3L05 : Algèbre Multilinéaire

Avant-propos.

Ce document sert de polycopié pour le cours avec l'intitulé ci-dessus. Il résulte de ce que j'ai donné ce cours pendant les années universitaires 2010–2011 et 2011–2012. La première année j'ai donné le cours en rédigeant des notes manuscrites ; la rédaction du cours sous forme informatique a eu lieu essentiellement en automne 2011, et a apporté des changements importants à son organisation. Le cours avait été donné les deux années précédentes respectivement par mes collègues Jean Souville et par Rupert Yu, et je leur suis reconnaissant de m'avoir transmis les documents qu'ils ont préparés pendant ce temps. Les deux exemples m'ont montré qu'il existe un vaste éventail de possibilités pour présenter la matière, pourtant classique, qui est l'objet de ce cours.

J'ai cherché à trouver une présentation personnelle qui tienne compte à la fois du niveau limité des connaissances acquises dans le cours d'algèbre linéaire de la première année universitaire (qui est commun aux différents parcours de sciences et mathématiques), et de la nécessité pour les étudiants de mathématiques de se familiariser avec des notions plus structurelles en algèbre (comme les sous-espaces, choix de bases, homomorphismes, quotients). La maîtrise de ces notions est réputée plus difficile que celles des méthodes calculatoires, mais j'ai essayé de limiter le degré d'abstraction utilisé. Si certaines définitions sont formulées dans un cadre général (espaces vectoriels sur un corps commutatif K, idéaux dans $K[X]$, déterminant d'une matrice carrée à coefficients dans un anneau commutatif), c'est parce que cette généralité s'avère directement utile dans le cours.

Je considère comme but principal de ce cours d'aborder les questions autour des problèmes des vecteurs propres, et plus généralement les notions de "réduction d'endomorphismes" (une expression consacrée en France ; je n'ai jamais compris dans quel sens un endomorphisme peut être réduit). Les polynômes et déterminants figurent dans ce cours principalement comme outils pour atteindre ce but, la notion de polynôme caractéristique touchant de façon évidente aux deux sujets. Malgré son rôle utilitaire, le traitement des déterminants est assez complet, et il peut être considéré comme but secondaire du cours ; cela se justifie non seulement par le titre du cours, mais aussi par le fait qu'aucun cours ultérieur ne traitera les déterminants en profondeur, pendant qu'on se servira d'eux à diverses occasions. Pour les polynômes la situation est différente, et pour cette raison leur traitement se limite aux connaissances utiles pour ce cours. Je regrette d'ailleurs que le chapitre 3 fasse une interruption si importante des considérations de l'algèbre linéaire, et pour parler des choses qu'on pourrait supposer déjà familières, mais il est nécessaire d'assurer un fondement solide sur les polynômes, pour pouvoir en faire un nombre d'applications (notamment les polynômes d'un endomorphisme et la définition de son polynôme minimal). J'ai profité de l'occasion pour renforcer dans ce chapitre d'autres connaissances de base qui à un moment ou un autre doivent être acquises, notamment concernant les structures quotient.

L'organisation du cours de 2010–2011 était basée sur l'idée de ne pas attendre l'introduction des déterminants (et donc du polynôme caractéristique) avant d'aborder la recherche des valeurs propres, le faisant dans un premier temps à l'aide du polynôme minimal. Mais cette approche me semble maintenant une erreur pédagogique : le fait que la définition complète du polynôme caractéristique demande plus de préparation que celle du polynôme minimal (et que son degré est parfois plus élevé) ne semble en rien dissuader les étudiants de l'utiliser (même avant qu'il ne soit introduit dans le cours !). Pour cette raison l'introduction du polynôme minimal est maintenant reportée au dernier chapitre du cours.

En rédigeant, le texte est devenu beaucoup plus long que mes notes manuscrites ne l'étaient, ou que je ne l'avais imaginé au départ. Je reconnais avoir un style de discours peu succinct, et être incapable d'écrire dans le style définition–théorème–preuve souvent trouvé dans les textes français. Aussi de nombreuses remarques que je jugeait importantes de faire quelque part se sont glissées dans le texte au fur et à mesure. Je souligne que le texte est fait pour expliquer, et non pas pour être appris par cœur. J'espère que le lecteur saura reconnaître les quelques énoncés importants à retenir tels quels (souvent marqués "théorème") ; en cas de doute, un résumé des objectifs du cours est donné à la fin. Sinon, divers livres sous le titre *Algèbre Linéaire* consultables à la BU peuvent servir de référence complémentaire.

9 novembre 2011 *Marc van Leeuwen*

Introduction.

Ce cours est une continuation du cours de l'algèbre linéaire de la première année. On développera les techniques de l'algèbre linéaire avec le but notamment d'étudier les *endomorphismes* d'un espace vectoriel (toujours supposé de dimension finie), c'est-à-dire les applications linéaires de l'espace vers lui-même, qui dans une base de l'espace peuvent être exprimées par une matrice carrée. Dans cette étude les *valeurs propres* d'un endomorphisme, et les vecteurs propres associés, joueront un rôle important. Pour la recherche effective de ces valeurs propres, on aura besoin de certains *polynômes*, notamment du polynôme caractéristique d'un endomorphisme, dont la définition fait intervenir l'opération du *déterminant*. Un chapitre du cours sera alors dédié à une courte introduction aux polynômes du point de vue algébrique (par opposition à l'étude des fonctions polynomiales), et un autre au développement de la notion du déterminant (dont le traitement en première année était limité aux aspects calculatoires et aux matrices de petite taille). C'est d'ailleurs le caractère multilinéaire du déterminant qui a inspiré le titre de ce cours, un titre qui ne couvre donc qu'une petite partie du contenu. Le dernier chapitre du cours fera une synthèse des notions abordés dans le cadre de l'étude des endomorphismes.

Chapitre 1. Rappels de l'algèbre linéaire.

Dans ce premier chapitre on fera un rappel des notions de l'algèbre linéaire qui ont été introduites (ou devraient l'avoir été) en première année. Mais si en première année il y avait un accent sur le calcul vectoriel et matriciel (dont la maîtrise est bien sûr une condition nécessaire pour la compréhension de l'algèbre linéaire), on se rendra vite compte que cela ne suffit pas, et que pour progresser on aura besoin d'une approche plus conceptuelle. Loin d'être redondant pour ceux qui ont réussi le cours de première année, ce chapitre servira à revoir les notions de ce point de vue, dont l'expérience montre qu'il est en général considéré comme plus difficile à appréhender.

1.1. ***Espaces vectoriels, sous-espaces, combinaisons linéaires, applications linéaires.***

Dans l'algèbre linéaire on commence par fixer une fois pour toute un ensemble K de *scalaires*, qui sont des valeurs numériques qui peuvent être additionnées, soustraites, multipliées et divisées entre elles (à l'exception de la division par 0 qui n'est pas définie), et par lesquelles les vecteurs pourront être multipliés. La raison de cette abstraction est que différents ensembles (plus précisément ; différents corps commutatifs ; une notion qui sera définie de façon précise dans le chapitre 3) de scalaires peuvent être utilisés sans que cela change la description de l'algèbre linéaire ; on peut penser notamment aux ensembles de nombres rationnels $\mathbf{Q}$, aux nombres réels $\mathbf{R}$, ou aux nombres complexes $\mathbf{C}$.

Un espace vectoriel E sur K (ou un K-espace vectoriel pour faire court) est un ensemble de valeurs mathématiques appelées vecteurs, qui peuvent être additionnés et soustraits entre eux, ainsi que multipliés par des scalaires. La nature précise des vecteurs n'est pas spécifiée : si les ensembles K^n des n-uplets de scalaires (pour $n \in \mathbf{N}$ fixe) forment les exemples les plus connus des K-espaces vectoriels, il y a des applications importantes où les espaces vectoriels sont des ensembles de fonctions, ou des suites, ou des polynômes, ou des solutions de systèmes d'équations, ou d'autres objets encore. Ce qui rend possible l'étude de toute cette diversité de possibilités au même temps est que l'algèbre linéaire s'intéresse uniquement aux opérations indiquées (addition, multiplication scalaire), et aux relations que peuvent être exprimés en termes de ces opérations. Il faudra supposer bien sûr que quelque propriétés de base (des axiomes) sont vérifiés pour pouvoir raisonner en toute généralité. Ces propriétés sont notamment que les opérations sont définies pour tous les arguments du bon type (deux vecteurs peuvent toujours être additionnés, sans exception, et le résultat sera un vecteur), et que certaines égalités sont toujours vérifiées (comme la loi distributive $\lambda(v+w) = \lambda v + \lambda w$ pour $\lambda \in K$ et $v, w \in E$) ; la liste complète de ces axiomes est longue mais bien connue et on la ne répétera pas ici.

Si on a une collection de vecteurs $v_1, \ldots, v_l \in E$ et des scalaires $\lambda_1, \ldots, \lambda_l \in K$, on peut former la *combinaison linéaire* $\lambda_1 v_1 + \cdots + \lambda_l v_l$, qui est un vecteur de E.* Un *sous-espace vectoriel* de E est un

* Une subtilité est que parfois, en parlant de "combinaison linéaire", on désigne l'expression $\lambda_1 v_1 + \cdots + \lambda_l v_l$

sous-ensemble S contenant le vecteur nul $\vec{0}$, et qui est fermé pour l'addition (la somme de deux vecteurs de S est toujours dans S) et pour la multiplication scalaire (un multiple scalaire d'un vecteur de S est toujours dans S). Il en découle que S est aussi fermé pour les combinaisons linéaires (si $v_1, \ldots, v_l \in S$ alors on a aussi $\lambda_1 v_1 + \cdots + \lambda_l v_l \in S$, quels que soient les scalaires $\lambda_1, \ldots, \lambda_l \in K$), et les sous-espaces vectoriels sont les seuls sous-ensembles non vides de E qui sont fermés pour les combinaisons linéaires.

La notion de sous-espace vectoriel est fondamentale, d'une part parce qu'ils sont eux-mêmes des espaces vectoriels (comme le suggère leur nom), mais aussi parce qu'ils sont nécessaires dans la description de par exemple les solutions d'un système d'équations linéaires (ou encore de l'image et le noyau d'une application linéaire). Un sous-espace vectoriel est un ensemble, qui permet en général plusieurs descriptions, dont aucune n'est privilégiée (car une énumération de *tous* leurs éléments n'est en général pas possible). Il y a deux types de description principaux d'un sous-espace vectoriel : une description par des *générateurs*, des vecteurs particuliers du sous-espace dont on peut déduire tous ses vecteurs, et une description par des équations, le sous-espace étant l'ensemble de vecteurs qui vérifient ces équations. Une description par générateurs prend la forme $S = \mathrm{Vect}(v_1, \ldots, v_l)$, qui est par définition l'ensemble de toutes les combinaisons linéaires $\lambda_1 v_1 + \cdots + \lambda_l v_l$ des vecteurs $v_1, \ldots, v_l$ (les générateurs). On peut vérifier facilement qu'un tel ensemble $\mathrm{Vect}(v_1, \ldots, v_l)$ est toujours un sous-espace vectoriel, qu'il contient $v_1, \ldots, v_l$, et que c'est le plus petit tel sous-espace : un sous-espace vectoriel qui contient $v_1, \ldots, v_l$ contiendra forcément $\mathrm{Vect}(v_1, \ldots, v_l)$ tout entier.

Une autre notion fondamentale est celle d'une application linéaire entre des K-espaces vectoriels. (C'est sans doute pour traiter de façon uniforme un grand nombre d'opérations très disparates comme la différentiation et intégration de fonctions, évaluation de polynômes, rotations et projections dans l'espace, mais qui ont le caractère *linéaire* en commun, que l'algèbre linéaire comme discipline abstraite a été développée.) Si E, E' sont deux K-espaces vectoriels, une application $f : E \to E'$ est dite (K-)linéaire si elle vérifie

$$\begin{aligned} f(x+y) &= f(x) + f(y) \\ f(\lambda x) &= \lambda f(x) \end{aligned} \tag{1}$$

pour tout $x, y \in E$ et tout $\lambda \in K$. Une façon alternative de caractériser les application linéaires est qu'elles sont compatibles avec les combinaisons linéaires :

$$f(\lambda_1 v_1 + \cdots + \lambda_l v_l) = \lambda_1 f(v_1) + \cdots + \lambda_l f(v_l) \tag{2}$$

pour tout $v_1, \ldots, v_l \in E$ et tout $\lambda_1, \ldots, \lambda_l \in K$. Un exemple typique d'une application linéaire est l'opération de la combinaison linéaire elle-même, avec une liste de vecteurs fixée, qu'on décrit ainsi. Fixons des vecteurs $\vec{v}_1, \ldots, \vec{v}_l$, dans un K-espace qu'on appelle E' car ce sera l'espace d'arrivée. L'espace de départ sera K^l, dont les éléments sont les l-uplets $(c_1, \ldots, c_l)$ de scalaires ; comme on le sait bien, c'est un K-espace vectoriel avec addition et multiplication scalaire composante par composante :

$$\begin{aligned} (c_1, \ldots, c_l) + (d_1, \ldots, d_l) &= (c_1 + d_1, \ldots, c_l + d_l) \quad \text{et} \\ \lambda(c_1, \ldots, c_l) &= (\lambda c_1, \ldots, \lambda c_l). \end{aligned} \tag{3}$$

L'application de combinaison linéaire $f : K^l \to E'$ est définie par $f : (c_1, \ldots, c_l) \mapsto c_1\vec{v}_1 + \cdots + c_l\vec{v}_l$. On vérifie facilement qu'elle vérifie les identités de (1) ; par exemple pour la première identité on a

$$\begin{aligned} f : ((c_1, \ldots, c_l) + (d_1, \ldots, d_l)) &= f(c_1 + d_1, \ldots, c_l + d_l) = (c_1 + d_1)\vec{v}_1 + \cdots + (c_l + d_l)\vec{v}_l \\ &= (c_1\vec{v}_1 + \cdots + c_l\vec{v}_l) + (d_1\vec{v}_1 + \cdots + d_l\vec{v}_l) = f(c_1, \ldots, c_l) + f(d_1, \ldots, d_l). \end{aligned}$$

Dans les identités qui caractérisent les applications linéaires $E \to E'$, les additions et les multiplications scalaires dans les seconds membres ont lieu dans l'espace vectoriel E'. La notion d'application linéaire n'a besoin d'aucune relation particulière entre les espaces E et E' (sauf qu'ils utilisent le même corps de scalaires K), mais il est très bien possible que les deux espaces soient en fait le même. Ce cas

elle-même plutôt que le vecteur qu'elle désigne ; notamment une "combinaison linéaire non triviale" est une telle expression avec au moins un des λ_i non nul, ce qui ne veut pas dire qu'elle désigne un vecteur non nul.

particulier $E = E'$ est d'un intérêt particulier, et on appelle alors f un *endomorphisme* du K-espace vectoriel E. L'étude des endomorphismes de E est un sujet majeur de ce cours. Par rapport aux applications linéaires en général, le cas des endomorphismes est relativement simple dans la mesure où il n'y a qu'un seul type de vecteurs à considérer, mais on verra qu'il est aussi plus riche, à cause des relations qui sont possibles entre les vecteurs avant et après l'application de f.

1.2. *Familles génératrices d'un sous-espace, liées ou libres, bases.*

On appellera $[v_1, \dots, v_l]$ une famille génératrice du sous-espace S si on a $\mathrm{Vect}(v_1, \dots, v_l) = S$ (et en particulier c'est une famille génératrice de l'espace E tout entier si $\mathrm{Vect}(v_1, \dots, v_l) = E$: tout vecteur s'écrit comme combinaison linéaire de vecteurs de la famille). (Une famille de vecteurs se distingue d'un ensemble de vecteurs par le fait que ses membres ont une place dans la famille, de sorte que $[x, y, z]$ ne soit pas la même famille que $[y, z, x]$, et par le fait que plusieurs membres peuvent être égaux. La terminologie est traditionnelle mais n'est pas de grande importance ; on pourrait lire "liste" au lieu de "famille" si elle est finie, ce qui sera toujours le dans ce cours. Aussi, on a choisi de noter les familles de vecteurs entre crochets, pour éviter la confusion possible avec les éléments de K^n notés avec parenthèses.)

La description $\mathrm{Vect}(v_1, \dots, v_l)$ d'un sous-espace par générateurs n'est pas unique, et on a même beaucoup de choix pour les générateurs v_i (par exemple, même dans le cas $\mathrm{Vect}(v)$ d'un seul générateur, le sous-espace est formé de tous les multiples scalaires de v, et le sous-espace $\mathrm{Vect}(\lambda v)$ engendré par un tel multiple (non nul) sera le même que $\mathrm{Vect}(v)$). On ne cherchera donc pas *la meilleure* description d'un sous-espace par générateurs. Mais dans une description donnée, on peut au moins se demander si l'un des générateurs n'est pas redondant, c'est-à-dire que l'espace engendré ne changerait pas si l'on enlevait ce générateur. On voit facilement qu'un générateur est redondant dans ce sens si (et seulement si) il s'écrit comme combinaison linéaire des *autres* générateurs, et si c'est la cas pour l'un (au moins) des v_i, on dira que la famille $[v_1, \dots, v_l]$ de vecteurs est *liée*. Dans ce cas on peut prendre cette combinaison linéaire donnant v_i (mais dans quelle expression v_i lui-même n'intervient pas), et y rajouter un terme "$-v_i$", pour trouver une combinaison linéaire de $v_1, \dots, v_l$ dont la valeur est $\vec{0}$, mais quelle combinaison linéaire est non-triviale (car au moins le coefficient $\lambda_i = -1$ de v_i n'est pas nul). On pourra alors donner une caractérisation alternative des familles liées : une famille de vecteurs est liée si on peut former une combinaison linéaire non-triviale de ses vecteurs dont la valeur est $\vec{0}$.

