. Par conséquent si $v = v_1 + v_2$ avec $v_1 \in V_1$ et $v_2 \in V_2$, on a obligatoirement $v_1 = \pi_1(v)$ et $v_2 = \pi_2(v)$. Autrement dit, la somme $V_1 + V_2$ est directe et π_1, π_2 sont les projecteurs correspondants.

4.2.2. Théorème de décomposition des noyaux. *Si $P_1, \ldots, P_l \in K[X]$ sont premiers entre eux 2 à 2, et $\phi \in \mathrm{End}(E)$, alors*

$$\mathrm{Ker}((P_1 \ldots P_l)[\phi]) = \mathrm{Ker}(P_1[\phi]) \oplus \cdots \oplus \mathrm{Ker}(P_l[\phi]),$$

et chaque projecteur de la somme sur l'un des facteurs est la restriction d'un polynôme en ϕ.

Il s'agit d'une généralisation directe du lemme, prouvée par récurrence sur l en utilisant celui-ci. Pour $l < 2$ il n'y a rien à démontrer, et pour $l \geq 2$ on peut appliquer le lemme avec $P = P_1 \ldots P_{l-1}$ et $Q = P_l$, qui sont premiers entre eux grâce à l'hypothèse sur les facteurs P_i et la proposition 3.3.6, pour obtenir la décomposition $\mathrm{Ker}((P_1 \ldots P_l)[\phi]) = \mathrm{Ker}(P_1 \ldots P_{l-1})[\phi]) \oplus \mathrm{Ker}(P_l[\phi])$. L'hypothèse de récurrence fournit la décomposition supplémentaire en somme directe du premier sous-espace $\mathrm{Ker}(P_1 \ldots P_{l-1})[\phi])$ nécessaire pour avoir la décomposition du théorème. Pour les projecteurs, le polynôme donnant celui sur $\mathrm{Ker}(P_l[\phi])$ est le second donné par le lemme ; pour les autres, on multiplie les polynômes donnés par l'hypothèse de récurrence par le premier donné par le lemme. (Leur existence est plus intéressant que de savoir ces polynômes concrets.)

Comme annoncé, ce résultat permet de mieux comprendre le théorème 2.4.4. Dans le cas d'un polynôme annulateur de ϕ qui est scindé, les différents facteurs $X - \lambda_i$ sont tous irréductibles, donc ils sont premiers entre eux si et seulement s'ils sont distincts. Quand c'est le cas, le théorème de décomposition des noyaux s'applique, et donne comme résultat une décomposition de l'espace en somme directe de sous-espaces E_{λ_i}. Certains des ces sous-espaces peuvent éventuellement être de dimension 0 ; les autres sont des sous-espaces propres, et le fait que leur somme remplit E dit précisément que ϕ est diagonalisable avec ces valeurs propres.

4.3. *Sous-espaces caractéristiques, trigonalisation.*

4.3.1. Définition. *Soit ϕ un endomorphisme d'un K-espace E de dimension finie, et λ une valeur propre de ϕ. Le sous-espace caractéristique de ϕ pour λ est $\tilde{E}_\lambda = \mathrm{Ker}\big((\phi - \lambda \mathbf{I}_E)^{m_\lambda}\big)$, où m_λ est la multiplicité de ϕ comme racine du polynôme minimal μ_ϕ,*

Le sous-espace caractéristique $\tilde{E}_\lambda$ inclut le sous-espace propre E_λ (car les vecteurs annulés par $\phi - \lambda \mathbf{I}_E$ sont *a fortiori* annulés par $(\phi - \lambda \mathbf{I}_E)^{m_\lambda}$). D'après la proposition 4.1.4, le polynôme minimal de la restriction $\phi|_{\tilde{E}_\lambda}$ est $(X - \lambda)^{m_\lambda}$. Donc $\tilde{E}_\lambda$ inclut *strictement* E_λ si $m_\lambda > 1$ (c'est-à-dire si λ est une racine multiple de μ_ϕ, ce qui implique que ϕ n'est pas diagonalisable).