C'est le cas contraire qui est plus intéressant, c'est-à-dire où aucun des vecteurs $v_1, \dots, v_l$ ne s'exprime comme combinaison linéaire des autres vecteurs de la famille ; dans ce cas on dit que $[v_1, \dots, v_l]$ est une *famille libre*. Une description $S = \mathrm{Vect}(v_1, \dots, v_l)$ d'un sous-espace par générateurs est donc non redondante si $[v_1, \dots, v_l]$ est une famille libre de vecteurs. On peut caractériser les familles libres $[v_1, \dots, v_l]$ par la propriété que *la seule* combinaison linéaire $\lambda_1 v_1 + \dots + \lambda_l v_l$ dont la valeur est $\vec{0}$ est la combinaison triviale, avec $\lambda_1 = \dots = \lambda_l = 0$. Par conséquent, un vecteur ne peut s'écrire que d'une façon au plus comme combinaison linéaire de vecteurs d'une famille libre donnée : s'il y avait deux écritures distinctes, leur soustraction produirait une combinaison non triviale de valeur $\vec{0}$, mais celle-ci n'existe pas.

Une *base* d'un sous-espace S est une famille libre et génératrice de S. C'est donc précisément une famille $[v_1, \dots, v_l]$ telle que $S = \mathrm{Vect}(v_1, \dots, v_l)$ soit une description non redondante de S par générateurs. En particulier, une base de E est une famille libre $[v_1, \dots, v_l]$ de vecteurs telle que tout vecteur $x \in E$ s'écrive comme combinaison linéaire $x = \lambda_1 v_1 + \dots + \lambda_l v_l$ de vecteurs de la famille ; comme on vient d'observer pour une famille libre, cette écriture sera unique. Les scalaires λ_i qui figurent dans cette écriture s'appellent les *coordonnées* de x dans la base $[v_1, \dots, v_l]$.

Un résultat fondamental est que, pour tout sous-espace S qui admet une famille finie de générateurs, toute description non redondante de S par générateurs (pour laquelle le choix est en général vaste) fait intervenir *le même nombre* de générateurs, quel nombre est appelé la *dimension* de S et noté $\dim(S)$. Une telle famille de générateurs est libre, et donc précisément une base de S ; le résultat dit donc que toutes les bases de S ont précisément $\dim(S)$ vecteurs. (Il faut savoir qu'il existe des espaces vectoriels qui n'admettent aucune famille finie de générateurs, et en particulier pas de base finie ; ceci arrive notamment pour des espaces définis de façon très générale, comme ceux de toutes les fonctions différentiables $\mathbf{C} \to \mathbf{C}$, de toutes le suites infinies de scalaires, ou encore de tous les polynômes en X (à coefficients dans K).

Ces espaces, dits de dimension infinie, ont des propriétés bien plus compliquées que ceux de dimension finie, et ils ne seront pas étudiés dans ce cours. Mais ils ont néanmoins leur utilité comme source de sous-espaces de dimension finie, comme par exemple ceux formés des solutions de certaines équations.)

Une observation pratique est que, si on sait que une famille $[v_1, \dots, v_i]$ est libre mais qu'en rajoutant le vecteur v_{i+1} on obtient une famille liée $[v_1, \dots, v_{i+1}]$, alors on a $\mathrm{Vect}(v_1, \dots, v_i) = \mathrm{Vect}(v_1, \dots, v_{i+1})$, c'est-à-dire le dernier générateur v_{i+1} est redondant par rapport aux vecteurs précédents. (Il existe une combinaison linéaire non triviale de $v_1, \dots, v_{i+1}$ qui vaut $\vec{0}$, et dans celle-ci le coefficient de v_{i+1} ne peut pas être nul car $[v_1, \dots, v_i]$ est libre ; en divisant par ce coefficient et en isolant le terme v_{i+1} on voit que $v_{i+1} \in \mathrm{Vect}(v_1, \dots, v_i)$.) Ainsi on a une méthode systématique pour extraire d'une famille quelconque de générateurs $[v_1, \dots, v_l]$ de S une base de S : en commençant avec la famille vide (qui est libre), on considère dans l'ordre les vecteurs $v_1, \dots, v_l$; si avec les vecteurs retenus précédemment il forme une famille libre on le retient comme élément de la base à sélectionner, et sinon (il est redondant par rapport aux vecteurs précédents et) on ne le sélectionne pas. Par ce procédé on montre le résultat suivant, qui est une formulation modeste (car en dimension finie) du théorème dit "de la base incomplète".

1.2.1. Théorème. *Soit S un sous-espace d'un K-espace vectoriel E. Si $L = [v_1, \dots, v_l]$ est une famille libre de vecteurs de S, et $G = [w_1, \dots, w_g]$ une famille génératrice de S (donc $S = \mathrm{Vect}(w_1, \dots, w_g)$), alors il existe une base de S de la forme $B = [v_1, \dots, v_l, w_{i_1}, \dots, w_{i_k}]$ avec $k \geq 0$ et $1 \leq i_1 < \dots < i_k \leq g$, c'est-à-dire formée à partir de L tout entier en complétant par une partie de G.*

Pour le prouver, il suffit de commencer avec la famille $L \cup G = [v_1, \dots, v_l, w_1, \dots, w_g]$, qui est certainement génératrice de S car sa partie G l'est déjà, et d'y appliquer le procédé ci-dessus de sélection d'une base de S ; comme sa partie initiale L est libre, les vecteurs de L seront tous retenus pour la base.

Ce résultat montre clairement (ce qui d'ailleurs est évident par d'autre considérations aussi) que le nombre l d'éléments d'une famille libre L dans S vérifie toujours $l \leq \dim(S)$ et que le nombre g d'éléments d'une famille génératrice G de S vérifie toujours $g \geq \dim(S)$ (car on pourra prendre $L = \emptyset$).

1.3. *Expression dans une base, matrices d'applications linéaires.*

On a déjà observé que, étant donnée une base $\mathcal{B} = [b_1, \dots, b_n]$ d'un espace vectoriel E, tout vecteur $v \in E$ peut être écrit de façon unique comme combinaison linéaire $v = x_1 b_1 + \dots + x_n b_n$ des vecteurs de cette base. Il sera pratique de noter

$$(x_1, \dots, x_n)_{\mathcal{B}} = x_1 b_1 + \dots + x_n b_n$$

ce vecteur. L'application linéaire $K^n \to E$ qui envoie $(x_1, \dots, x_n) \mapsto (x_1, \dots, x_n)_{\mathcal{B}}$ est donc bijective (surjective à cause de l'existence d'une écriture, injective à cause de son unicité). Il existe donc une application réciproque $E \to K^n$, qui à $v \in E$ associe le n-uplet $(\lambda_1, \dots, \lambda_n)$ de ses coordonnées dans la base $\mathcal{B}$. Cette réciproque est linéaire, comme c'est toujours le cas de la réciproque d'une application linéaire bijective. Les applications linéaires et bijectives sont aussi appelées *isomorphismes* de K-espaces vectoriels. Leur importance vient du fait que, à l'aide d'un isomorphisme $E \to E'$ et l'isomorphisme réciproque $E' \to E$, toute question d'algèbre linéaire concernant l'espace E peut être traduit en une question équivalente concernant E', et vice versa. Tout K-espace E de dimension n possède (au moins) une base de n éléments, et donc un isomorphisme $E \to K^n$, ce qui explique l'importance des espaces K^n comme exemples de K-espaces de dimension finie.

Dans le cas d'un espace vectoriel K^n, les n scalaires qui constituent les vecteurs peuvent aussi être interprétés comme leur coefficients dans une base particulière $\mathcal{E} = [e_1, \dots, e_n]$, dite canonique, de K^n. Cela veut dire que l'isomorphisme $K^n \to K^n$ d'expression en coordonnées sur cette base est l'identité : on a $(x_1, \dots, x_n) = x_1 e_1 + \dots + x_n e_n = (x_1, \dots, x_n)_{\mathcal{E}}$, quels que soient $x_1, \dots, x_n \in K$. Pour connaître e_i, il suffit de prendre $x_i = 1$ et tous les autres scalaires nuls, car on aura dans ce cas $x_1 e_1 + \dots + x_n e_n = e_i$; on a donc $e_i = (0, \dots, 0, 1, 0, \dots, 0)$ où la coordonnée 1 se trouve à la i-ème place.

Les espaces de la forme K^n sont donc munis d'une base préférée, ce qui n'est pas le cas des espaces vectoriels en général. Mais la possibilité de choisir une base adaptée à une situation particulière (par exemple la donnée d'un endomorphisme de E) sera un outil important, et pour cela il est préférable de ne pas fixer une base une fois pour toutes (ou d'étudier seulement les espaces K^n, ce qui revient au même).

1.3 Expression dans une base, matrices d'applications linéaires

Une application linéaire $f : E \to E'$ est souvent initialement donnée d'une façon qui utilise la nature précise des éléments de E et de E' ; à titre d'exemple, pour $K = \mathbf{R}$ et avec E égal l'espace de polynômes réels quadratiques en X, et $E' = \mathbf{R}$ (un espace vectoriel de dimension 1), on pourra considérer l'application linéaire $f : E \to \mathbf{R}$ qui consiste à évaluer les polynômes en le nombre $\sqrt{2}$, c'est-à-dire $f : P(X) \mapsto P(\sqrt{2})$. Mais pour pouvoir appliquer les méthodes de l'algèbre linéaire, on a besoin d'un type de description qui soit indépendante de la nature des espaces E, E'. Pour cela on pourra se servir de bases de ces espaces (qu'on suppose de dimension finie). Une telle description utilise la propriété fondamentale suivante.

1.3.1. Proposition. *Si $\mathcal{B} = [b_1, \dots, b_m]$ est une base de E, alors une application linéaire $f : E \to E'$ est entièrement déterminée par la donnée de ses valeurs dans les vecteurs de la base $\mathcal{B}$, c'est-à-dire par les éléments $f(b_1), \dots, f(b_m)$ de E'. Réciproquement, si l'on donne un m-uplet quelconque de vecteurs $v_1, \dots, v_m$ de E', il existe une application linéaire $f : E \to E'$ telle que $f(b_i) = v_i$ pour $i = 1, \dots, n$.*

Pour prouver cette proposition, il suffit d'utiliser la compatibilité d'une application linéaire avec les combinaisons linéaires : l'égalité $f(\lambda_1 b_1 + \cdots + \lambda_m b_m) = \lambda_1 f(b_1) + \cdots + \lambda_m f(b_m)$ donne la valeur de $f(v)$ d'un vecteur quelconque $v = (\lambda_1, \dots, \lambda_n)_{\mathcal{B}} \in E$. En plus, si on se donne des vecteurs $v_1, \dots, v_m \in E'$, l'application qui a $(\lambda_1, \dots, \lambda_n)_{\mathcal{B}}$ associe $\lambda_1 v_1 + \cdots + \lambda_m v_m$ est toujours linéaire.

Pour une description explicite d'une application linéaire $f : E \to E'$ en termes de scalaires seulement, on aura, en plus de la base $\mathcal{B}$ de E, aussi besoin d'une base $\mathcal{B}' = [b'_1, \dots, b'_n]$ de E', pour exprimer les vecteurs $f(b_1), \dots, f(b_m)$ en coordonnées dans cette base.

1.3.2. Définition. *La matrice d'une application linéaire $f : E \to E'$, par rapport à un couple de bases $\mathcal{B} = [b_1, \dots, b_m]$ de E et $\mathcal{B}' = [b'_1, \dots, b'_n]$, est la matrice $n \times m$*

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,m} \\ a_{2,1} & a_{2,2} & \dots & a_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \dots & a_{n,m} \end{pmatrix},$$

dont la colonne j contient les coordonnées dans la base $\mathcal{B}'$ de l'image par f du j-ème vecteur de base de $\mathcal{B}$, c'est-à-dire $f(b_j) = (a_{1,j}, \dots, a_{m,j})_{\mathcal{B}'}$, pour $j = 1, \dots, m$.

On fera attention au fait qu'en donnant la taille d'une matrice, on mentionne le nombre de lignes avant le nombre de colonnes, et que ce nombre de lignes est la dimension de l'espace *d'arrivée*.

La définition décrit la matrice d'une application linéaire f donnée, mais la pratique est également importante pourvoir décrire réciproquement l'application f à partir de sa matrice A (toujours par rapport à des bases de E et E' fixées). On trouve (en utilisant la linéarité de f) que

$$f\big((x_1, \dots, x_m)_{\mathcal{B}}\big) = (y_1, \dots, y_n)_{\mathcal{B}'} \qquad \text{où } y_i = a_{i,1}x_1 + \cdots + a_{i,m}x_m \text{ pour } i = 1, \dots, n. \tag{4}$$

L'opération de cette application f peut être réalisée en trois étapes : d'abord on exprime le vecteur $v \in E$ auquel f est appliqué en coordonnées (l'isomorphisme $E \to K^m$ déterminé par la base $\mathcal{B}$), puis on transforme le m-uplet de coordonnées $(x_1, \dots, x_m)$ en le n-uplet de coordonnées $(y_1, \dots, y_n)$ à l'aide de la matrice A, et on transforme finalement ce n-uplet en un vecteur de E' par combinaison linéaire des vecteurs de la base $\mathcal{B}'$ (l'isomorphisme $K^n \to E'$ déterminé par la base $\mathcal{B}'$). Seulement l'étape du milieu dépend de A, et elle ne dépend de rien d'autre. En effet, cette application linéaire $L_A : K^m \to K^n$ est celle dont A est la matrice *par rapport aux bases canoniques* de K^m et de K^n. La situation est ainsi :

$$\begin{array}{ccc} v \in E & \xrightarrow{f} & f(v) \in E' \\ \updownarrow \mathcal{B} & & \updownarrow \mathcal{B}' \\ (x_1, \dots, x_m) \in K^m & \xrightarrow{L_A} & (y_1, \dots, y_n) \in K^n \end{array}$$

L'application L_A est déterminée par (4) ; en écrivant les listes d'éléments de K verticalement on a

$$L_A : \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \mapsto \begin{pmatrix} a_{1,1}x_1 + \cdots + a_{1,m}x_m \\ a_{2,1}x_1 + \cdots + a_{2,m}x_m \\ \vdots \\ a_{n,1}x_1 + \cdots + a_{n,m}x_m \end{pmatrix}. \tag{5}$$

Cette opération est prise comme définition de la multiplication à gauche d'une "colonne" par A :

$$A \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = \begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ a_{2,1} & \dots & a_{2,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,m} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} \overset{\text{déf}}{=} \begin{pmatrix} a_{1,1}x_1 + \dots + a_{1,m}x_m \\ a_{2,1}x_1 + \dots + a_{2,m}x_m \\ \vdots \\ a_{n,1}x_1 + \dots + a_{n,m}x_m \end{pmatrix}. \tag{6}$$

La raison de la présenter L_A comme opérant sur des matrices à une colonne par la multiplication à gauche par A, est que ce produit est un cas particulier du produit matriciel, qui est définie de façon à correspondre à la composition d'applications linéaire. Concrètement, supposons qu'en plus de cette matrice A de taille $n \times m$ on ait une autre matrice B de taille $m \times l$ pour un certain l, qui détermine une application linéaire $L_B : K^l \to K^m$. On peut alors former la composée $L_A \circ L_B : K^l \to K^n$ (dans une composée c'est toujours l'application écrite à droite, ici L_B, qui agit en premier : $(L_A \circ L_B)(x) = L_A(L_B(x))$). La composition d'applications linéaires donne toujours une application linéaire, et on peut donc exprimer $L_A \circ L_B$ par une matrice $n \times l$ par rapport aux bases canoniques de K^l et de K^n, quelle matrice sera par définition le produit matriciel $A \cdot B$, autrement dit on aura $L_{A \cdot B} = L_A \circ L_B$.

Pour déterminer les coefficients de ce produit $A \cdot B$, on utilise la définition 1.3.2 : ses colonnes sont formées par les images des vecteurs de la base de départ, exprimés dans la base d'arrivée. Comme les bases considérées ici sont toutes canoniques, l'expression sur la base d'arrivée est une opération sans effet, et on peut simplement dire que la k-ème colonne d'une matrice M *est égale à* l'image $L_M(e_k)$ du k-ème vecteur de la base canonique de l'espace de départ. En particulier k-ème colonne de $A \cdot B$ est égale à l'image $L_{A \cdot B}(e_k) = (L_A \circ L_B)(e_k) = L_A(L_B(e_k))$ du vecteur e_k de la base canonique de K^l. Or, le vecteur $L_B(e_k) \in K^m$ est égal à la k-ème colonne de la matrice B. Il suffit donc d'utiliser (5) pour L_A appliqué à $L_B(e_k)$, c'est-à-dire en prenant pour les coefficients x_j les coefficients $b_{j,k}$ de la k-ème colonne de la matrice B, pour $j = 1, \dots, m$. On obtient la définition du produit matriciel :

$$\begin{pmatrix} a_{1,1} & \dots & a_{1,m} \\ a_{2,1} & \dots & a_{2,m} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \dots & a_{n,m} \end{pmatrix} \cdot \begin{pmatrix} b_{1,1} & b_{1,2} & b_{1,3} & \dots & b_{1,l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{m,1} & c_{m,2} & c_{m,3} & \dots & c_{m,l} \end{pmatrix} \overset{\text{déf}}{=} \begin{pmatrix} c_{1,1} & c_{1,2} & c_{1,3} & \dots & c_{1,l} \\ c_{2,1} & c_{2,2} & c_{2,3} & \dots & c_{2,l} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n,1} & c_{n,2} & c_{n,3} & \dots & c_{n,l} \end{pmatrix}, \tag{7}$$

où les coefficients du produit sont donnés par $c_{i,k} = a_{i,1}b_{1,k} + \dots + a_{i,m}b_{m,k} = \sum_{j=1}^{m} a_{i,j}b_{j,k}$ pour $i = 1, 2, \dots, n$ et $k = 1, 2, 3, \dots, l$. Le produit matriciel peut également être utilisé pour décrire la composition d'applications linéaires dont les matrices sont données par rapport à des bases autres que les bases canoniques d'espaces de la forme K^n. Pour cela il est essentiel que pour l'expression des matrices on se serve dans chacun des espaces concernés toujours d'une même base. Les questions liées à l'utilisation de bases différentes dans un même espace seront abordées dans la section suivante.

La matrice de l'identité $\mathrm{id}_E : E \to E$, qui vérifie $\mathrm{id}_E(x) = x$ pour tout $x \in E$, par rapport à une base $\mathcal{B} = (b_!, \dots, b_n)$ de E utilisée au départ comma à l'arrivée, est toujours de la même forme. En effet sa colonne j contient les coordonnées de $\mathrm{id}_E(b_j) = b_j$ dans la base $\mathcal{B}$, qui sont tous nuls sauf celui à la position j et qui est 1. On trouve donc la matrice $n \times n$ avec des coefficients 1 partout sur la diagonale principale, et des coefficients 0 partout ailleurs ; on l'appelle la matrice identité id_n.

Si $f : E \to E'$ est un isomorphisme d'espaces vectoriels (ce qui nécessite $\dim(E) = \dim(E')$) et $g : E' \to E$ l'isomorphisme réciproque, on a $g \circ f = \mathrm{id}_E$ et $f \circ g = \mathrm{id}_{E'}$. Si $\mathcal{B}$ est une base de E et $\mathcal{B}'$ une base de $\mathcal{B}'$, les matrices P de f par rapport à $\mathcal{B}, \mathcal{B}'$ et Q de g par rapport à $\mathcal{B}', \mathcal{B}$ vérifieront donc $Q \cdot P = \mathrm{id}_n$ et $P \cdot Q = \mathrm{id}_n$. On appelle ces matrices inverses l'une de l'autre, et qu'on note $Q = P^{-1}$ ou $P = Q^{-1}$. Une matrice qui possède un inverse est dit inversible. Pour qu'une matrice soit inversible il est nécessaire qu'elle soit carrée, mais ce n'est pas suffisant. Mais il est utile de savoir que si les matrices P et Q sont carrées et $P \cdot Q = \mathrm{id}_n$, alors on aura automatiquement $Q \cdot P = \mathrm{id}_n$, et donc $Q = P^{-1}$.

1.4. *Changement de base, classification d'applications linéaires par le rang.*

Quand on utilise différentes bases dans un même espace, des questions se posent concernant les conversions nécessaires entre les descriptions en coordonnées de vecteurs et d'applications linéaires. Bien

que ces conversions aient toujours un caractère linéaire (la conversion d'une somme de vecteurs est la somme de leurs conversions, etc.), et fassent intervenir des matrices, elles ne correspondent pas à des applications linéaires entre des espaces vectoriels dans le sens considéré auparavant, ou plus précisément elles correspondent à l'identité $\mathrm{id}_E : E \to E$ (car leur but n'est pas de changer les vecteurs, juste leurs descriptions). La comparaison avec des conversions de quantités entre différentes unités de mesure est pertinente : la valeur reste la même, mais la représentation numérique change. En effet, une telle conversion peut être vue comme un exemple d'un changement de base, dans un espace de dimension 1. À titre d'exemple, l'ensemble de longueurs dans la réalité forme un $\mathbf{R}$-espace vectoriel (l'addition de longueurs et la multiplication scalaire étant définies), et le choix d'une unité de mesure comme le centimètre y fixe une base (d'un seul élément) ; la conversion d'une longueur exprimé en cette unité en une unité différente (le pouce) fera intervenir un facteur numérique de conversion, mais la vraie longueur ne change pas. Dans le cas général d'un changement de base, une matrice inversible remplacera le facteur de conversion.

Considérons donc un espace vectoriel E de dimension n, muni de deux bases $\mathcal{B}, \mathcal{B}'$; pour faciliter le discours on appellera $\mathcal{B}$ la base originale, et $\mathcal{B}'$ la nouvelle base. On peut illustrer la situation par le diagramme suivant.

$$\begin{array}{ccc} E & \overset{\mathrm{id}_E}{\longleftrightarrow} & E \\ \updownarrow{\scriptstyle \mathcal{B}} & & \updownarrow{\scriptstyle \mathcal{B}'} \\ K^n & \underset{d}{\overset{c}{\rightleftarrows}} & K^n \end{array} \tag{8}$$

Les flèches c, d dans la ligne en bas représentent les conversions entre les coordonnées dans les deux bases ; par définition l'effet de suivre une telle flèche est la même que de passer d'abord en haut vers E (interprétation des coordonnées dans la base indiquée), puis après avoir traversé la flèche id_E de repasser en bas (expression du vecteur de E en coordonnées dans l'autre base). Comme elles sont obtenues par la composition d'isomorphismes linéaires, les flèches c et d représentent des isomorphismes l'espace K^N vers lui-même. Elles correspondent donc (directement) à des matrices, dont l'une est l'inverse de l'autre.

Ces matrices sont les matrices de conversion entre les deux bases, dites *matrices de passage* entre ces bases. Une question de convention se pose : quelle des deux matrices faut-il associer au passage de l'ancien base $\mathcal{B}$ à la nouvelle base $\mathcal{B}'$? Du point de vue les applications linéaires $c, d : K^n \to K^n$ associées aux matrices, le choix naturel est de prendre la matrices correspondant à c, car c convertit les coordonnées dans la base $\mathcal{B}$ en les coordonnées du même vecteur dans la base $\mathcal{B}'$. Mais si l'on pense aux bases elles-mêmes, il est le plus naturel d'exprimer les vecteurs de la nouvelle base en coordonnées dans la base originale, et cela donne les colonnes de la matrice correspondante à d : le vecteur $b'_j \in E$ de la nouvelle base $\mathcal{B}'$ correspond par la flèche verticale à droite au vecteur e_j de la base canonique de K^n, et son expression dans la base originale $\mathcal{B}$ (la flèche verticale à gauche) donne donc $d(e_j) \in K^n$, c'est-à-dire la colonne j de la matrice correspondant à d. Dans la terminologie française, c'est la deuxième argumentation qui l'a emporté, et on appelle matrice de passage de $\mathcal{B}$ vers $\mathcal{B}'$ la matrice correspondant à la flèche d. Cela est regrettable en vue de fait que dans le diagramme cette flèche va plutôt de $\mathcal{B}'$ vers $\mathcal{B}$. Ceci dit, la convention correspond à celle utilisée dans la vie courante pour le passage du franc (ou d'autres anciennes monnaies) à l'euro, qu'on décrit toujours par le facteur (6,55957) qui exprime la nouvelle base (l'euro) en termes de l'ancienne base (le franc), et non pas par le facteur inverse (0,152449) par lequel il faudrait multiplier un montant en francs pour trouver le montant correspondant en euros.

1.4.1. Définition. *Si $\mathcal{B}, \mathcal{B}'$ sont deux bases d'un espace vectoriel E de dimension n, la matrice de passage de $\mathcal{B}$ vers $\mathcal{B}'$ est la matrice $n \times n$ dont la colonne j contient les coordonnées dans la base $\mathcal{B}$ du j-ème vecteur de la base $\mathcal{B}'$. En multipliant (une colonne contenant) les coordonnées d'un vecteur $x \in E$ dans la base $\mathcal{B}'$ à gauche par cette matrice, on obtient les coordonnées de x dans la base $\mathcal{B}$.*

Une fois qu'il est clair comment il faut convertir des coordonnées d'un vecteur pour l'exprimer dans une autre base, l'effet d'un changement de base sur des matrices exprimées par rapport à ces bases est facile à déduire. Par exemple, supposons qu'on connaît la matrice A d'une application $f : E \to E'$ par rapport aux bases $\mathcal{B}$ de E et $\mathcal{B}'$ de E', et que $\mathcal{C}$ est une autre base de E, telle que la matrice (inversible) P est la matrice de passage de $\mathcal{B}$ à $\mathcal{C}$. Alors si l'on cherche la matrice B de f par rapport aux bases $\mathcal{C}$ et $\mathcal{B}'$, on peut remarquer que cette matrice doit opérer sur une colonne de coordonnées exprimés dans la base $\mathcal{C}$.

Il convient donc de convertir en coordonnées sur la base $\mathcal{B}$ en multipliant par P, à quel point on peut appliquer la matrice A pour obtenir l'image par f, exprimée dans la base $\mathcal{B}'$; au total on a $B = A \cdot P$. Par un même type de raisonnement, on voit qu'un changement de base à l'arrivée avec matrice de passage Q a pour effet sur la matrice une multiplication à gauche par la matrice Q^{-1} (car cette fois on obtient d'abord les coordonnées sur l'ancienne base de E', qu'il faut convertir en coordonnées sur la nouvelle base).

1.4.2. Proposition. *Si A est la matrice de $f : E \to E'$ par rapport aux bases $\mathcal{B}$ de E et $\mathcal{B}'$ de E', et si $\mathcal{C}$ et $\mathcal{C}'$ sont d'autres bases de E respectivement de E', et si P et Q sont les matrices de passage de $\mathcal{B}$ vers $\mathcal{C}$ respectivement de $\mathcal{B}'$ vers $\mathcal{C}'$, alors la matrice de f rapport aux bases $\mathcal{C}$ et $\mathcal{C}'$ sera $Q^{-1} \cdot A \cdot P$.*

Si une application linéaire est représentée par une matrice A par rapport à un couple de bases (une au départ et une à l'arrivée), et par une matrice B par rapport à un autre couple de bases, on dit que A et B sont des "matrices équivalentes". En vue de la proposition cela veut dire qu'il existe des matrices inversibles P, Q telles que $B = Q^{-1} \cdot A \cdot P$, et on en déduit qu'il s'agit en effet d'une relation d'équivalence. Deux matrices équivalentes sont évidemment de la même taille ; on verra ci-dessous que parmi les matrices d'un taille donnée il n'y a qu'un très petit nombre (et en particulier toujours un nombre fini) de classes d'équivalence de matrices pour cette relation.

On peut associer à chaque application linéaire $f : E \to E'$ deux sous-espaces, l'un dans E' et l'autre dans E, qui donnent des renseignements importants concernant f. Dans E', il s'agit de l'image $\mathrm{Im}(f)$ se f, le sous-espace des vecteurs sont l'image par f d'au moins une vecteur de E ; en formule

$$\mathrm{Im}(f) = \{\, f(v) \mid v \in E \,\}.$$

Dans E il s'agit du noyau $\mathrm{Ker}(f)$ de f, le sous-espace des vecteurs que f envoie sur $\vec{0} \in E'$; en formule

$$\mathrm{Ker}(f) = \{\, v \in E \mid f(v) = \vec{0} \,\}.$$

L'image et le noyau de f ne sont pas liés de façon directe, après tout ce sont des sous-espaces de différents espaces E', E. Mais il existe une relation fondamentale entre leurs *dimensions* et celle de E.

1.4.3. Théorème du rang. *Soit E un espace vectoriel de dimension finie, et $f : E \to E'$ linéaire, alors* $\dim(\mathrm{Im}(f)) + \dim(\mathrm{Ker}(f)) = \dim(E)$. *Le nombre* $\dim(\mathrm{Im}(f))$ *est appelé le rang* $\mathrm{rg}(f)$ *de f.*

Preuve. Soit $l = \dim(\mathrm{Ker}(f))$ et $n = \dim(E)$. On choisit une base $L = [b_1, \ldots, b_l]$ du sous-espace $\mathrm{Ker}(f)$ de E. Le théorème de la base incomplète permet de compléter la famille libre L de vecteurs de E à une base $\mathcal{B} = [b_1, \ldots, b_n]$ de E (on pourra prendre n'importe quelle base de E pour la famille génératrice G dans laquelle le théorème sélectionne les vecteurs supplémentaires $b_{l+1}, \ldots, b_n$). Tout vecteur de $\mathrm{Im}(f)$ est combinaison linéaire des vecteurs $f(b_1), \ldots, f(b_n)$, et comme par construction $f(b_i) = \vec{0}$ pour tout $i \leq l$, on a en fait $\mathrm{Im}(f) = \mathrm{Vect}(f(b_{l+1}), \ldots, f(b_n))$. Montrons que les vecteurs $f(b_{l+1}), \ldots, f(b_n)$ forment une famille libre dans E', et donc une base de $\mathrm{Im}(f)$; cela entraînera $\dim(\mathrm{Im}(f)) = n - l$ et le théorème. Supposons que $\vec{0} = \lambda_{l+1} f(b_{l+1}) + \cdots + \lambda_n f(b_n) = f(\lambda_{l+1} b_{l+1} + \cdots + \lambda_n b_n)$, alors on a $\lambda_{l+1} b_{l+1} + \cdots + \lambda_n b_n \in \mathrm{Ker}(f) = \mathrm{Vect}(b_1, \ldots, b_l))$. Mais une combinaison linéaire d'une partie $b_{l+1}, \ldots, b_n$ de la base $\mathcal{B}$ ne peut être aussi une combinaison linéaire de la partie complémentaire $b_1, \ldots, b_l$ de $\mathcal{B}$, que si les deux combinaisons sont triviales, donc en particulier $(\lambda_{l+1}, \ldots, \lambda_n) = (0, \ldots, 0)$. □

Si A est la matrice de f par rapport aux bases $\mathcal{B}$ de E et $\mathcal{B}'$ de E', alors $\mathrm{rg}(f) = \mathrm{rg}(L_A)$: la seconde flèche verticale dans le diagramme

$$\begin{array}{ccc} E & \xrightarrow{f} & E' \\ \uparrow{\scriptstyle \mathcal{B}} & & \uparrow{\scriptstyle \mathcal{B}'} \\ K^m & \xrightarrow{L_A} & K^n \end{array}$$

est un isomorphisme qui fait correspondre $\mathrm{Im}(f)$ à $\mathrm{Im}(L_A)$. Or $\mathrm{Im}(L_A)$ ne dépend que de la matrice A, ce qui permet de définir $\mathrm{rg}(A) = \dim(\mathrm{Im}(L_A)) = \mathrm{rg}(f)$. Ce rang est égal au nombre d'une famille maximale *libre* de colonnes de A (car c'est une base de $\mathrm{Im}(L_A) \subseteq K^n$). Le rang de deux matrices équivalentes est le même : c'est $\mathrm{rg}(f)$ si l'une et l'autre représentent f sur des couple de bases convenable.

1.4.4. Théorème. *Deux matrices sont équivalentes si et seulement si elles ont la même taille et la même rang. Pour $n, m, r \in \mathbf{N}$, il existe des matrices de taille $n \times m$ et de rang r si et seulement si $r \leq \min(n, m)$, dont un exemple est la matrice $M_{n,m,r}$ de taille $n \times m$ dont les seuls coefficients non nuls sont r coefficients* 1*, situés sur les r premières positions de la diagonale principale.*

Preuve. On a déjà vu que pour que deux matrices soient équivalentes il est nécessaire qu'elles aient la même taille $n \times m$ et le même rang r ; or on aura $r \leq \min(n, m)$ car $r = \dim(\mathrm{Im}(L_A))$ ne peut dépasser ni la dimension m de l'espace de départ de L_A, ni celle n de son espace d'arrivée. Pour conclure, il suffit de montrer qu'une telle matrice A est équivalente à la matrice $M_{n,m,r}$ (dont on voit facilement qu'elle a rang r). Soit $f : E \to E'$ une application linéaire dont A est la matrice par rapport à un certain couple de bases ; on choisira un autre couple de bases par rapport auquel la matrice de f est $M_{n,m,r}$. Dans E on choisit une base comme dans la preuve du théorème du rang (dans cette preuve on a $r = n - l$), mais dont on prend les vecteurs dans l'ordre $b_{l+1}, \ldots, b_n, b_1, \ldots, b_l$ (mettant ceux qui engendrent $\mathrm{Ker}(f)$ à la fin, pour que les colonnes nulles soient à droite) ; dans E' on commence avec les images $f(b_{l+1}), \ldots, f(b_n)$ des r premiers vecteurs de la base de E, dont on a vu qu'elles forment une famille libre, et qu'on complète à une base de E' (encore le théorème de la base incomplète). □

1.5. *Endomorphismes, et un nouveau problème de classification.*

Le théorème 1.4.4, et sa démonstration assez directe, montrent comment la liberté d'adapter les bases des espaces vectoriels à une situation donnée permet de simplifier la description de celle-ci (concrètement : la matrice de l'application linéaire). La problématique principale de ce cours est très similaire, mais avec la différence suivante : on ne considère pas une application linéaire $f : E \to E'$ entre des espaces distincts, mais un endomorphisme de E, c'est-à-dire application linéaire $f : E \to E$ vers l'espace E lui-même.

Cette différence a pour conséquence qu'il n'est plus raisonnable de choisir séparément une base au départ et à l'arrivée de f : quand on parle d'une matrice d'un endomorphisme, on suppose (sans mention explicite du contraire) que *la même* base est utilisée au départ et à l'arrivée. Cela est important entre autres par la possibilité d'itérer f pour obtenir ses puissances comme $f^3 : v \mapsto f(f(f(v)))$; si l'on veut que la matrice de f^n soit le produit matriciel A^n de n copies de la matrice A de f, il est essentiel (comme toujours pour un produit matriciel) que la base utilisée à l'arrivée coïncide avec celle au départ.

On est donc amené à définir, pour les matrices carrées uniquement, une relation plus fine que celle d'être des matrices équivalentes : deux matrices carrées A, B sont dites *semblables* si elles peuvent être obtenues comme les matrices d'un même endomorphisme de E, chacune par rapport à une base de E. Si la matrice de passage de la première base vers la seconde est P, on aura donc $B = P^{-1} \cdot A \cdot P$. La relation d'être des matrices semblables est (aussi) une relation d'équivalence, mais a ne pas confondre avec la relation d'être des matrices équivalentes. Pour voir combien ces deux relations sont différentes, il suffit de regarder le cas $A = \mathrm{id}_n$: comme $P^{-1} \cdot \mathrm{id}_n \cdot P = P^{-1} \cdot P = \mathrm{id}_n$ pour toute matrice $n \times n$ inversible P, la matrice id_n n'est semblable à *aucune* autre matrice, bien qu'elle soit équivalente (d'après le théorème 1.4.4) à toute matrice $n \times n$ de rang n, c'est-à-dire inversible. Ce cas est extrême (car la plupart des classes de matrices semblables contiennent une multitude de matrices), mais il est indicatif du fait qu'on a à faire avec une relation beaucoup plus fine que celle d'être équivalentes. En fait on n'arrivera pas dans ce cours à décrire complètement la relation d'être semblables. Ceci dit, la polynôme caractéristique d'une matrice (ou d'un endomorphisme), dont la définition est un but important de ce cours, donnera une information assez détaillée, et qui est toujours la même pour deux matrices semblables (on dit que c'est un invariant de similitude, tout comme le rang est un invariant des classes de matrices équivalentes). Dans la plupart des cas, le fait que deux matrices *ne sont pas semblables* pourra être attesté par une différence entre leur polynômes caractéristiques.