4.3.2. Théorème. *Si μ_ϕ est scindé, alors $E = \bigoplus_{i=1}^{k} \tilde{E}_{\lambda_i}$, une décomposition en somme directe des sous-espaces caractéristiques $\tilde{E}_{\lambda_1}, \ldots, \tilde{E}_{\lambda_k}$ pour les valeurs propres distinctes $\lambda_1, \ldots, \lambda_k$ de ϕ.*

Preuve. On décompose le polynôme minimal μ_ϕ en facteurs de la forme $(X - \lambda)^{m_\lambda}$, où on a regroupé tous les facteurs irréductibles identiques (donc m_λ est la multiplicité de λ comme racine de μ_ϕ). Ainsi les facteurs regroupés sont premiers entre eux grâce à la proposition 3.3.6, et on peut appliquer le théorème 4.2.2 à cette décomposition. Sa conclusion donne précisément celle du théorème actuel, en tenant compte de la définition des sous-espaces caractéristiques.

Le résultat principal de ce chapitre est l'existence de cette décomposition de E en somme directe de sous-espaces ϕ-stables $\tilde{E}_\lambda$. Par contraste, l'ensemble des sous-espaces propres E_λ forme bien somme directe de sous-espaces ϕ-stables, mais celle-ci ne remplit pas l'espace E dans le cas non diagonalisable, et elle n'est alors pas d'une grande utilité. Bien que la décomposition en sous-espaces caractéristiques soit définie en termes du polynôme minimal, la proposition suivante montre qu'elle peut être trouvée à l'aide de n'importe quel polynôme annulateur, notamment (grâce au théorème de Cayley-Hamilton) à l'aide du polynôme caractéristique.

4.3.3. Proposition. *Si $P = (X-a_1)^{e_1}\dots(X-a_k)^{e_k}$ est un polynôme annulateur de ϕ (avec les a_i distincts), alors on a une décomposition en somme directe $E = V_1 \oplus \cdots \oplus V_k$, où $V_i = \mathrm{Ker}((\phi - a_i\mathbf{I}_E)^{e_i})$. Les V_i de dimension non nulle sont des sous-espaces caractéristiques. Plus précisément ce sont les V_i pour lequel a_i est une valeur propre λ, et dans ces cas $V_i = \tilde{E}_\lambda$; aussi $e_i \geq m_\lambda$, où m_λ est l'exposant utilisée pour cette valeur propre dans la définition de $\tilde{E}_\lambda$.*

La démonstration du théorème 4.3.2 s'applique ici aussi pour obtenir la décomposition $E = V_1 \oplus \cdots \oplus V_k$, les seuls faits utilisés étant d'avoir un polynôme annulateur (et donc $E = \mathrm{Ker}(0) = \mathrm{Ker}(P[\phi])$) et d'en utiliser une décomposition en facteurs premier entre eux. Le reste de l'énoncé dit que c'est essentiellement la même décomposition que celle du théorème 4.3.2 (qui existe car μ_ϕ, qui divise le polynôme annulateur P, est scindé). Si a_i n'est pas valeur propre de ϕ, alors $\phi - a_i\mathbf{I}_E$ est inversible et donc $V_i = \{0\}$. Supposons donc que $a_i = \lambda$ soit une valeur propre. Comme μ_ϕ divise P, la multiplicité de $X-\lambda$ dans P est au moins aussi grande que celle dans μ_ϕ : on a $e_i \geq m_\lambda$, et donc $V_i \supseteq \tilde{E}_\lambda$. Il reste à montrer l'inclusion dans l'autre sens. Or, le polynôme minimal de $\phi|_{V_i}$ divise à la fois $(X-\lambda)^{e_i}$ (par définition de V_i) et μ_ϕ (car $V_i \subseteq E$) donc il divise $\mathrm{pgcd}((X-\lambda)^{e_i}, \mu_\phi) = (X-\lambda)^{m_\lambda}$. Alors $(X-\lambda)^{m_\lambda}[\phi|_{V_i}] = 0$ et donc $V_i \subseteq \tilde{E}_\lambda$.