Chapitre 2. Vecteurs propres, valeurs propres.

Pour trouver des propriétés caractéristiques d'un endomorphisme ϕ, qui soient indépendantes d'une base particulière et s'appliquent donc à toute la classe de similitude des matrices qui peuvent représenter f, il est possible de tester pour l'existence de vecteurs qui aient une position particulière par rapport à ϕ. Par exemple, il est possible de tester s'il existe des vecteurs non nuls qui soient annulés par ϕ, c'est à dire si $\dim(\operatorname{Ker}\phi) > 0$. Mais cette possibilité ne se produit que très rarement, et ne donne donc peu d'information pour la plupart des endomorphismes (qui sont inversibles). On peut profiter du fait que ϕ est un endomorphisme, en comparant les vecteurs v avec leur propre image $\phi(v)$. La position particulière que v peut avoir est que $\phi(v)$ soit un multiple scalaire de v. On ne s'intéresse pas au cas $v = \vec{0}$ (pourquoi?), et l'existence d'un tel vecteur, ainsi que la valeur du scalaire $\lambda \in K$ tel que $\phi(v) = \lambda v$, donnent alors un renseignement important sur ϕ. Cela nous conduit aux notions de vecteur propre et de valeur propre.

2.1. *Définition et premières propriétés.*

2.1.1. Définition. *Soit E un K-espace vectoriel et $\phi \in \operatorname{End}(E)$. Si $v \in E$ est non-nul et vérifie $\phi(v) = \lambda v$ pour un scalaire $\lambda \in K$, on appelle v un vecteur propre de ϕ et λ une valeur propre de v.*

Comme $v \neq \vec{0}$, il existe au plus une valeur propre λ associé à v, qu'on appelle la valeur propre de v, et v s'appelle un vecteur propre pour λ. Pour une valeur λ donnée, les vecteurs propres, s'ils existent, ne sont pas uniques: les vecteur propres pour λ sont les vecteur non nuls dans le sous-espace $\operatorname{Ker}(\phi - \lambda \operatorname{id}_E)$ de E (on appelle $\lambda \operatorname{id}_E$ l'homothétie de E de facteur λ). Cet espace $\operatorname{Ker}(\phi - \lambda \operatorname{id}_E)$ est appelé l'*espace propre* de λ (qui contient donc le vecteur nul qui lui n'est pas vecteur propre). Il est parfois utile d'admettre la phrase "espace propre pour λ" même si λ n'est peut-être pas une valeur propre; dans ce dernier cas cet "espace propre" $\operatorname{Ker}(\phi - \lambda \operatorname{id}_E)$ est réduit à $\{\vec{0}\}$ et ne contient donc aucun vecteur propre. Les espaces propres sont des exemples d'espaces ϕ-stables, une notion importante dans la suite, définie ainsi.

2.1.2. Définition. *Un sous-espace F de E est ϕ-stable pour $\phi \in \operatorname{End}(E)$ si $\phi(F) \subseteq F$.*

Les sous-espaces stables F sont ceux pour lesquels on peut déduire par restriction de ϕ à F un endomorphisme de F (si F est un sous-espace non ϕ-stable, on peut restreindre ϕ à une application linéaire $F \to E$, mais cela ne donne pas un endomorphisme de F). La restriction de ϕ à l'espace propre de ϕ pour λ est l'homothétie de cet espace de facteur λ. Réciproquement tout sous-espace ϕ-stable F tel que la restriction de λ à F soit une homothétie de facteur λ est contenu dans l'espace propre de ϕ pour λ. Pour un espace de dimension 1, les seuls endomorphismes possibles sont les homothéties, donc un vecteur $v \neq \vec{0}$ est un vecteur propre si et seulement si $\operatorname{Vect}(v)$ est ϕ-stable, et tous les sous-espaces ϕ-stables de E de dimension 1 sont de cette forme. On verra que souvent les espaces propres (s'il y en a) sont eux-mêmes de dimension 1, et dans ce cas ils forment la totalité des sous-espaces ϕ-stables de E de dimension 1. Mais si (une possibilité à l'extrême opposé) ϕ est une homothétie, il n'y aura qu'une seule valeur propre, avec pour espace propre E tout entier, et *tous* les sous-espaces (en particulier ceux de dimension 1) seront ϕ-stables.

Comme un exemple simple de vecteurs et valeurs propres, considérons dans un espace de dimension 2, un endomorphisme ϕ dont la matrice par rapport à une certaine base $\mathcal{B} = [b_1, b_2]$ est $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, autrement dit on a $\phi(b_1) = b_2$ et $\phi(b_2) = b_1$. Alors b_1 et b_2 ne sont visiblement pas des vecteurs propres, mais $b_1 + b_2$ en est un, avec valeur propre 1, car $\phi(b_1 + b_2) = b_2 + b_1 = 1(b_1 + b_2)$. Un autre vecteur propre est $b_1 - b_2$, avec valeur propre -1 cette fois-ci, car $\phi(b_1 - b_2) = b_2 - b_1 = -1(b_1 - b_2)$. On vérifie facilement que les espaces propres pour $\lambda = 1$ et pour $\lambda = -1$ sont respectivement les droites vectorielles $\operatorname{Vect}(b_1 + b_2)$ et $\operatorname{Vect}(b_1 - b_2)$. Or on verra que le nombre de valeurs propres ne peut pas dépasser la dimension de l'espace, donc on a ainsi décrit toutes les valeurs propres de ϕ, et les espaces propres correspondants.

Si on change cet exemple légèrement, en prenant l'endomorphisme ρ dont la matrice est $R = \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right)$, la situation devient différente. Pour qu'un vecteur $xb_1 + yb_2$ soit vecteur propre avec valeur propre λ, il faut avoir $\rho(xb_1 + yb_2) = -yb_1 + xb_2 = \lambda xb_1 + \lambda yb_2$, et donc $\lambda x = -y$ et $\lambda y = x$, dont on déduit $\lambda^2 x = -x$ et $\lambda^2 y = -y$. Comme au moins un de x, y doit être non nul, il est nécessaire que $\lambda^2 = -1$, ce qui est impossible si $K = \mathbf{R}$ (ou si $K = \mathbf{Q}$). Si par contre on considérait un espace vectoriel complexe

($K = \mathbf{C}$), on a les possibilités $\lambda = \mathbf{i}$ et $\lambda = -\mathbf{i}$, qui sont alors effectivement des valeurs propres de ρ, avec pour vecteurs propres (par exemple) $(1, -\mathbf{i})_{\mathcal{B}} = b_1 - \mathbf{i}b_2$ pour $\lambda = \mathbf{i}$, et $(1, \mathbf{i})_{\mathcal{B}} = b_1 + \mathbf{i}b_2$ pour $\lambda = -\mathbf{i}$.

Cet exemple est une première indication que dans la recherche des valeurs propres, on est confronté à des équations ($\lambda^2 = -1$) qui sont polynomiales, et non pas linéaires, malgré leur origine dans un problème d'algèbre linéaire. Cela expliquera que (une partie de) l'algèbre des polynômes sera traitée dans ce cours.

2.2. *Diagonalisation.*

2.2.1. Définition. *Une matrice diagonale est une matrice carrée A dont tous les coefficients hors de la diagonale principale sont nuls : on a $A_{i,j} = 0$ si $i \neq j$. Une matrice diagonalisable est une matrice semblable à une matrice diagonale. Un endomorphisme ϕ d'un K-espace vectoriel E de dimension finie est diagonalisable si sa matrice par rapport à une base de E quelconque est diagonalisable, ce qui veut dire qu'il existe une base de E telle que la matrice de ϕ par rapport à cette base soit diagonale.*

Comme dans la matrice A d'un endomorphisme ϕ par rapport à une base $\mathcal{B} = [b_1, \ldots, b_n]$, la colonne j contient les coordonnées de $\phi(b_j)$ dans la base $\mathcal{B}$, la condition que les coefficients de cette colonne qui ne sont pas sur la diagonale principale soient nuls veut dire que b_j est un vecteur propre de ϕ, pour la valeur propre égale au coefficient $A_{j,j}$ sur la diagonale. Par conséquent, la condition que A soit diagonale est équivalente à celle que $b_1, \ldots, b_n$ sont tous des vecteurs propres de ϕ. On trouve ainsi :

2.2.2. Proposition. *Un endomorphisme ϕ d'un K-espace vectoriel E de dimension finie est diagonalisable si et seulement si E possède une base entièrement constituée de vecteurs propres pour ϕ.* □

Si un endomorphisme ϕ est diagonalisable, sur une base $\mathcal{B} = [b_1, \ldots, b_n]$ de vecteur propres et avec des valeurs propres correspondantes $\lambda_1, \ldots, \lambda_n$ (ce sont les coefficients diagonaux de la matrice de ϕ par rapport à $\mathcal{B}$), alors l'effet d'appliquer ϕ sur un vecteur en termes de ses coordonnées dans $\mathcal{B}$ est de multiplier chaque coordonnée i par la valeur propre correspondante λ_i, c'est-à-dire

$$\phi\big((x_1, \ldots, x_n)_{\mathcal{B}}\big) = (\lambda_1 x_1, \ldots, \lambda_n x_n)_{\mathcal{B}}. \tag{9}$$

Cette description permet de trouver tous les vecteurs propres de ϕ : pour que $v = (x_1, \ldots, x_n)_{\mathcal{B}} \neq \vec{0}$ soit vecteur propre pour une valeur propre λ, il faut que $\lambda_i x_i = \lambda x_i$ pour $i = 1, \ldots, n$, autrement dit les valeurs propres λ_i à toutes les positions i où $x_i \neq 0$ (ce qui est le cas pour au moins un i) doivent être égales à λ. En particulier λ se trouve parmi les valeurs $\lambda_1, \ldots, \lambda_n$: une matrice diagonale qui représente ϕ met en évidence *toutes* les valeurs propres de ϕ. On voit qu'un vecteur dont une seule coordonnée est non nulle est toujours un vecteur propre (on obtient ainsi tous les multiples non nuls des vecteurs de la base $\mathcal{B}$). Mais il peut y en avoir d'autres, si certaines parmi les valeurs $\lambda_1, \ldots, \lambda_n$ sont égales. Pour une valeur propre λ, sa multiplicité parmi les coefficients diagonaux $\lambda_1, \ldots, \lambda_n$ donne la dimension de l'espace propre pour λ: un vecteur de cet espace peut avoir des coordonnées quelconques dans toutes les positions i avec $\lambda_i = \lambda$ (et doit avoir des coordonnées nulles dans les autres positions).

Remarque. On peut déduire de ce qui précède qu'une famille F de vecteurs propres de ϕ pour des *valeurs propres distinctes* est toujours libre. En effet, comme n'importe quelle famille de vecteurs, F est génératrice d'un certain sous-espace S, et il existe (d'après le théorème 1.2.1) une partie F' de F qui soit une *base* de S. Alors S est clairement ϕ-stable, et l'endomorphisme de S donné par restriction de ϕ est diagonalisable sur la base F'. Mais cet endomorphisme n'a donc pas d'autres valeurs propres que celles des membres de F', pendant qu'un membre du complément $F \setminus F'$ serait un vecteur propre de ϕ dans S pour une telle autre valeur propre. Cette contradiction montre que $F \setminus F' = \emptyset$, donc $F = F'$ est libre.

Une matrice diagonalisable est similaire à une matrice diagonale, mais celle-ci n'est pas (en général) unique. Ceci dit, les choix pour une telle matrice sont limités : deux matrices diagonales qui sont semblables ont le même ensemble de coefficients diagonaux (car c'est l'ensemble des valeurs propres de l'endomorphisme) et pour chaque valeur propre λ, sa multiplicité parmi ces coefficients diagonaux est le même dans les deux matrices (car c'est la dimension de l'espace propre pour λ). Par conséquent, deux matrices diagonales sont semblables seulement si l'une peut être obtenue de l'autre par une permutation

de ses coefficients diagonaux, et le changement de base qui effectue la même permutation des vecteurs de la base montre que deux telles matrices sont effectivement semblables.

Trouver une base des diagonalisation d'un endomorphisme ϕ, si elle existe, est notamment très utile pour décrire les puissances ϕ^n de l'endomorphisme. On déduit facilement de (9) que

$$\phi^k\big((x_1,\dots,x_n)_{\mathcal{B}}\big) = (\lambda_1^k x_1,\dots,\lambda_n^k x_n)_{\mathcal{B}} \qquad \text{pour tout } k\in\mathbf{N}, \tag{10}$$

autrement dit, la puissance D^k d'une matrice diagonale D est la matrice diagonale obtenue de D en remplaçant chaque coefficient diagonal par sa puissance k-ème. Pour une matrice générale, l'expression pour sa puissance k-ème qu'on peut déduire de celle du produit matriciel est *beaucoup* plus compliquée.

Considérons un exemple concret. Dans un $\mathbf{Q}$-espace E de dimension 2, muni d'une base $\mathcal{B}=[b_1,b_2]$, on considère un endomorphisme ϕ aux valeurs propres -3 et 2, mais avec vecteurs propres correspondants $v_1=(1,1)_{\mathcal{B}}=b_1+b_2$ et $v_2=(2,1)_{\mathcal{B}}=b_2$. Le matrice de passage de la base $\mathcal{B}$ à la base $[v_1,v_2]$ est donc $P=\left(\begin{smallmatrix}1&2\\1&1\end{smallmatrix}\right)$, dont l'inverse est $P^{-1}=\left(\begin{smallmatrix}-1&2\\1&-1\end{smallmatrix}\right)$, et l'endomorphisme ϕ, dont la matrice par rapport à la base $[v_1,v_2]$ est par construction $\mathrm{Mat}_{[v_1,v_2]}(\phi)=D=\left(\begin{smallmatrix}-3&0\\0&2\end{smallmatrix}\right)$, aura comme matrice A par rapport à la base $\mathcal{B}$:

$$A=\mathrm{Mat}_{\mathcal{B}}(\phi)=P\cdot D\cdot P^{-1}=\begin{pmatrix}1&2\\1&1\end{pmatrix}\cdot\begin{pmatrix}-3&0\\0&2\end{pmatrix}\cdot\begin{pmatrix}-1&2\\1&-1\end{pmatrix}=\begin{pmatrix}7&-10\\5&-8\end{pmatrix}.$$

[La transformation de la matrice par rapport à la nouvelle base $[v_1,v_2]$ vers la matrice par rapport à l'ancienne base $\mathcal{B}$ est l'opposée que celle décrite dans la proposition 1.4.2, d'où l'inverse de P est utilisé *à droite*. Pour éviter la recherche (pas encore abordée dans ce cours) des valeurs propres et des vecteurs propres d'une matrice donnée, on les à simplement imposés au début de notre construction, ce qui est bien sûr une démarche dans une direction pas très réaliste. Le lecteur prendra néanmoins soin de vérifier qu'on a effectivement obtenu que $\phi(v_1)=-3v_1$ et que $\phi(v_2)=2v_2$.] Par conséquent on a pour $n\in\mathbf{N}$:

$$\mathrm{Mat}_{[v_1,v_2]}(\phi^n)=D^n=\begin{pmatrix}(-3)^n&0\\0&2^n\end{pmatrix}$$

et donc

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

On peut vérifier cette formule facilement par récurrence, mais elle se laisse difficilement deviner à partir de quelques valeurs explicites de A^n : pour la trouver il était nécessaire de connaître les valeurs propres et les vecteurs propres de A. Par contre, ce qu'on peut constater facilement par le calcul explicite des puissances A^n, est que ses colonnes tendent vers des multiples du vecteur propre $v_1=(1,1)_{\mathcal{B}}=b_1+b_2$. On a par exemple $A^{15}=\left(\begin{smallmatrix}14414443&-28763350\\14381675&-28730582\end{smallmatrix}\right)$; la proximité de ses colonnes à $-(-3)^{15}v_1$ respectivement à $2\times(-3)^{15}v_1$ s'explique par le fait que le facteur $(-3)^{15}=-14\,348\,907$ est beaucoup plus grand en valeur absolue que $2^{15}=32\,768$. Il est un phénomène général que, si A est une matrice diagonalisable sur le nombres réels, alors pour presque tous les vecteurs v l'image itérée $A^n\cdot v$ s'«approche» (dans un sens qui reste à préciser) de l'espace propre pour la valeur propre qui est la plus grande en valeur absolue. Ce constat est très important en analyse numérique, et est à la base de nombreux algorithmes qui veulent éviter (car trop coûteuse) notamment la détermination d'une base de diagonalisation. Mais ces considérations étant d'une nature différente de l'approche algébrique de ce cours, on en dira pas plus.