L'utilité de cette proposition est qu'on peut trouver les sous-espaces caractéristiques même sans connaître explicitement le polynôme minimal μ_ϕ. Celui-ci est utilisé dans la définition 4.3.1 car il fallait bien indiquer un exposant concret pour $\phi - \lambda\mathbf{I}_E$. Mais le noyau en question ne change pas si l'on utilise un exposant $e > m_\lambda$, comme le montre la proposition.

En revanche $(X-\lambda)^{m_\lambda}$ est le polynôme *minimal* de $\phi|_{\tilde{E}_\lambda}$ (proposition 4.1.4), donc on ne peut pas utiliser un exposant $e < m_\lambda$: le noyau $\mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^e\big)$ serait alors plus petit que $\tilde{E}_\lambda$. En fait, tous les exposants $e \leq m_\lambda$ donnent comme noyaux des sous-espaces *différents* :

4.3.4. Proposition. *Si λ est une valeur propre de ϕ avec multiplicité m_λ comme racine de μ_ϕ, alors on a des inclusions strictes $\mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^{i-1}\big) \subset \mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^i\big)$ pour $0 < i \leq m_\lambda$.*

Les inclusions faibles $\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^{i-1}\big) \subseteq \mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^i\big)$ sont claires (le polynôme à droite est un multiple de celui à gauche). Il suffit donc de montrer qu'elles sont strictes, ce qui revient à établir pour chaque i qu'il existe un vecteur dans la différence $\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^{i+1}\big)\setminus\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^i\big)$, c'est-à-dire un vecteur v_i tel que $(\phi - \lambda\mathbf{I}_E)^i(v_i) = 0$ mais $(\phi - \lambda\mathbf{I}_E)^{i-1}(v_i) \neq 0$ (l'indice i de v_i compte son «nombre de vies restantes» par rapport aux applications répétées de $\phi-\lambda\mathbf{I}_E$). Pour $i = m_\lambda$ l'inclusion est bien stricte, car $(X-\lambda)^{m_\lambda}$ est le polynôme minimal de $\phi|_{\tilde{E}_\lambda}$ comme on vient de le remarquer ; on peut donc choisir un vecteur $v_{m_\lambda} \in \tilde{E}_\lambda \setminus \mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^{m_\lambda-1}\big)$. Ce vecteur ayant le nombre maximal m_λ de vies restantes, il est facile d'en déduire un vecteur v_i avec $i < m_\lambda$ de vies restantes : il suffit d'appliquer $m_\lambda - i$ fois l'endomorphisme $\phi - \lambda\mathbf{I}_E$ pour qu'il en reste $m_\lambda - (m_\lambda - i) = i$. Autrement dit on pose $v_i = (\phi - \lambda\mathbf{I}_E)^{m_\lambda - i}(v_{m_\lambda})$ pour $0 < i < m_\lambda$, et on vérifie facilement que $v_i \in \mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^{i+1}\big) \setminus \mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^i\big)$.

En complément de ce côté strict des inclusions, on peut démontrer (mais on ne le fera pas ici ; les courageux peuvent l'essayer comme exercice) que dans ces inclusions la différence (non nulle) des deux dimensions va en décroissant (au sens large) quand i va en croissant de 1 à m_λ.

Avant de poursuivre, on veut régler un détail technique : le théorème 4.3.2 a pour hypothèse que μ_ϕ est scindé, mais on aimerait parfois la remplacer par la condition que le polynôme *caractéristique* χ_ϕ soit scindé. C'est possible d'après le théorème de Cayley-Hamilton, qui dit que μ_ϕ divise χ_ϕ (tout diviseur d'un polynôme scindé est scindé). Mais sans utiliser ce théorème (qu'on n'a pas encore démontré), on peut aussi raisonner ainsi. Supposons χ_ϕ scindé sur K, mais μ_ϕ non scindé. Alors μ_ϕ a un facteur irréductible F dans $K[X]$ avec $\deg(F) \geq 2$, et donc sans racine dans K. Le polynôme minimal de la restriction $\phi|_V$ de ϕ à $V = \ker(F[\phi])$

est F (proposition 4.1.4), et $\phi|_V$ n'a donc pas de valeurs propres dans K ; alors son polynôme caractéristique $\chi_{\phi|_V}$ est aussi sans racine dans K, et il n'est pas constant (car $\deg(\chi_{\phi|_V}) = \dim(V) > 0$, voir théorème 4.1.2). Mais cela contredit le fait (proposition 2.3.9) que $\chi_{\phi|_V}$ divise le polynôme χ_ϕ, qui était supposé scindé. En conclusion, si χ_ϕ est scindé, alors μ_ϕ l'est aussi.