Un autre exemple illustrera ce qui se passe quand une matrice n'a pas de valeurs propres réelles, mais est diagonalisable sur les nombres complexes. En reprenant la matrice $R=\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$ de ce type qu'on a vue avant, on peut observer qu'il s'agit d'un quart de tour, d'où $R^2=\left(\begin{smallmatrix}-1&0\\0&-1\end{smallmatrix}\right)$, $R^3=\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)$, $R^4=\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$ est la matrice identité, et pour toute autre puissance de R on trouvera la même matrice qu'en remplaçant l'exposant par son reste modulo 4. Mais même si dans ce cas on voit les puissances de la matrice tout de suite, on peut aussi (en supposant que la matrice correspond à un endomorphisme d'un espace *complexe* de dimension 2) les déterminer en utilisant la base de vecteurs propres $v_1=(1,-\mathbf{i})_{\mathcal{B}}$ et $v_2=(1,-\mathbf{i})_{\mathcal{B}}$

(où $\mathcal{B}$ est la base par rapport à laquelle la matrice est R). La matrice de passage de $\mathcal{B}$ à $[v_1, v_2]$ est $P = \left(\begin{smallmatrix} 1 & 1 \\ -\mathbf{i} & \mathbf{i} \end{smallmatrix}\right)$ avec inverse $P^{-1} = \frac{1}{2}\left(\begin{smallmatrix} 1 & \mathbf{i} \\ 1 & -\mathbf{i} \end{smallmatrix}\right)$; alors

$$\mathrm{Mat}_{[v_1,v_2]}(\phi) = D = \begin{pmatrix} \mathbf{i} & 0 \\ 0 & -\mathbf{i} \end{pmatrix} = P^{-1} \cdot R \cdot P = \frac{1}{2} \begin{pmatrix} 1 & \mathbf{i} \\ 1 & -\mathbf{i} \end{pmatrix} \cdot \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ -\mathbf{i} & \mathbf{i} \end{pmatrix}$$

et donc

$$\begin{aligned} R^n = P \cdot D^n \cdot P^{-1} &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -\mathbf{i} & \mathbf{i} \end{pmatrix} \cdot \begin{pmatrix} \mathbf{i}^n & 0 \\ 0 & (-\mathbf{i})^n \end{pmatrix} \cdot \begin{pmatrix} 1 & \mathbf{i} \\ 1 & -\mathbf{i} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} \mathbf{i}^n + (-\mathbf{i})^n & \mathbf{i}^{n+1} + (-\mathbf{i})^{n+1} \\ -(\mathbf{i}^{n+1} + (-\mathbf{i})^{n+1}) & \mathbf{i}^n + (-\mathbf{i})^n \end{pmatrix} \end{aligned}$$

On peut observer que, à cause du fait que (les valeurs propres) $\mathbf{i}$ et $-\mathbf{i}$ sont conjuguées complexes, ces matrices sont réelles pour tout n. En plus, les termes sur le diagonal s'annulent pour n impair, et les autres termes pour n pair, et on retrouve les puissances de R mentionnées ci-dessus, qui sont périodiques modulo 4 en fonction de n. Ce dernier fait est une particularité due au fait que la puissance 4 des valeurs propres est 1 (on dit qu'elles sont des 4-èmes racines de l'unité). La relation entre la matrice diagonale et la matrice correspondante sur la base $\mathcal{B}$ devient plus claire si l'on remplace le couple de valeurs propres par un couple de conjugués plus général, qu'on écrira sous la forme exponentielle $(re^{\mathbf{i}\theta}, re^{-\mathbf{i}\theta})$. On aura alors

$$D = \begin{pmatrix} re^{\mathbf{i}\theta} & 0 \\ 0 & re^{-\mathbf{i}\theta} \end{pmatrix}$$

et

$$\begin{aligned} P \cdot D^n \cdot P^{-1} &= \frac{1}{2} \begin{pmatrix} 1 & 1 \\ -\mathbf{i} & \mathbf{i} \end{pmatrix} \cdot \begin{pmatrix} r^n e^{n\mathbf{i}\theta} & 0 \\ 0 & r^n e^{-n\mathbf{i}\theta} \end{pmatrix} \cdot \begin{pmatrix} 1 & \mathbf{i} \\ 1 & -\mathbf{i} \end{pmatrix} \\ &= \frac{1}{2} \begin{pmatrix} r^n(e^{n\mathbf{i}\theta} + e^{-n\mathbf{i}\theta}) & r^n(\mathbf{i}e^{n\mathbf{i}\theta} - \mathbf{i}e^{-n\mathbf{i}\theta}) \\ -r^n(\mathbf{i}e^{n\mathbf{i}\theta} - \mathbf{i}e^{-n\mathbf{i}\theta}) & r^n(e^{n\mathbf{i}\theta} + e^{-n\mathbf{i}\theta}) \end{pmatrix} = r^n \begin{pmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{pmatrix}. \end{aligned}$$

Ainsi les puissances d'un couple de conjugués complexes sur cette base particulière de vecteurs propres se traduit sur la base $\mathcal{B}$ en la puissance correspondante de la similitude composée d'une rotation par θ et une homothétie de facteur r. En règle générale, si les puissances d'une matrice réelle (appliquée à un vecteur) manifestent un comportement de rotation ou d'oscillation, éventuellement en combinaison avec une croissance ou décroissance exponentielle, cela est souvent une indication d'un couple de conjugués complexes comme valeurs propres, si l'on interprète la matrice comme une matrice complexe.

2.3. *Existence de valeurs propres.*

On a vu que les vecteurs propres peuvent rendre la description d'un endomorphisme ϕ plus transparente, surtout quand on peut trouver une base constituée de vecteurs propres (le cas où ϕ est diagonalisable). Mais on a aussi vu dans les exemples que, du moins pour les endomorphismes de $\mathbf{R}$-espaces vectoriels, on peut ne pas avoir des vecteurs propres du tout. Dans cette section on montrera que cette situation ne se produit pas pour les endomorphismes de $\mathbf{C}$-espaces vectoriels (de dimension finie et non nulle) : un tel endomorphisme admet au moins un vecteur propre. La propriété du corps $\mathbf{C}$ qui le distingue à cet égard de $\mathbf{R}$ (ou de $\mathbf{Q}$) est le fait que $\mathbf{C}$ est *algébriquement clos* : tout polynôme de degré > 0 (à coefficients dans $\mathbf{C}$) possède au moins une racine dans $\mathbf{C}$ (c'est l'énoncé du Théorème de d'Alembert–Gauss, prouvé en 1814 par le mathématicien suisse Jean-Robert Argand, et qu'on admettra ici).

Il peut paraître étrange que cette propriété de polynômes joue un rôle pour la question d'existence de vecteurs propres, car rien dans la définition de ces derniers, ou des valeurs propres associées, fait référence à des polynômes. Néanmoins, on verra dans la suite qu'on peut associer de façon naturelle un polynôme à ϕ, de telle sorte que les valeurs propres de ϕ sont précisément les racines de ce polynôme. Il y aura même deux façons *différentes* de le faire, à savoir d'associer à ϕ son polynôme caractéristique χ_ϕ ou son polynôme minimal μ_ϕ. Mais si ces constructions sont importantes, on n'aura pas besoin de les détailler juste pour montrer l'existence d'un vecteur propre, car on peut obtenir cette existence par un argument qui est certes un peu astucieux, mais qui montre bien comment les polynômes interviennent en faisant correspondre l'indéterminée X à l'endomorphisme ϕ, et ses puissances X^n au puissances (les itérées) ϕ^n

de ϕ. On remarque en passant que le polynôme constant $1 = X^0$ correspondra à l'endomorphisme ϕ^0, défini par convention comme $\phi^0 = \mathrm{id}_E$, l'identité de l'espace E sur lequel opère ϕ, quel que soit ϕ.

Avant de pouvoir donner l'argument, il nous faudra un minimum de terminologie et de résultats sur les polynômes, qu'on résume ici ; les polynômes seront traités plus systématiquement dans le chapitre 3, où on trouvera des justifications (d'ailleurs fort simples) de ces résultats. Un polynôme P à coefficients dans K, noté $P \in K[X]$, peut être donné par une expression construite à partir des scalaires de K ainsi que d'une indéterminée X en les combinant par addition, soustraction et multiplication. Pour un scalaire $a \in K$, une opération $K[X] \to K$ est définie, dite de substitution de a pour X, dont le résultat est obtenu en remplaçant dans l'expression systématiquement les occurrences de X par a, et en effectuant ensuite les opérations dans K. On appelle a une racine de $P \in K[X]$ si cette substitution appliquée à P résulte en la valeur $0 \in K$. En particulier a est racine du polynôme $X - a$, ainsi que de tout polynôme $(X - a)Q$ obtenu en multipliant $X - a$ par un autre polynôme $Q \in K[X]$. La réciproque est aussi vraie : si a est racine de P, alors il existe $Q \in K[X]$ tel qu'on puisse écrire $P = (X - a)Q$.

Dans ce qui suit ϕ sera une endomorphisme d'un espace vectoriel E, et on fixera un vecteur $v \in E$. On associera alors à chaque polynôme $P \in K[X]$ un vecteur qui sera noté $P \cdot_\phi v$, et qu'on obtient en remplaçant dans chaque terme de P l'indéterminée X par ϕ et en faisant opérer l'endomorphisme ainsi obtenu sur v ; la somme des vecteurs résultant ainsi des termes de P donne $P \cdot_\phi v$. Par exemple $(X^5 - 2X + 7) \cdot_\phi v = \phi^5(v) - 2\phi(v) + 7v \in E$. La propriété cruciale de cette opération est que l'opération successive sur v de deux polynômes est équivalente à l'opération par le produit des deux polynômes :

$$P \cdot_\phi (Q \cdot_\phi v) = (PQ) \cdot_\phi v. \tag{11}$$

C'est une propriété purement formelle, c'est à dire elle découle de la similarité des manipulations dont consiste l'application de $P \cdot_\phi$ et celles qui interviennent dans la multiplication d'un polynôme par P. Par exemple, on peut opérer avec $P = 3X - 4$ sur l'exemple ci-dessus, pour trouver d'un coté

$$\begin{aligned}(3X - 4) \cdot_\phi (\phi^5(v) - 2\phi(v) + 7v) &= 3\phi(\phi^5(v) - 2\phi(v) + 7v) - 4(\phi^5(v) - 2\phi(v) + 7v) \\ &= 3\phi^6(v) - 6\phi^2(v) + 21\phi(v) - 4\phi^5(v) + 8\phi(v) - 28v,\end{aligned}$$

et d'autre côté pour le produit $P \times (X^5 - 2X + 7)$:

$$\begin{aligned}(3X - 4)(X^5 - 2X + 7) &= 3X(X^5 - 2X + 7) - 4(X^5 - 2X + 7) \\ &= 3X^6 - 6X^2 + 21X - 4X^5 + 8X - 28,\end{aligned}$$

dont l'opération sur v donne clairement le même résultat que ci-dessus. On laisse au lecteur assidu le soin de formuler une preuve de (11), en remarquant qu'une preuve sera donnée dans un chapitre ultérieur.

2.3.1. Proposition. *Soit ϕ un endomorphisme d'un $\mathbf{C}$-espace vectoriel E de dimension finie et non nulle. Alors E contient (au moins) un vecteur propre de ϕ.*

Preuve. Soit v un vecteur non nul (c'est ici qu'on utilise $\dim(E) \neq 0$). En itérant l'application de ϕ à partir de v, on obtient une suite de vecteurs $v = \phi^0(v), \phi(v), \phi^2(v), \phi^3(v), \ldots$ de E. Des qu'on a obtenu plus de $\dim(E)$ vecteurs de cette suite, on est sûr d'avoir une famille liée (le nombre d'éléments d'une famille libre ne peut pas dépasser la dimension de l'espace). Soit $d \leq \dim(E)$ le plus petit nombre tel que la famille $[\phi^0(v), \ldots, \phi^d(v)]$ soit liée. On a $d > 0$ car $v \neq \vec{0}$, et par minimalité de d, la famille $[\phi^0(v), \ldots, \phi^{d-1}(v)]$ est libre. Ainsi $\phi^d(v) \in \mathrm{Vect}(\phi^0(v), \ldots, \phi^{d-1}(v))$, disons $\phi^d(v) = -a_0\phi^0(v) - \cdots - a_{d-1}\phi^{d-1}(v)$ pour certains $a_0, \ldots, a_{d-1} \in \mathbf{C}$. Posons $P = a_0X^0 + \cdots + a_{d-1}X^{d-1} + X^d \in \mathbf{C}[X]$, de sorte que cette équation s'écrive sous la forme $P \cdot_\phi v = \vec{0}$. D'après le théorème de d'Alembert–Gauss P possède (au moins) une racine λ, ce qui veut dire que $P = (X - \lambda)Q$ pour un certain polynôme $Q \in \mathbf{C}[X]$, qui sera de degré $d - 1$ et unitaire (son coefficient de X^{d-1} est 1). Alors $\vec{0} = P \cdot_\phi v = (X - \lambda) \cdot_\phi (Q \cdot_\phi v) = (\phi - \lambda\,\mathrm{id}_E)(Q \cdot_\phi v)$. Ceci veut dire que le vecteur $Q \cdot_\phi v$ est dans l'espace propre $\mathrm{Ker}(\phi - \lambda\,\mathrm{id}_E)$ de ϕ pour λ. Or, par définition $w = Q \cdot_\phi v$ est une combinaison linéaire des vecteurs $\phi^0(v), \ldots, \phi^{d-1}(v)$, et même une combinaison linéaire non-triviale (car son coefficient de $\phi^{d-1}(v)$ est 1) ; comme la famille $[\phi^0(v), \ldots, \phi^{d-1}(v)]$ est libre, cela veut dire que $w \neq \vec{0}$, et w est donc un vecteur propre de ϕ (pour la valeur propre λ). □

Remarquons d'emblée que cette proposition ne permettra pas de trouver dans tous les cas *une base* de vecteurs propres : une telle base n'existe pas toujours si $\dim(E) \geq 2$. Le plus simple exemple est avec $\dim(E) = 2$ et un endomorphisme ϕ tel que $\phi(b_1) = b_2$ et $\phi(b_2) = \vec{0}$ pour une certaine base $[b_1, b_2]$; alors ϕ^2 est l'application nulle (on dit que ϕ est nilpotent si l'une de ses puissances est nulle), ce qui exclut toute autre valeur propre que 0. Mais l'espace propre $\mathrm{Ker}(\phi)$ pour $\lambda = 0$ est $\mathrm{Vect}(b_2)$ qui n'est que de dimension 1, donc il n'y a pas de base de E formée de vecteurs propres (pour $\lambda = 0$). On voit de la même façon qu'aucun endomorphisme nilpotent n'est diagonalisable, sauf l'endomorphisme nul.

Ces exemples d'endomorphismes non diagonalisables se généralisent facilement à d'autres valeurs propres que $\lambda = 0$, en ajoutant $\lambda \,\mathrm{id}_E$ à l'endomorphisme ϕ ; l'endomorphisme ainsi obtenu aura λ comme unique valeur propre sans que son espace propre soit l'espace entier. En formant une somme directe d'espaces avec un endomorphisme agissant séparément sur chaque facteur, on peut aussi produire des exemples d'endomorphismes non diagonalisables qui admettent plusieurs valeurs propres. Ce qu'on peut néanmoins observer dans tous ces exemples est que certaines valeurs propres ont un "territoire", dans le sens d'y exclure toute autre valeur propre, qui est plus grand que leur espace propre. On introduira plus tard la notion d'espace caractéristique pour formaliser cette idée.

Malgré ces exemples, les cas non diagonalisables seront exceptionnels parmi les endomorphismes d'une $\mathbf{C}$-espace vectoriel. Sans pouvoir prouver cela pour l'instant, on peut l'illustrer à l'aide de la démonstration de la proposition 2.3.1. Soit $S = \mathrm{Vect}(\phi^0(v), \ldots, \phi^{d-1}(v))$ le sous-espace engendré par les d premières puissances de ϕ appliquées à ϕ. En fait, on a vu qu'on a aussi $\phi^d(v) \in S$, et on en déduit que S est ϕ-stable, et contient donc *tous* les vecteurs de la forme $\phi^i(v)$ (et très souvent on aura $d = \dim(E)$, et donc $S = E$ tout entier). En prenant une racine λ du polynôme P, qui est de degré $d = \dim(S)$, on a construit un vecteur propre w pour λ de ϕ, et on voit facilement que $w \in S$. Mais en itérant l'application du théorème d'Alembert–Gauss, on peut factoriser P entièrement comme produit de facteurs de degré 1, c'est-à-dire écrire $P = (X - \lambda_1) \cdots (X - \lambda_d)$ pour certains $\lambda_1, \ldots, \lambda_d \in \mathbf{C}$ (parmi lesquels se trouve la racine λ). Comme elle l'a fait pour λ, la construction fournit un vecteur propre pour chacune de ces valeurs propres λ_i. Si l'on suppose que les λ_i soient *tous distincts*, on aura ainsi une famille de d vecteurs propres pour d valeurs propres différentes, et une telle famille étant libre, on aura une *base* de S : l'endomorphisme de S obtenu par restriction de ϕ sera diagonalisable. Ce n'est donc que dans les cas où P possède au moins une racine multiple, ce qui est exceptionnel pour un polynôme complexe "pris au hasard", que cet endomorphisme de S peut être non diagonalisable. (Mais en revanche, on verra plus tard que dans ce cas, cet endomorphisme sera en effet toujours non diagonalisable, tout comme ϕ lui-même.)

2.4. *Exemples d'application des vecteurs propres.*

Dans cette section on donnera quelques exemples dans lesquels il sera possible de trouver toutes les valeurs propres d'un endomorphisme, et qui illustrent l'utilité de cette démarche.

Les exemples concerneront des espaces de *suites récurrentes linéaires*. Elles forment des sous-espaces de l'espace $K^{\mathbf{N}}$ des suites infinies $(a_i)_{i \in \mathbf{N}} = (a_0, a_1, a_2, \ldots)$ de scalaires $a_i \in K$. Comme les ensembles K^n des suites d'une longueur finie n, l'ensemble $K^{\mathbf{N}}$ possède des opérations d'addition et de multiplication scalaire, définies terme par terme, qui lui donnent la structure de K-espace vectoriel (mais de dimension infinie dans le cas de $K^{\mathbf{N}}$). Une relation de récurrence linéaire est une condition de la forme

$$a_{i+d} = c_0 a_i + c_1 a_{i+1} + \cdots + c_{d-1} a_{i+d-1} \qquad \text{pour tout } i \in \mathbf{N} \tag{12}$$

déterminée par un d-uplet de constantes $c_0, \ldots, c_{d-1} \in K$, et les suites de $K^{\mathbf{N}}$ qui vérifient cette relation forment l'ensemble E de suites récurrentes défini par elle. Si $d = 1$ on obtient la relation qui définit les *suites géométriques* de raison c_0, et le plus célèbre exemple d'une suite récurrente linéaire est celle de Fibonacci, obtenue pour $d = 2$ et $c_0 = c_0 = 1$: chaque terme est la somme de ses deux prédécesseurs.