4.3.5. Définition. *On appelle $\phi \in \mathrm{End}(E)$ trigonalisable s'il existe une base $\mathcal{B}$ de E tel que $\mathrm{Mat}_{\mathcal{B}}(\phi)$ soit une matrice triangulaire supérieure. On appelle $\mathcal{B}$ une base de trigonalisation,*

4.3.6. Proposition. *Chaque restriction $\phi|_{\tilde{E}_\lambda}$ de ϕ à un sous-espace caractéristique est trigonalisable, et les coefficients diagonaux de la matrice triangulaire en question sont tous λ.*

Preuve. Pour la base de trigonalisation on peut commencer avec une base de l'espace propre $E_\lambda = \mathrm{Ker}(\phi - \lambda\mathbf{I}_E)$, l'étendre à une base de $\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^2\big)$, puis à une base de $\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^3\big)$, et ainsi de suite, jusqu'à l'obtention d'une base $\mathcal{B}$ de $\mathrm{Ker}\big((\phi-\lambda\mathbf{I}_E)^{m_\lambda}\big) = \tilde{E}_\lambda$. L'image par $\phi-\lambda\mathbf{I}_E$ de chaque vecteur de $\mathcal{B}$ est située dans le sous-espace engendré par les vecteurs *précédents* de $\mathcal{B}$, d'où la matrice $\mathrm{Mat}_{\mathcal{B}}((\phi - \mathbf{I}_E)|_{\tilde{E}_\lambda})$ est triangulaire *strictement* supérieure. La proposition suit.

4.3.7. Corollaire. *Si μ_ϕ scindé (ou si χ_ϕ est scindé), alors ϕ est trigonalisable sur K.*

Il suffit de décomposer l'espace comme dans le théorème 4.3.2, et de choisir dans chaque espace caractéristique $\tilde{E}_\lambda$ une base de trigonalisation (proposition 4.3.6). La matrice de ϕ par rapport à la réunion de ces bases est (diagonale en blocs, et) triangulaire supérieure.

D'après le théorème 4.3.2, le polynôme caractéristique χ_λ se décompose comme le produit des polynômes caractéristiques des restrictions de ϕ aux sous-espaces caractéristiques, et d'après la proposition 4.3.6, le sous-espace $\tilde{E}_\lambda$ contribue ainsi le facteur $(X-\lambda)^{\dim(\tilde{E}_\lambda)}$ à χ_λ. On a donc :

4.3.8. Proposition. *La dimension de $\tilde{E}_\lambda$ est égale à la multiplicité de λ comme racine de χ_λ.*

Finalement, le polynôme caractéristique d'une matrice triangulaire T est toujours scindé (car $\chi_T = \prod_{i=1}^n (X - T_{i,i})$ où les $T_{i,i}$ sont les coefficients diagonaux), ce qui donne après changement de base vers une base de trigonalisation la réciproque du corollaire 4.3.7. Ainsi on a :

4.3.9. Théorème. *Les conditions suivantes sont équivalentes :*
(i) χ_ϕ est scindé sur K,
(ii) μ_ϕ scindé sur K,
(iii) ϕ est trigonalisable sur K.

4.4. *Quelques approches du théorème de Cayley-Hamilton.*

Jusqu'ici on a annoncé le théorème 2.3.10 de Cayley-Hamilton (qui dit que le polynôme caractéristique χ_ϕ est toujours un polynôme annulateur de ϕ), mais on ne l'a pas démontré (ni utilisé). Dans cette dernière section du cours on indique quelques parmi les très nombreuses approches pour démontrer ce théorème.

Une des raisons de la variété des approches est que déjà il y a différentes manières de formuler l'énoncé. Par exemple, dire que χ_ϕ est polynôme annulateur de ϕ équivaut, d'après la proposition 4.1.1, à dire que la polynôme minimal μ_ϕ divise χ_ϕ. Mais on peut aussi prendre le point de vue que $\chi_\phi[\phi] = 0 \in \mathrm{End}(E)$ veut dire concrètement que $\chi_A[A] = 0 \in \mathrm{Mat}_n(K)$ pour la matrice (carrée) de A par rapport à une base de E, ce qui doit donc être vrai pour *toutes* les matrices carrées A. Sous cette forme il est clair qu'on peut remplacer K par un corps plus grand, par exemple par $\mathbf{C}$ si $K = \mathbf{Q}$ ou $K = \mathbf{R}$, car cela ne change pas la matrice $\chi_A[A]$.

L'énoncé $\chi_A[A] = 0$ affirme un système de n^2 identités algébriques compliquées (pour chacun des coefficients de la matrice $\chi_A[A]$) en n^2 variables (à savoir les coefficients de A). Ce dernier point de vue ne donne pas une piste pour directement prouver ces identités (sauf pour des valeurs de n concrètes, tel $n = 2$), mais permet des réflections de ce genre : si ces identités sont valables pour toutes les valeurs des variables, alors elles doivent être une conséquence d'identités algébriques générales (telles que la commutativité, la distributivité), et ne pas dépendre du corps K utilisé du tout (techniquement : on pourra remplacer K par un anneau commutatif).

• *Raisonnement par la densité des matrices diagonalisables.*

Un argument sans calcul algébrique pour prouver le théorème pour le cas $K = \mathbf{C}$ est basé sur l'observation que l'ensemble des matrices $A \in \mathrm{Mat}_n(\mathbf{C})$ pour lesquelles $\chi_A[A] \neq 0$ (dont on veut montrer qu'il est vide) est certainement *ouvert* : si A est dans cet ensemble, l'un au moins des coefficients de $\chi_A[A]$ est non nul, et pour des perturbations suffisamment petites de A il le restera. Il suffira donc de montrer que $\chi_A[A] = 0$ est valable pour A dans un ensemble *dense* dans $\mathrm{Mat}_n(\mathbf{C})$, c'est-à-dire qui rencontre tout sous-ensemble ouvert et non vide. Il est facile à voir que $\chi_A[A] = 0$ est valable quand A est diagonalisable, car $\chi_A[A]$ agit par le scalaire $\chi_A[\lambda]$ sur le sous-espace propre E_λ, et chaque valeur propre λ de A est une racine de χ_A. Or l'ensemble des matrices diagonalisables est dense dans $\mathrm{Mat}_n(\mathbf{C})$, car il inclut l'ensemble des A pour lesquels χ_A est à racines simples, qui est déjà dense (cela demande une preuve, qu'on omet ici).

• *Raisonnement par la trigonalisation.*

L'argument ci-dessus, même s'il est convainquant, n'est pas tout à fait satisfaisant, car il montre une identité algébrique par un argument d'approximation (la densité), ce qui ne devait pas être nécessaire. Tout en restant dans le cas $K = \mathbf{C}$, on peut éviter cette approximation en étendant l'argument qui montre $\chi_A[A] = 0$ au cas où A est trigonalisable (au lieu de diagonalisable), car *toute* matrice carrée complexe est trigonalisable (corollaire 4.3.7). Si A est une matrice triangulaire supérieure, on a $\chi_A = \prod_{i=1}^n (X - a_{i,i})$ où les $a_{i,i}$ sont les coefficients diagonaux de A. Or on peut montrer assez facilement, par récurrence sur n, que dans ce cas le produit matriciel $(A - a_{1,1}\mathbf{I})(A - a_{2,2}\mathbf{I}) \dots (A - a_{n,n}\mathbf{I})$ est nul (le produit des $n-1$ premiers facteurs est nul en dehors de sa dernière colonne par l'hypothèse de récurrence, et la dernière ligne du facteur final $A - a_{n,n}\mathbf{I}$ est nulle), ce qui prouve $\chi_A[A] = 0$ pour A triangulaire, ou trigonalisable.