La condition de (12) pour un $i \in \mathbf{N}$ fixé est une relation linéaire sur $K^{\mathbf{N}}$ (parce qu'elle décrit le noyau de l'application linéaire $K^{\mathbf{N}} \to K$ donnée par $(a_i)_{i \in \mathbf{N}} \mapsto c_0 a_i + c_1 a_{i+1} + \cdots + c_{d-1} a_{i+d-1} - a_{i+d}$) et définit donc un sous-espace de $K^{\mathbf{N}}$ (ce noyau). L'intersection de tous ces sous-espaces obtenus quand i parcourt $\mathbf{N}$ est E, qui est donc aussi un sous-espace de $K^{\mathbf{N}}$. Si deux suites de E coïncident en leur d premiers termes, on montre par récurrence sur i qu'elles coïncident en chaque position i, et sont donc des

suites identiques. Aussi, quel que soit $(v_0, \ldots, v_{d-1}) \in K^d$, on peut poser $a_i = v_i$ pour $i < d$ et définir les nombre a_i pour $d \leq i$ récursivement par (12), pour obtenir une suite qui appartient à E. Ainsi l'application $E \to K^d$ qui associe à une suite infinie ses d premiers termes, c'est-à-dire $(a_i)_{i\in\mathbf{N}} \mapsto (a_0, \ldots, a_{d-1})$, est une bijection ; comme c'est aussi clairement un application linéaire, c'est un isomorphisme de K-espaces vectoriels, ce qui montre en particulier que $\dim(E) = d$. Un tel isomorphisme peut être vu comme celui qui associe aux des suites de E leurs coordonnées dans une certaine base $\mathcal{B}$ de E ; explicitement, l'élément $\mathbf{b}_k$ de $\mathcal{B}$, pour $0 \leq k < d$, est la suite $(a_i)_{i\in\mathbf{N}}$ dont les d premiers termes sont tous nuls, sauf a_k qui est 1, et dont les autres termes sont déterminés par (12). Concrètement, pour la relation de récurrence de Fibonacci, les deux suites suivantes qui constituent la base $\mathcal{B}$:

$$\mathbf{b}_0 = (1, 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, \ldots),$$
$$\mathbf{b}_1 = (0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, \ldots).$$

Le fait que $\mathbf{b}_1$ est obtenue par un décalage de $\mathbf{b}_0$ est particulier à cet exemple, comme on peut voir en changeant le récurrence en $a_{i+2} = 2a_i + a_{i+1}$, auquel cas on obtiendrait comme base le couple de suites

$$\mathbf{b}'_0 = (1, 0, 2, 2, 6, 10, 22, 42, 86, 170, 342, 682, \ldots),$$
$$\mathbf{b}'_1 = (0, 1, 1, 3, 5, 11, 21, 43, 85, 171, 341, 683, \ldots).$$

En revanche, le fait que le décalage d'une suite de E (c'est-à-dire la suite obtenue en supprimant le premier terme et en numérotant à nouveau les termes restants) appartient à E est vrai en général, et facile à comprendre : pour que la suite décalée vérifie (12) à l'indice i, il suffit que la suite originale vérifie (12) à l'indice $i + 1$, ce qui est le cas par hypothèse. (Ce qui est en jeu ici est le fait que les coefficients $c_0, \ldots, c_{d-1}$ de (12) sont *constants* ; si au contraire ils dépendaient de i, la relation déterminerait toujours un sous-espace vectoriel de $K^{\mathbf{N}}$ de dimension d, mais le décalage d'une suite de ce sous-espace ne serait pas forcément dans ce sous-espace.) Le résultat de ce décalage s'écrit donc comme une combinaison linéaire des suites $\mathbf{b}_0, \ldots, \mathbf{b}_{d-1}$ de la base $\mathcal{B}$; par exemple le décalage de $\mathbf{b}_1$ ci-dessus est $\mathbf{b}_0 + \mathbf{b}_1$.

Comme l'opération de décalage est aussi linéaire, elle définit un endomorphisme de E, et c'est cet endomorphisme δ qui permettra de mieux comprendre le comportement des suites de E. On a défini le décalage comme une opération qui supprime le premier terme et déplace les autres termes une place vers la gauche, mais il peut aussi être utile de le visualiser comme l'opération qui consiste d'avancer vers la droite le long la suite, qui elle reste fixe (mais dont on perd de vue le terme initial), un peu comme dans un vieux film la projection en arrière plan d'une route qui s'approche du spectateur suggère que la scène s'avance sur cette route. Ainsi les puissances δ^n décriront le calcul des termes successifs de la suite.

La matrice de δ par rapport à la base $\mathcal{B}$ est simple à déterminer. Les coordonnées dans $\mathcal{B}$ d'une suite $\mathbf{s}$ sont ses d premiers termes (dans l'ordre). Or, les $d - 1$ premiers termes de la suite décalée $\delta(\mathbf{s})$ sont parmi les d premiers termes de $\mathbf{s}$, car ce sont ces termes sauf le terme initial (qui est supprimé par δ). La seule coordonnée de $\delta(\mathbf{s})$ dans $\mathcal{B}$ qui demande un calcul est donc la dernière (des d). Elle est égale au terme à l'indice $d - 1$ de $\delta(\mathbf{s})$, et donc au terme à l'indice d de $\mathbf{s}$, autrement dit si $\mathbf{s} = (a_i)_{i\in\mathbf{N}}$, c'est a_d. Ce terme est déterminé par la relation (12) pour $i = 0$; si on prend pour $\mathbf{s}$ le vecteur $\mathbf{b}_k$ de la base $\mathcal{B}$ (avec $0 \leq k < d$), les termes a_j dans le second membre de (12) seront tous nuls sauf a_k qui est 1, et on obtient dans ce cas $a_d = c_k$. Ainsi on trouve comme matrice de δ par rapport à la base $\mathcal{B}$:

$$\mathrm{Mat}_{\mathcal{B}}(\delta) = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ddots & \vdots \\ 0 & 0 & 0 & \ddots & 0 \\ \vdots & \vdots & \vdots & \ddots & 1 \\ c_0 & c_1 & c_2 & \ldots & c_{d-1} \end{pmatrix} \tag{13}$$

L'opération de décalage est bien définie dans l'espace $K^{\mathbf{N}}$ tout entier, et dire pour $\mathbf{s} = (a_i)_{i\in\mathbf{N}}$ que sa suite décalée $(a_{i+1})_{i\in\mathbf{N}}$ est égale à un multiple $\lambda\mathbf{s}$ de $\mathbf{s}$ veut dire que les termes de $\mathbf{s}$ vérifient $a_{i+1} = \lambda a_i$ pour tout $i \in \mathbf{N}$, autrement dit que $\mathbf{s}$ est une suite géométrique de raison λ. Le problème de trouver le vecteurs propres de δ est donc celui de déterminer quelles suites géométriques sont dans le sous-espace

E de $K^{\mathbf{N}}$, autrement dit quelles suites géométriques (non nulles) vérifient (12). Si on met $a_k = a_0\lambda^k$ pour $k \in \mathbf{N}$ dans (12), avec $a_0 \neq 0$, alors tous les termes sont un multiple de $a_0\lambda^i$; on peut constater que la condition pour un i quelconque est une conséquence de celle pour $i = 0$, quelle condition devient, après simplification par a_0 :

$$\lambda^d = c_0 + c_1\lambda^1 + \cdots + c_{d-1}\lambda^{d-1}, \tag{14}$$

autrement dit on obtient la condition que λ soit racine du polynôme $Q = X^d - c_{d-1}X^{d-1} - \cdots - c_1X^1 - c_0$. Dans le cas concret de la récurrence de Fibonacci, ce polynôme est $X^2 - X - 1$, qui a deux racines dans $\mathbf{R}$, à savoir $\frac{1+\sqrt{5}}{2}$ (le nombre d'or, φ, approximativement 1,618) et $\frac{1-\sqrt{5}}{2} = 1 - \varphi = -\frac{1}{\varphi} \approx -0{,}618$).

Dans le cas où le polynôme Q possède d racines *distinctes* $\lambda_1, \ldots, \lambda_d$ dans K, l'endomorphisme δ admet donc une famille de d vecteurs propres associés à ces d valeurs propres, pour lesquels on peut prendre les d suites géométriques $\mathbf{g}_{\lambda_i} = (1, \lambda_i, \lambda_i^2, \lambda_i^3, \ldots) = (\lambda_i^k)_{k\in\mathbf{N}}$ pour $i = 1, \ldots, d$. Une telle famille étant toujours libre (voir la remarque après la proposition 2.2.2), elle forme dans ce cas une base de l'espace E des suites vérifiant la relation de récurrence, dont on sait qu'il est de dimension d. Par conséquent, l'endomorphisme δ est diagonalisable si Q possède d racines distinctes dans K ; comme on a vu que les seuls vecteurs propres possibles pour l'opération de décalage sont les suites géométriques (et aucun espace propre ne peut donc avoir une dimension plus grande que 1), on peut ajouter que réciproquement, si Q ne possède pas d racines distinctes dans K, alors δ n'est pas diagonalisable.

Pour le cas de la récurrence de Fibonacci, on a trouvé effectivement deux valeurs propres distinctes, donc les deux suites géométriques $\mathbf{g}_1 = \left((\frac{1+\sqrt{5}}{2})^k\right)_{k\in\mathbf{N}}$ et $\mathbf{g}_2 = \left((\frac{1-\sqrt{5}}{2})^k\right)_{k\in\mathbf{N}}$ forment une base de E. Pour exprimer ces deux vecteurs en coordonnées dans la base $\mathcal{B} = [\mathbf{b}_0, \mathbf{b}_1]$ il suffit de prendre les deux premiers termes de ces suites, et on a donc $\mathbf{g}_1 = \mathbf{b}_0 + (\frac{1+\sqrt{5}}{2})\mathbf{b}_1 = (1, \frac{1+\sqrt{5}}{2})_{\mathcal{B}}$ et $\mathbf{g}_2 = \mathbf{b}_0 + (\frac{1-\sqrt{5}}{2})\mathbf{b}_1 = (1, \frac{1-\sqrt{5}}{2})_{\mathcal{B}}$. La matrice de passage P de $\mathcal{B}$ vers la base $[\mathbf{g}_1, \mathbf{g}_2]$ des vecteurs propres est donc

$$P = \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix}, \qquad \text{dont la matrice l'inverse est} \qquad P^{-1} = \frac{1}{\sqrt{5}} \begin{pmatrix} \frac{-1+\sqrt{5}}{2} & 1 \\ \frac{1+\sqrt{5}}{2} & -1 \end{pmatrix}.$$

Les nombres de Fibonacci F_n sont les termes de la suite vérifiant la relation de récurrence $F_{n+2} = F_n + F_{n+1}$et avec termes initiales $F_0 = 0$ et $F_1 = 1$, autrement dit $(F_n)_{n\in\mathbf{N}} = \mathbf{b}_1$. On peut les exprimer en termes des puissances de δ, car F_n est le terme initial de $\delta^n(\mathbf{b}_1)$; concrètement, en utilisant l'expression (13) la matrice de δ par rapport à la base $\mathcal{B}$, on a

$$\mathrm{Mat}_{\mathcal{B}}(\delta) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{et donc} \qquad (F_n) = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}^n \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

(la matrice colonne à droite est celle des coordonnées de $\mathbf{b}_1$ dans la base $\mathcal{B} = [\mathbf{b}_0, \mathbf{b}_1]$, la matrice ligne à gauche sert à extraire la première coordonnée dans cette base, qui est égale au terme initial de la suite).

La diagonalisation de δ par rapport à la base $[\mathbf{g}_1, \mathbf{g}_2]$ permet de trouver une formule explicite pour F_n. On a

$$\mathrm{Mat}_{\mathcal{B}}(\delta) = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} = P \cdot \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix} \cdot P^{-1}$$

et donc

$$\begin{aligned}
(F_n) &= \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{pmatrix} \cdot \begin{pmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{pmatrix}^n \cdot \frac{1}{\sqrt{5}} \begin{pmatrix} \frac{-1+\sqrt{5}}{2} & 1 \\ \frac{1+\sqrt{5}}{2} & -1 \end{pmatrix} \cdot \begin{pmatrix} 0 \\ 1 \end{pmatrix} \\
&= \frac{1}{\sqrt{5}} \begin{pmatrix} 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} \left(\frac{1+\sqrt{5}}{2}\right)^n & 0 \\ 0 & \left(\frac{1-\sqrt{5}}{2}\right)^n \end{pmatrix} \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\
&= \frac{1}{\sqrt{5}} \left(\left(\frac{1+\sqrt{5}}{2}\right)^n - \left(\frac{1-\sqrt{5}}{2}\right)^n \right)
\end{aligned}$$

On trouve dans cette expression les puissances des deux valeurs propres de δ, dont le première φ est plus grand que 1, pendant que la seconde $-\frac{1}{\varphi}$ est (négative et) plus petite en valeur absolue que 1. Par conséquent, le premier terme dominera largement le second quand n devient grand, et on pourra conclure qu'alors $F_n \approx \varphi^n/\sqrt{5}$ est une très bonne approximation, et que $\lim_{n\to\infty} F_{n+1}/F_n = \varphi = \frac{1+\sqrt{5}}{2}$.

Comme autre exemple on peut considérer un problème d'équations différentielles en analyse ; même si le contexte est assez différent, sa nature algébrique sera assez semblable à celle de l'exemple précédent. On cherche à connaître l'espace E des fonctions $f : \mathbf{R} \to K$, avec $K = \mathbf{R}$ ou $K = \mathbf{C}$, suffisamment dérivables, qui vérifient une équation différentielle linéaire d'ordre d à coefficients constants :

$$f^{(d)}(t) = c_o f(t) + c_1 f'(t) + c_2 f^{(2)}(t) + \cdots + c_{d-1} f^{(d-1)}(t) \qquad \text{pour tout } t \in \mathbf{R}, \tag{15}$$

dont on voit facilement que c'est une espace K-vectorielle. Une telle équation avec $d = 2$ est très connue en mécanique : elle exprime l'accélération $f''(t)$ instantanée d'un particule comme une fonction linéaire de la position $f(t)$ et de la vitesse $f'(t)$ instantanées. Dans cette application les constantes c_0 et c_1 seront négatives en règle générale : $-c_0$ correspond à une force élastique qui repousse le particule vers la position d'équilibre (la valeur 0), quelle force est proportionnelle à la déviation de cette position, et $-c_1$ correspond à une force de frottement opposée à la vitesse qui tend à amortir le mouvement (l'hypothèse que cette force varie de façon *linéaire* avec la vitesse n'est réaliste que pour certains mécanismes de frottement, mais si l'amortissement est faible l'approximation peut néanmoins être bonne). Pour les deux coefficients, la constante donnant la force par unité de distance, respectivement par unité de vitesse, doit encore être divisée par la masse du particule pour donner l'accélération nécessaire dans l'équation (15). Dans le cas ou $c_0 < 0$ et $c_1 = 0$ (absence de frottement) on parle de l'équation d'un oscillateur harmonique, et dans le cas $c_0 < 0$ et $c_1 < 0$ on parle de l'équation d'un oscillateur harmonique amorti.

On admettra ici le fait que deux solutions f, g de cette équation qui pour un certain x_0 vérifient $f(t_0) = g(t_0)$ et $f^{(i)}(t_0) = g^{(i)}(t_0)$ pour $i = 1, \ldots d-1$ seront identiques partout, ce qu'on sait montrer en analyse. Par conséquent l'application linéaire $E \to K^d$ donnée par $f \mapsto \big(f(t_0), f'(t_0), \ldots, f^{(d-1)}(t_0)\big)$ est injective, et en particulier E est de dimension finie au plus d (en fait le raisonnement en analyse montrera aussi l'existence de solutions, c'est-à-dire la surjectivité de l'application, et donc on aura toujours $\dim(E) = d$.) En prenant la dérivée de l'équation (15), on voit que si f en est une solution, f' en est une solution aussi. L'opération de dérivation est donc un endomorphisme δ de E. Un vecteur propre de δ pour une valeur propre λ est une fonction f qui vérifie $f' = \lambda f$, autrement dit c'est une fonction qui vérifie l'équation (15) pour $d = 1$ et $c_0 = \lambda$; on sait que les seules telles fonction sont des multiples scalaires de la fonction exponentielle $t \mapsto e^{\lambda t}$. Pour que cette fonction appartienne à E, il faut que λ vérifie la même équation (14) qu'on a vue pour les suites récurrentes linéaires, c'est-à-dire que λ soit racine du polynôme $Q = X^d - c_{d-1}X^{d-1} - \cdots - c_1 X^1 - c_0$. En fait, on peut vérifier que pour la base $\mathcal{B}$ telle que l'isomorphisme linéaire $E \to K^d$ ci-dessus avec $t_0 = 0$ coïncide avec celle de l'expression en coordonnées dans $\mathcal{B}$, la matrice de δ par rapport à $\mathcal{B}$ est aussi donnée par (13). On peut conclure, comme dans le cas des suites récurrentes, que δ sera diagonalisable si et seulement si Q possède d racines distinctes dans K.

En considérant le cas de l'oscillateur harmonique, on aura $Q = X^2 - c_0$ avec $c_0 < 0$; c'est un polynôme quadratique à deux racines complexes, distinctes et conjuguées complexes, $\lambda = \sqrt{-c_0}\,\mathbf{i}$ et $\overline{\lambda} = -\sqrt{-c_0}\,\mathbf{i}$. Pour que δ soit diagonalisable il faut donc dans ce prendre $K = \mathbf{C}$, c'est-à-dire considérer des fonctions à valeurs complexes. On aura alors comme bas de vecteurs propres $\mathcal{E} = [e_\lambda, e_{\overline{\lambda}}]$ où $e_\lambda : t \mapsto e^{\lambda t}$ et $e_{\overline{\lambda}} : t \mapsto e^{\overline{\lambda} t}$, avec comme matrice de passage de $\mathcal{B}$ vers cette base

$$P = \begin{pmatrix} 1 & 1 \\ \lambda & \overline{\lambda} \end{pmatrix}, \qquad \text{dont la matrice l'inverse est} \qquad P^{-1} = \frac{1}{2}\begin{pmatrix} 1 & 1/\lambda \\ 1 & 1/\overline{\lambda} \end{pmatrix}.$$

Ici, ce ne ont pas tellement les puissances δ^n qui nous intéressent, mais l'expression explicite des vecteurs de la base $\mathcal{B} = [\mathbf{b}_0, \mathbf{b}_1]$, qu'on peut trouver en les exprimant dans la base $\mathcal{E}$ des vecteurs propres, expressions dont les coefficients sont données par les colonnes de la matrice P^{-1} :

$$\mathbf{b}_0 = \frac{1}{2}(1,1)_{\mathcal{E}} = \frac{1}{2}(e_\lambda + e_{\overline{\lambda}}) : t \mapsto \frac{1}{2}(e^{\lambda t} + e^{\overline{\lambda} t}) = \cos(\sqrt{-c_0}t)$$
$$\mathbf{b}_1 = \frac{1}{2\sqrt{-c_0}\,\mathbf{i}}(1,-1)_{\mathcal{E}} = \frac{1}{2\sqrt{-c_0}\,\mathbf{i}}(e_\lambda - e_{\overline{\lambda}}) : t \mapsto \frac{1}{2\sqrt{-c_0}\,\mathbf{i}}(e^{\lambda t} - e^{\overline{\lambda} t}) = \frac{1}{\sqrt{-c_0}}\sin(\sqrt{-c_0}t)$$

Donc les solutions à valeurs réelles de l'équation de l'oscillateur harmonique sont les combinaisons linéaires réelles des fonctions périodiques $t \mapsto \cos(\sqrt{-c_0}t)$ et $t \mapsto \sin(\sqrt{-c_0}t)$ (le facteur $1/\sqrt{-c_0}$ dans l'expression pour $\mathbf{b}_1$ vient du fait que par définition $\mathbf{b}_1'(0) = 1$).