• *Raisonnement par l'étude des espaces caractéristiques.*

D'après la proposition 4.3.6, chaque espace caractéristique $\tilde{E}_\lambda$ d'un endomorphisme ϕ, de dimension disons d, admet une base telle que la matrice B de la restriction $(\phi - \lambda\mathbf{I})|_{\tilde{E}_\lambda}$ est triangulaire strictement supérieure, c'est-à-dire triangulaire supérieure et avec coefficients 0 sur la diagonale principale. Cette valeur propre contribue un facteur $(X - \lambda)^d$ au polynôme caractéristique χ_ϕ, donc pour montrer que $\chi_\phi[\phi]$ s'annule sur $\tilde{E}_\lambda$ il suffit de montrer que $B^d = 0$ pour une telle matrice, ce qui se fait facilement (par récurrence sur la taille d de la matrice). Une façon un peu différente d'arriver à la même conclusion que $(\phi - \lambda\mathbf{I})^d$ s'annule sur $\tilde{E}_\lambda$, est d'observer que ce dernier est par définition annulé par $(\phi - \lambda\mathbf{I})^{m_\lambda}$ avec m_λ comme dans la proposition 4.3.4, d'où il suffit de montrer que $m_\lambda \leq d = \dim\big(\mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^{m_\lambda}\big)\big)$; mais cela découle de la proposition citée, car la dimension des espaces $\mathrm{Ker}\big((\phi - \lambda\mathbf{I}_E)^i\big)$ monte par au moins 1 pour chaque inclusion stricte.

• *En utilisant les propriétés de la matrice compagnon.*

Les preuves précédents ont un défaut esthétique dans la mesure qu'elles utilisent un argument structurelle (une décomposition de l'espace E sur lequel agit ϕ), mais seulement après avoir utilisé un argument algébrique formelle pour justifier de remplacement du corps K par $\mathbf{C}$ (ou

plus généralement par un corps $K' \supseteq K$ suffisamment grand pour que μ_ϕ se scinde sur K', quel corps existe toujours) pour s'assurer que ladite décomposition existe. On peut éviter ce remplacement et travailler directement avec le corps K, mais alors on ne pourra pas se servir des sous-espaces caractéristiques, qui peuvent ne pas exister (si χ_ϕ et μ_ϕ sont sans racines dans K).

L'argument pour prouver que $\chi_\phi[\phi] = 0$ sera par récurrence forte sur $n = \dim(E)$; comme hypothèse de récurrence on suppose donc que pour toute matrice carrée M de taille $< n$ on ait $\chi_M[M] = 0$. Le cas où $\dim(E) = 0$ est trivial car l'unique endomorphisme d'un tel espace est nul ; on peut donc supposer $\dim(E) > 0$ et choisir un vecteur non nul $v \in E$ quelconque. Puis on forme une famille de vecteurs $[v, \phi(v), \phi^2(v), \ldots]$, en continuant tant que cette famille est libre. On s'arrête donc quand une telle image ϕ^d se trouver dans le sous-espace $V = \mathrm{Vect}\big(v, \phi(v), \ldots, \phi^{d-1}(v)\big)$ engendré par les vecteurs précédents. Par construction $\mathcal{B}_V = [v, \phi(v), \ldots, \phi^{d-1}(v)]$ est une base de V, qui est ϕ-stable (car il contient les images par ϕ de tous les vecteurs de $\mathcal{B}_V$), et en fait le plus petit sous-espace ϕ-stable V qui contient v. L'expression $\phi^d(v) = c_0 v + c_1\phi(v) + \cdots + c_{d-1}\phi^{d-1}(v)$ nous donne le polynôme unitaire $P = X^n - c_{n-1}X^{n-1} - \cdots - c_0X^0$ du plus petit degré tel que $P[\phi](v) = 0$.