Dans le cas de l'oscillateur harmonique amorti, on aura $Q = X^2 - c_1X - c_0$ avec $c_0 < 0$ et $c_1 < 0$, un polynôme quadratique de discriminant $\Delta = c_1^2 + 4c_0$. Si $|\frac{c_1}{2}| < \sqrt{-c_0}$, ce discriminant sera négatif comme pour oscillateur harmonique, et les racines de Q seront encre deux nombres conjugués complexes, mais elles ne sont plus purement imaginaires, car elles ont un partie réelle $\frac{c_1}{2} < 0$. Comme vecteurs propres on aura dans ce cas les fonctions exponentielles complexes $t \mapsto \exp(\frac{c_1 \pm \sqrt{-\Delta}\,\mathbf{i}}{2}t)$, dont on peut former les combinaisons linéaires à valeurs réelles $t \mapsto \exp(\frac{c_1}{2}t)\cos(\frac{\sqrt{-\Delta}}{2}t)$ et $t \mapsto \exp(\frac{c_1}{2}t)\sin(\frac{\sqrt{-\Delta}}{2}t)$. (Dans ce cas la première de ces deux solutions n'est pas proportionnelle à la solution $\mathbf{b}_0$, car sa dérivée en $t = 0$ n'est pas nulle ; on laisse au lecteur l'exercice de calculer l'inverse de la matrice de passage, qui exprime $\mathbf{b}_0, \mathbf{b}_1$ dans la base $\mathcal{E}$.) Par rapport à l'oscillateur harmonique non amorti, on observe la présence d'un facteur décroissante exponentielle $\exp(\frac{c_1}{2}t)$, ainsi qu'un facteur $\frac{\sqrt{-\Delta}}{2}$ déterminant la fréquence des oscillations qui est légèrement plus petit que le facteur $\sqrt{-c_0}$ (le fait d'amortir l'oscillateur baisse aussi sa fréquence).

Si par contre $|\frac{c_1}{2}| > \sqrt{-c_0}$, on aura $\Delta > 0$ et les racines seront deux nombres réels distincts qu'on désignera $\lambda_0 = \frac{c_1+\sqrt{\Delta}}{2}$ et $\lambda_1 = \frac{c_1-\sqrt{\Delta}}{2}$; comme c_0 et c_1 sont négatifs on aura $\lambda_1 < \lambda_0 < 0$. La solution générale sera donc de la forme $t \mapsto ae^{\lambda_0 t} + be^{\lambda_1 t}$ pour des constantes réelles a, b, donc il n'y aura pas d'oscillation (la solution pas au plus une fois par 0). Dans une telle solution le terme $ae^{\lambda_0 t}$ dominera quand t est grand (sauf si $a = 0$) donc l'évolution ressemble beaucoup à une décroissance exponentielle.

Le cas $|\frac{c_1}{2}| = \sqrt{-c_0}$ est spécial car $\Delta = 0$, et Q possède une racine double $\frac{c_1}{2}$. Cela veut dire que δ n'est pas diagonalisable, même avec $K = \mathbf{C}$. Dans ce cas spécial on parle d'un amortissement critique. On peut vérifier que, à part de la solution exponentielle $t \mapsto \exp(\frac{c_1}{2}t)$, on a comme solution indépendante dans ce cas $t \mapsto t\exp(\frac{c_1}{2}t)$, donc la solution générale sera de la forme $t \mapsto (a + bt)\exp(\frac{c_1}{2}t)$ avec $a, b \in K$.

Chapitre 3. Corps et anneaux, arithmétique, polynômes.

Dans ce chapitre on fera un survol de divers sujets concernant des structures algébriques. Ce sont, de façon très générale, des ensembles munis de certains opérations, qui vérifient certaines propriétés appelées axiomes. Le nom de la structure (on verra corps, anneau, groupe, monoïde, mais il en existe beaucoup d'autres) correspond à la collection d'axiomes, et pourra s'appliquer à divers ensembles vérifiant ces axiomes (corps des nombres rationnels, réels ou complexes, anneau d'entiers, anneau de polynômes, groupe de permutations, groupe d'isométries, ...). L'utilité de cet approche abstraite se trouve dans sa puissance : tant que des propriétés qu'on peut formuler pour une certaine structure sont déduites en utilisant uniquement ses axiomes, elles sont valables pour toutes les instances de cette structure, sans avoir besoin d'être démontrées séparément pour chaque cas. Cependant, il n'est pas notre but ici de développer de telles théories en détail, mais juste de donner un premier aperçu des structures qui existent et dont on se servira de quelques instances simples.

3.1. *Définition de corps et anneaux.*

On a déjà évoqué la notion de corps commutatif au début de ce cours, car elle est un préalable à la notion d'espace vectoriel.

3.1.1. Définition. *Un anneau est un ensemble R muni d'opérations '$+$' : $R \times R \to R$, '$\times$' : $R \times R \to R$, ainsi que des constantes notées $0, 1 \in R$, tel que, pour tout $a, b, c \in R$:*

(1) $0 + a = a = a + 0$ (0 est élément neutre pour l'addition);
(2) il existe un élément, noté $-a$, tel que $a + -a = 0 = -a + a$ (existence de symétriques pour l'addition);
(3) $a + (b + c) = (a + b) + c$ (associativité de l'addition);
(4) $a + b = b + a$ (commutativité de l'addition);
(5) $1 \times a = a = a \times 1$ (1 est élément neutre pour la multiplication);
(6) $a \times (b \times c) = (a \times b) \times c$ (associativité de la multiplication);
(7) $a \times (b + c) = (a \times b) + (a \times c)$ et $(a + b) \times c = (a \times c) + (b \times c)$ (distributivité à gauche et à droite).

On appelle R un anneau commutatif si en plus pour tout $a, b \in R$:

(8) $a \times b = b \times a$ (commutativité de la multiplication).

On appelle un anneau R (commutatif ou non) un corps si

(9) $1 \neq 0$, et pour tout $a \in R$ avec $a \neq 0$ il existe $b \in R$ tel que $a \times b = 1 = b \times a$ (existence de symétriques pour la multiplication).

Les conditions (1)–(4) expriment le fait que R muni seulement de '+' et de '0' forme un *groupe commutatif*, dit groupe additif de R. (La notion de groupe existe aussi sans exiger la commutativité, donc seulement avec les conditions (1)–(3), et selon le groupe en question d'autres notations pour l'opération, l'élément neutre et le symétrique peuvent être utilisés; cependant la notation avec '+', '0', et '$-a$' est réservée aux groups commutatifs.) Les conditions (5)–(6) expriment le fait que R muni de '$\times$' et de '1' forme un *monoïde*, dit monoïde multiplicatif de R; dire que R est commutatif revient à dire que son monoïde multiplicatif est commutatif. La distributivité exprime une compatibilité entre le group additif et le monoïde multiplicatif, qui fait tout l'intérêt de la notion d'anneau. Dans un corps K les éléments *non nuls* forment un groupe (noté multiplicativement) appelé le groupe multiplicatif $K^\times$ du corps K.

Un anneau commutatif qui est aussi un corps est appelé un corps commutatif. Les deux conditions d'être un corps ou non et d'être commutatif ou non sont indépendants, et les quatre cas de figure possibles se présentent. Parmi les anneaux qui ne sont pas des corps, l'exemple le plus basique et le plus important est l'anneau $\mathbf{Z}$ des entiers (dits relatifs, en France) qui est commutatif, tout comme les anneaux de polynômes $K[X]$ pour un les corps commutatifs K (et il en existe de très nombreux autres exemples d'anneaux commutatifs qui ne sont pas des corps). Le seul type d'exemple qui nous intéressera d'un anneau qui est ni commutatif ni un corps est $\mathrm{End}(E)$ pour un K-espace vectoriel E (de dimension ≥ 2), oú la multiplication est définie par la composition d'endomorphismes (mais il faut admettre que la seule structure d'anneau est insuffisante pour bien comprendre $\mathrm{End}(E)$). Parmi les corps, c'est aussi le cas commutatif nous intéresse le plus (on a vu les corps commutatifs $\mathbf{Q}$, $\mathbf{R}$, $\mathbf{C}$, et on verra bientôt d'autres exemples), mais il faut au moins mentionner le corps non-commutatif $\mathbf{H}$ des quaternions de Hamilton, une extension des nombres complexes qui est de dimension 4 (d'où le nom) comme $\mathbf{R}$-espace vectoriel.

Certaines règles de calcul habituelles ne se trouvent pas parmi les axiomes, mais sont néanmoins valables dans tous les anneaux car elles peuvent être déduites des axiomes ; notamment la multiplication (à gauche ou à droite) par 0 donne toujours 0. En fait toutes les règles habituelles qui ont la forme d'une égalité, comme par exemple $(x+y)(x-y) = x^2 - y^2$ (ou x^2 est une abréviation pour xx) sont valables dans tout anneau commutatif. Dans ce cours on utilisera ces propriétés sans les détailler davantage.

3.2. *Structures quotient.*

Cette section a pour but de donner une motivation et introduction aux structures quotients, pour ceux qui n'en ont pas vues auparavant, ou qui s'interrogent sur la démarche. Elle ne fait pas partie de la matière propre de ce cours, pour lequel la connaissance des anneaux quotients de $\mathbf{Z}$ et de $K[X]$, décrits explicitement plus après, suffit.

Une méthode importante pour définir de nouvelles structures mathématiques à partir de structures déjà connues est la formations de structures quotient, dont on va donner comme premier exemple l'anneau $\mathbf{Z}/n\mathbf{Z}$ des entiers modulo un entier n. Une telle définition consiste d'une part à construire les éléments de la nouvelle structure, et d'autre part de munir cette ensemble avec des opérations. Le point essentiel d'une structure quotient est que ses éléments ne sont pas des éléments de l'ancienne structure (comme il est le cas pour des sous-structures) mais des *classes* de tels éléments. Par exemple on définir à partir de l'anneau $\mathbf{Z}$ des entiers une nouvelle structure qui n'a que deux éléments, à savoir les classes des nombres pairs respectivement des nombres impairs. Ensuite des opérations sur cet ensemble de classes seront définies, le plus souvent en utilisant les opérations de la structure de départ (c'est l'intérêt de la construction), mais dans ce cas il sera nécessaire de vérifier que ces opérations sont compatibles avec la partition en classes (par exemple on pourra définir une addition sur les l'ensemble des deux classes de parité, car la parité d'une somme $x+y$ d'entiers est déterminé si on connaît les parités de x et de y).

Un ensemble quotient peut être formé dès qu'on a une *partition* d'un ensemble U, qui est par définition un ensemble $\{ C_i \mid i \in I \}$ de parties C_i de U, appelées classes, tel que (1) aucune classe n'est vide, (2) la réunion $\bigcup_{i\in I} C_i$ des classes est égale à U (donc chaque élément de U appartient à une classe), et (3) deux classes distincts sont toujours *disjointes* : elles ont une intersection vide (donc aucun élément de U appartient à plus d'une seule classe). Le plus souvent une telle partition n'est pas définie en donnant directement les classes, mais indirectement pas la spécification d'une *relation d'équivalence* sur U, qui dit pour chaque paire d'éléments si ils appartiennent à la même classe ou non. Si '$\sim$' désigne une relation sur U, il s'agit d'une relation d'équivalence si sont vérifiés pour tout $u, v, w \in U$: (1) $u \sim u$ (réflexivité), (2) si $u \sim v$ alors aussi $v \sim u$ (symétrie), et (3) si $u \sim v$ et $v \sim w$ alors $U \sim w$ (transitivité).

3.2.1. Proposition. *Si '$\sim$' est une relation d'équivalence sur U, alors il existe une partition unique $\{ C_i \mid i \in I \}$ de U telle que deux éléments u, v de U appartiennent à une classe commune si et seulement si $u \sim v$, autrement dit, telle que pour toute classe C_i et tout $u \in C_i$ on ait $\{ v \in U \mid u \sim v \} = C_i$. Réciproquement, si $\{ C_i \mid i \in I \}$ est une partition de U, alors la relation '$\sim$' d'appartenir à une même classe, définie par $u \sim v \iff \exists i \in I : \{u, v\} \subseteq C_i$, est une relation d'équivalence sur U.*

Preuve. Pour la première partie, les classes seront les ensembles d'éléments de la forme "tout ce qui est équivalent à un élément u fixé", autrement dit les ensembles $\{ v \in U \mid u \sim v \}$ pour $u \in U$, et la partition est donc $P = \{ \{ v \in U \mid u \sim v \} \mid u \in U \}$. Attention, les classes qu'on obtient en faisant varier u dans U sont souvent égales les unes aux autres, et on utilise implicitement que dans l'*ensemble* P de classes les occurrences multiples d'une même classe seront réduites à une seule copie de chaque classe. Vérifions les conditions pour avoir une partition : (1) la classe $\{ v \in U \mid u \sim v \}$ contient u par réflexivité, et n'est donc pas vide; (2) chaque $u \in U$ appartient à la classe $\{ v \in U \mid u \sim v \}$, et donc la réunion de toutes les classes, et cette réunion, étant incluse dans U, est donc égale à U; (3) on montrera que si deux classes $C_u = \{ v \in U \mid u \sim v \}$ et $C_w = \{ v \in U \mid w \sim v \}$ ont un élément v en commun, alors elles sont égales (et donc par contraposée, si elles sont distinctes, elle n'ont aucun élément en commun). En fait on montrera que $C_u = C = C_w$ où $C = \{ v' \in U \mid v \sim v' \}$, ce qui suffira, et en plus les deux cas sont identiques après échange de u, w, donc on ne traitera que celui de C_u. Par hypothèse $u \sim v$ et donc par symétrie $v \sim u$. Alors si $v' \in C_u$ on a $u \sim v'$ et donc $v \sim v'$ (transitivité, car $v \sim u$) c'est-à-dire $v' \in C$, et réciproquement si $v' \in C$ on a $v \sim v'$, donc $u \sim v'$ (transitivité, car $u \sim v$) c'est-à-dire $v' \in C_u$; ainsi $C_u = C$.

La seconde partie, pour montrer qu'une partition $\{ C_i \mid i \in I \}$ de U définit une relation d'équivalence sur U est moins fastidieuse. (1) Pour $u \in U$ il existe $i \in I$ tel que $u \in C_i$ car $\bigcup_{i \in I} C_i = U$, et $\{u, u\} \subseteq C_i$ montre que $u \sim u$. (2) La condition $u \sim v$ veut dire que $\exists i \in I : \{u, v\} \subseteq C_i$, et c'est *la même chose* que $\exists i \in I : \{v, u\} \subseteq C_i$ donc $v \sim u$. (3) Si $u \sim v$ et $v \sim w$ il existe i, j tels que $\{u, v\} \subseteq C_i$ et $\{v, w\} \subseteq C_i$, et les classes C_i, C_j n'étant pas disjointes (elles ont l'élément v en commun) sont égales, donc on a $\{u, w\} \subseteq \{u, v, w\} \subseteq C_i = C_j$, et $u \sim w$. □

Un exemple bien connu (même s'il n'est pas souvent présenté explicitement en tant que tel) où dans une définition on se sert d'un ensemble quotient, formé à l'aide d'une relation d'équivalence, est celle des nombres rationnels $\mathbf{Q}$. L'ensemble des nombres rationnels est formé à partir de celui $\mathbf{Z} \times \mathbf{Z}_{>0}$ des couples d'un entier p (le numérateur) et un entier strictement positif q (le dénominateur), mais de tels couple ne donnent pas tous des nombres rationnels distincts ; on forme donc la partition de $\mathbf{Z} \times \mathbf{Z}_{>0}$ en classes selon la relation $(p, q) \sim (r, s) \iff ps = qr$ qui exprime qu'on veut que $\frac{p}{q}$ et $\frac{r}{s}$ désignent le même nombre rationnel, et par définition $\mathbf{Q}$ est l'ensemble de ces classes (on identifie donc un nombre rationnel avec la classe de *tous* les couples (p, q) qui peuvent le représenter). Le fait que les classes sont grandes et compliquées n'est pas gênant, car dans la pratique on les désigne non pas en mentionnant tous leurs éléments, mais juste un élément librement choisi, dit représentant de la classe (les propriétés d'une partition garantissent que toute la classe peut être reconstruite étant donné n'importe quel représentant).

Le fait de travailler avec des représentants a de grands avantages (et est même indispensable comme le montre l'exemple de $\mathbf{Q}$), mais vient avec une responsabilité ; chaque fois qu'on définit une opération (fonction, relation) sur l'ensemble quotient en se servant de représentants, il faut prouver que le résultat de la définition ne dépend pas du choix de ces représentants, car si c'était le cas on aurait donné deux définitions contradictoires pour une même classe. Il est instructif de prouver pour $\mathbf{Q}$, d'abord que la relation ci-dessus est une relation d'équivalence, mais surtout que les formules habituelles définissant les opérations arithmétiques sont telles que le résultat (en tant que nombre rationnel, donc comme classe de couples!) ne dépend pas des choix des représentants de leurs opérandes. On peut d'ailleurs observer que la formation d'un ensemble quotient est essentiel pour définir le corps $\mathbf{Q}$: ces formules *ne définissent pas* une structure d'anneaux sur $\mathbf{Z} \times \mathbf{Z}_{>0}$, car les axiomes ne sont vérifiés que pour l'ensemble quotient.