Tout ce qu'on peut dire sur $\dim(V)$ est $0 < \dim(V) \leq \dim(E)$ et le plus souvent on aura $V = E$; dans ce cas avoir trouvé ce sous-espace ϕ-stable ne semble pas avoir un grand intérêt. Mais la matrice de la restriction $\phi|_V$ par rapport à $\mathcal{B}$ a la forme particulière

$$\begin{pmatrix} 0 & 0 & 0 & \ldots & c_0 \\ 1 & 0 & 0 & \ldots & c_1 \\ 0 & 1 & 0 & \ldots & c_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & c_{d-1} \end{pmatrix} = C_P \tag{2}$$

connue comme la matrice compagnon du polynôme P. Puisque $\mathrm{Ker}(P[\phi])$ est un sous-espace ϕ-stable qui contient v, on a $V \subseteq \mathrm{Ker}(P[\phi])$, autrement dit $P[\phi|_V] = 0$, et P est donc le polynôme minimal de $\phi|_V$ et de C_P. On montre également par un calcul directe que P est le polynôme *caractéristique* de C_P (c'est un exercice instructif). Ceci montre que dans le cas $V = E$ l'énoncé du théorème est vérifié, et même un peu plus car alors $\chi_\phi = \mu_\phi$, ce qui n'est pas vrai en général.

Pour le cas général il faut compléter la base $\mathcal{B}_V$ à une base $\mathcal{B}$ de E. Alors en divisant $\mathrm{Mat}_\mathcal{B}(\phi)$ en 4 blocs selon la partie initiale $\mathcal{B}_V$ de $\mathcal{B}$ et le reste, on obtient la forme en blocs

$$\mathrm{Mat}_\mathcal{B}(\phi) = \begin{pmatrix} C_P & B \\ \mathbf{0} & D \end{pmatrix} = A,$$

où D est une matrice carrée pour laquelle on peut appliquer l'hypothèse de récurrence, qui donne $\chi_D[D] = 0$. On sait aussi que $\chi_\phi = \chi_A = \chi_{C_P} \cdot \chi_D = P.\chi_D$. Un calcul direct montre que pour de telles matrice triangulaire en blocs que $\left(\begin{smallmatrix} X & * \\ \mathbf{0} & Y \end{smallmatrix}\right)^n = \left(\begin{smallmatrix} X^n & * \\ \mathbf{0} & Y^n \end{smallmatrix}\right)$ pour $n \in \mathbf{N}$ (où '$*$' désigne chaque fois un bloc différent, dont le contenu nous n'intéresse pas), et donc $Q[\left(\begin{smallmatrix} X & * \\ \mathbf{0} & Y \end{smallmatrix}\right)] = \left(\begin{smallmatrix} Q[X] & * \\ \mathbf{0} & Q[Y] \end{smallmatrix}\right)$ pour tout $Q \in K[X]$. Enfin, $\chi_\phi[\phi]$ calculé dans la base $\mathcal{B}$ sous la forme de matrices en blocs, est :

$$\begin{aligned} \chi_A[A] = (P.\chi_D)[A] = P[A] \cdot \chi_D[A] &= \begin{pmatrix} P[C_P] & * \\ \mathbf{0} & P[D] \end{pmatrix} \cdot \begin{pmatrix} \chi_D[C_P] & * \\ \mathbf{0} & \chi_D[D] \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{0} & * \\ \mathbf{0} & * \end{pmatrix} \cdot \begin{pmatrix} * & * \\ \mathbf{0} & \mathbf{0} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}. \end{aligned}$$

• *En utilisant l'algèbre purement formelle.*

Tous les arguments donnés jusqu'ici sont—à un degré variable—structurelles (ou géométriques), c'est à dire ils sont basés sur une considération de certains sous-espaces de l'espace vectoriel E qui dépendent de l'endomorphisme ϕ, et de certaines bases de E en rapport avec ces sous-espaces qui servent pour représenter ϕ par une matrice d'une forme particulière. Il existe cependant des preuves du théorème de Cayley-Hamilton qui se passent de toute approche structurelle, et qui démontrent $\chi_A[A] = 0$ pour toute matrice carrée A par une approche directe, n'utilisant que des manipulations d'expressions algébriques, basées sur leurs propriétés formelles (identités de différents types). On ne décrira pas de telles preuves ici, car le type d'algèbre formelle nécessaire (par exemple concernant les déterminants) n'est pas beaucoup abordé dans ce cours ; les introduire juste pour présenter une telle preuve ne semble pas très utile. Mais c'est une approche importante, et les preuves qu'elle fournit sont d'un certain point de vue les plus satisfaisantes et les plus générales parmi les preuves connues du théorème.

Table de matières.