Même si $\mathbf{Q}$ est basé sur un ensemble quotient, ce n'est pas une *structure quotient*, car la structure (de corps) de $\mathbf{Q}$ n'est pas déduite d'une structure similaire sur l'ensemble $\mathbf{Z} \times \mathbf{Z}_{>0}$ de base. Mais souvent le but de former un ensemble quotient de U est justement de transformer une structure de U en une structure similaire sur l'ensemble quotient. La possibilité de procéder ainsi dépend de propriétés du quotient, comme en montrera par l'exemple simple d'un quotient du groupe additif des nombres réels.

On imagine donc une partition de $\mathbf{R}$ en classes telle qu'on puisse additionner des classes en prenant la classe de la somme de deux représentants. Pour que cette addition de classes soit bien définie, il faut que cette *classe* (de la somme) ne dépende pas des représentants choisis. Donc en particulier si $x_1 \sim x_2$ on doit avoir pour tout y que $x_1 + y \sim x_2 + y$. Il en découle que la condition pour deux nombres réels d'être équivalents ou non ne dépend que de leur différence : on a $x_1 \sim x_2$ si et seulement si $x_1 - x_2 \sim 0$. Toute la partition sera donc connue une fois qu'on connaît la classe $N \subseteq \mathbf{R}$ de 0. On voit facilement que cette classe doit être fermée pour l'opposé ($x \in N \Rightarrow -x \in N$) et pour l'addition ($x, y \in N \Rightarrow x + y \in N$), autrement dit que c'est un sous-groupe du group additif de $\mathbf{R}$. Un exemple d'un tel sous-groupe est l'ensemble de tous les multiples *entiers* d'un nombre $m \in \mathbf{R}$ donné, c'est-à-dire l'ensemble $\{\, mk \mid k \in \mathbf{Z} \,\}$ qui est noté $m\mathbf{Z}$. Pour un tel $N = m\mathbf{Z}$, l'ensemble quotient correspondant, dit de classes modulo N ou modulo m, est noté $\mathbf{R}/N = \mathbf{R}/m\mathbf{Z}$, et il consiste des classes de la forme $x+N = x+m\mathbf{Z} = \{\, x + mk \mid k \in \mathbf{Z} \,\}$ pour $x \in \mathbf{R}$. On peut effectivement définir une structure de groupe sur $\mathbf{R}/N$ en opérant sur les représentants, c'est-à-dire tel qu'on ait $(x + N) + (y + N) = (x + y) + N$ et $-(xN) = (-x)N$ pour tout $x, y \in \mathbf{R}$ (et l'élément neutre de $\mathbf{R}/N$ est la classe $0 + N = N$).

L'exemple le plus connu d'un tel groupe quotient est pour $m = 2\pi$, où le groupe $\mathbf{R}/2\pi\mathbf{Z}$ modélise les angles de rotation dans un plan orienté : ces angles peuvent être représentés par des nombres réels et additionnés (en additionnant ces nombres) pour donner un nouvel angle bien défini, mais l'angle représenté par $\alpha \in \mathbf{R}$ est *identique* à ceux représentés par $\alpha + 2\pi$, $\alpha - 2\pi$, $\alpha + 4\pi$ etc. Cet exemple peut aussi illustrer la raison pour lequel on insiste sur la compatibilité des opérations avec la partition de l'ensemble. Si on peut additionner les angles, et donc aussi les doubler ou tripler (par addition itérée), on ne saura pas multiplier les angles par $\frac{1}{2}$ ou $\frac{1}{3}$ ou par un autre nombre non entier. Ce n'est pas que ces dernières opérations ne peuvent pas être effectuées sur des nombres réels représentant les angles, mais en multipliant deux représentants α et $\alpha+2\pi$ du *même* angle par t, on trouve respectivement $t\alpha$ et $t\alpha+2t\pi$, quels nombres représentent des angles *différents*, sauf dans le cas où $t \in \mathbf{Z}$. Il est donc a *fortiori* vain de chercher à multiplier des angles (éléments de $\mathbf{R}/2\pi\mathbf{Z}$) entre eux ; non seulement cela n'aurait pas de sens géométriquement, mais la définition d'une telle opération est mathématiquement impossible.

3.3. *L'anneau* $\mathbf{Z}$ *et ses quotients.*

Si A est un groupe commutatif quelconque et N un sous-groupe, alors on peut former l'ensemble quotient A/N et le munir d'une structure de groupe commutatif déduite de celle de A, c'est-à-dire vérifiant $(x + N) + (y + N) = (x + y) + N$ pour tout $x, y \in A$, exactement comme pour le cas $A = \mathbf{R}$ mentionné ci-dessus. Pour le cas $A = \mathbf{Z}$ une particularité se produit, à savoir que l'opération de multiplication définie dans $\mathbf{Z}$ est *aussi* compatible avec la formation du quotient, quel ensemble devient donc un anneau commutatif (quotient de $\mathbf{Z}$), et dans certain cas même un corps (que l'anneau de départ $\mathbf{Z}$ n'est pas!).

Si un sous-groupe S du group additif $\mathbf{Z}$ contient un nombre n, il contiendra aussi $-n$ et $2n$, et tous les multiples de n (ceci est aussi vrai dans $\mathbf{R}$ ou dans tout groupe noté additivement), c'est-à-dire le sous-ensemble $n\mathbf{Z}$ de $\mathbf{Z}$. Si S n'est pas réduit à $\{0\}$, il contiendra un plus petit élément strictement positif, qu'on appellera d, et on aura donc $d\mathbf{Z} \subseteq S$. Les multiples de d sont à distance d de leurs voisins (dans $d\mathbf{Z}$). S'il y avait un élément $s \in S$ non multiple de d, alors sa différence $h = s - dk$ avec le précédent multiple dk de d vérifierait $h \in S$ et $0 < h < d$, contredisant le choix de d, donc de tels s n'existent pas ; par conséquent $S = d\mathbf{Z}$. Ainsi on connaît *tous* les sous-groupes additifs de $\mathbf{Z}$: ce sont $\{0\}$, ainsi que les sous-groupes $n\mathbf{Z}$ pour $n = 1, 2, 3, \ldots$ (ou $1\mathbf{Z}$ est en fait égal à $\mathbf{Z}$ tout entier). Le sous-groupe trivial $\{0\}$ n'est pas intéressant pour former un quotient, car chaque classe du quotient sera réduit à un seul élément et on retrouvera le groupe $\mathbf{Z}$ (sous une forme très légèrement différente : la classe $\{k\}$ remplace k).

Calcul modulo n.

Soit $n > 0$ un entier, alors l'ensemble quotient $\mathbf{Z}/n\mathbf{Z} = \{\, i + n\mathbf{Z} \mid 0 \leq i < n \,\}$ (dont les éléments $i + n\mathbf{Z}$ sont appelées classes *modulo* n) est muni d'une addition vérifiant $(i + n\mathbf{Z}) + (j + n\mathbf{Z}) = (i + j) + n\mathbf{Z}$, ce qu'on peut écrire de façon plus transparente $\overline{i} + \overline{j} = \overline{i+j}$ si l'on note la classe $i + n\mathbf{Z}$ de i par $\overline{i}$. Cette structure modélise plus ou moins certains phénomènes cycliques du quotidien, comme par exemple l'horloge qui correspond (pour les heures entières) au calcul modulo 24, ou des compteurs finis (comme l'odomètre dans une voiture, ou un compteur d'eau ou d'électricité) qui, faute de pouvoir afficher des

nombres arbitrairement grands, reviennent à zéro après avoir atteint leur plus grande valeur possible (on remarque que les "entiers machine" dans les ordinateurs sont en fait des valeurs dans $\mathbf{Z}/n\mathbf{Z}$ ou n est un grand entier, en général une puissance de 2 comme $2^{32} = 4\,294\,967\,296$).

Ce qui rend $\mathbf{Z}/n\mathbf{Z}$ particulièrement intéressant mathématiquement, est que le quotient est aussi compatible avec la multiplication dans $\mathbf{Z}$: on peut munir $\mathbf{Z}/n\mathbf{Z}$ d'une multiplication qui vérifie $\overline{i} \times \overline{j} = \overline{ij}$, où comme ci-dessus $\overline{i}$ désigne la classe $i + n\mathbf{Z}$. Cela en fait ne marche que pour les quotients obtenus à partir de $\mathbf{Z}$, et non pas pour ceux basés sur $\mathbf{R}$ ou même sur $\mathbf{Q}$ (comme indiqué dans la section précédente). Pour vérifier qu'une telle multiplication existe dans $\mathbf{Z}/n\mathbf{Z}$, il faut montrer que la classe produit $\overline{i} \times \overline{j}$, qui doit être la classe $\overline{ij}$ du produit ij, est aussi la classe de tout autre produit de représentants des classes $\overline{i}$ et $\overline{j}$ (car en fait i et j ne sont que des représentants quelconques de leurs classes respectives, même si la notation $\overline{i}, \overline{j}$ les met en avant). D'autres représentants seront en général de la forme $i' = i + nk$ et $j' = j + nl$ avec $k, l \in \mathbf{Z}$, et pour ces représentants on a $i'j' = ij + nkj + nil + n^2kl = ij + n(kj + il + nkl)$, ce qui est un élément de la classe $\overline{ij} = ij + n\mathbf{Z}$, donc le produit $i'j'$ détermine bien la même classe modulo n que ij (on peut observer que dans cet argument il est essentiel non seulement que k, l, n soient entiers, mais aussi les représentants i, j eux-mêmes, d'où l'importance de "partir" de $\mathbf{Z}$ et non pas de $\mathbf{R}$ ou $\mathbf{Q}$).

Muni ainsi d'une addition et d'une multiplication, avec des éléments neutres $\overline{0}$ respectivement $\overline{1}$, on vérifie facilement que $\mathbf{Z}/n\mathbf{Z}$ est un anneau commutatif, quel que soit $n > 0$. En fait chacune des conditions (1)–(8) de la définition 3.1.1 pour $\mathbf{Z}/n\mathbf{Z}$ est une conséquence directe de la condition correspondante de $\mathbf{Z}$, par exemple (distributivité à gauche) $\overline{i}(\overline{j} + \overline{k}) = \overline{i}\overline{(j+k)} = \overline{i(j+k)} = \overline{ij + ik} = \overline{i}\,\overline{j} + \overline{i}\,\overline{k}$, pour tout $i, j, k \in \mathbf{Z}$. Le fait que les axiomes de 3.1.1 parlent d'éléments 0, 1, pendant que dans $\mathbf{Z}/n\mathbf{Z}$ ce sont les classes $\overline{0}$ et $\overline{1}$ de ces nombres peut prêter à confusion, mais on convient en général qu'en parlant d'un entier k comme élément de $\mathbf{Z}/n\mathbf{Z}$, on veut toujours dire la classe $\overline{k} = k + n\mathbf{Z}$. Ainsi on pourra dire, un peu paradoxalement, que $n = 0$ est vrai dans $\mathbf{Z}/n\mathbf{Z}$, mais faux dans $\mathbf{Z}$. L'anneau $\mathbf{Z}/n\mathbf{Z}$ a précisément n éléments, à savoir les classes modulo n de 0, 1, ..., $n-1$ (c'est facile à voir, mais on donnera néanmoins une preuve formelle plus tard). En particulier $\mathbf{Z}/1\mathbf{Z}$ n'a qu'un seul élément (qui est à la fois 0 et 1 dans cet anneau) ; c'est un anneau pas très intéressant appelé l'anneau trivial.

L'anneau trivial ne vérifie pas la première partie $1 \neq 0$ de la condition (9) de la définition 3.1.1, et n'est donc pas un corps. En revanche l'anneau $\mathbf{Z}/2\mathbf{Z}$, qui a $\overline{0}$ et $\overline{1}$ comme ses deux seuls éléments, vérifie la condition (9), et est donc un corps commutatif. Ici il suffisait de vérifier que dans $\mathbf{Z}/2\mathbf{Z}$ l'élément 1 est son propre inverse (c'est-à-dire symétrique multiplicatif), ce qui est vrai dans n'importe quel anneau. Il est un peu plus surprenant que $\mathbf{Z}/3\mathbf{Z}$ est aussi un corps, car à part $\overline{1} \times \overline{1} = \overline{1}$ on a aussi $\overline{2} \times \overline{2} = \overline{4} = \overline{1}$ dans $\mathbf{Z}/3\mathbf{Z}$, et $\overline{2}$ est aussi son propre inverse. Par contre $\mathbf{Z}/4\mathbf{Z}$ n'est pas un corps, car le résultat d'une multiplication par 2 dans $\mathbf{Z}/4\mathbf{Z}$ est toujours $\overline{0}$ ou $\overline{2}$, et jamais $\overline{1}$; l'élément 2 dans $\mathbf{Z}/4\mathbf{Z}$, qui n'est pas nul (contrairement à celui dans $\mathbf{Z}/2\mathbf{Z}$), n'a donc pas d'inverse. Ce qui est en jeu ici est que $2 \times 2 = 0$ dans $\mathbf{Z}/4\mathbf{Z}$, une multiplication de deux éléments non nuls qui produit l'élément 0. Une telle situation peut se produire dans certains anneaux, et dans ce cas ces éléments sont appelés des diviseurs de zéro ; un diviseur de zéro n'est jamais inversible car si $xy = 0$ avec $x \neq 0$ et $y \neq 0$ et si $zx = 1$, on aurait $0 = z0 = z(xy) = (zx)y = 1y = y$, une contradiction. Un corps, où tous les éléments non nuls ont un inverse, ne peut donc pas avoir de diviseurs de zéro. La réciproque est fausse, comme le montre l'exemple de $\mathbf{Z}$, qui n'a pas de diviseurs de zéro mais qui n'est certainement pas un corps. Un anneau commutatif qui n'a pas de diviseurs de zéro est appelé une *anneau intègre* ; c'est une importante catégorie d'anneaux avec de "meilleurs" propriétés que les anneaux commutatifs en général, mais néanmoins une catégorie beaucoup plus large que celle des corps commutatifs.

Si n est un nombre composé, disons $n = kl$ avec $k, l > 1$, alors (les classes de) k et l sont des diviseurs de zéro dans $\mathbf{Z}/n\mathbf{Z}$, et $\mathbf{Z}/n\mathbf{Z}$ n'est pas un anneau intègre (et certainement pas un corps). Dans le cas contraire (n est un *nombre premier*), on verra que $\mathbf{Z}/n\mathbf{Z}$ n'a aucun diviseur de zéro, et est donc un anneau intègre. Plus fort encore, $\mathbf{Z}/n\mathbf{Z}$ sera un corps, par le raisonnement suivant : pour $k \neq 0$ dans $\mathbf{Z}/n\mathbf{Z}$ l'opération $i \mapsto ik$ de multiplication par k est une application $\mathbf{Z}/n\mathbf{Z} \to \mathbf{Z}/n\mathbf{Z}$; elle est injective car $ij = ik$ entraîne $i(j-k) = 0$ et donc $j - k = 0$ car i n'est pas diviseur de zéro ; or une application injective d'un ensemble *fini* vers lui-même est toujours bijective, donc en particulier il existe i avec $ik = 1$ dans $\mathbf{Z}/n\mathbf{Z}$, et i est un inverse de k. On a donc une dichotomie pour les entiers $n \geq 2$: si n est composé, $\mathbf{Z}/n\mathbf{Z}$ possède des diviseurs de zéro ; sinon n est un nombre premier, et $\mathbf{Z}/n\mathbf{Z}$ est un corps commutatif.

La construction de $\mathbf{Z}/n\mathbf{Z}$ nous fournit une infinité de corps qu'on avait pas encore considérés, un pour chaque nombre premier (dont il est connu depuis l'antiquité qu'ils forment une collection infinie). Ces corps, tous finis, sont intéressants par le fait que leurs opérations peuvent être réalisés de façon exacte et très efficace ; en contraste, dans des corps infinis il est soit impossible de calculer de manière exacte ($\mathbf{R}$, $\mathbf{C}$), soit les opérations exactes peuvent prendre un temps non borné (c'est la cas dans $\mathbf{Q}$ quand numérateur et dénominateur deviennent très grands). En informatique on s'intéresse particulièrement au corps $\mathbf{Z}/2\mathbf{Z}$ et aux espaces vectoriels sur ce corps ; on remarque que les seuls multiplications scalaires seront alors par 0 et par 1, ce qui ne semble guère intéressant, mais le fait que tout vecteur v dans un tel espace doit vérifier $v + v = 2v = 0v = \vec{0}$ rend ces espaces vectoriels assez particuliers.

Mais la plus grand importance pour nous de la construction de $\mathbf{Z}/n\mathbf{Z}$ est le fait qu'elle fournit des informations sur $\mathbf{Z}$, et qu'elle est un modèle par laquelle on comprendra mieux la formation des quotients des anneaux de polynômes sur un corps, quels anneaux ont des propriétés très similaires à celles de $\mathbf{Z}$.

Division euclidienne.

Une opération clé pour comprendre la structure algébrique de $\mathbf{Z}$ est la division avec reste, appelée division euclidienne. C'est une opération fort simple, mais en vue de son rôle crucial il est utile de formuler et de prouver la propriété qui caractérise cette division de façon très précise.

3.3.1. Proposition. *Soit $a \in \mathbf{Z}$ et $n \in \mathbf{Z} \setminus \{0\}$. Alors il existe $q \in \mathbf{Z}$ et $r \in \{0, 1, \ldots, |n| - 1\}$ tels que $a = qn + r$, et le couple (q, r) est unique.*

Preuve. Comme (q, r) est une solution pour (a, n) si et seulement si $(-q, r)$ est une solution pour $(a, -n)$, il suffit de considérer le cas $n > 0$. Alors la condition $a = qn + r$ avec $0 \leq r < n$ veut dire que $qn \leq a < qn + n = (q + 1)n$, et réciproquement si $qn \leq a < (q + 1)n$ alors $r = a - qn$ vérifiera les conditions. Il suffit donc de montrer qu'il existe un q unique avec $qn \leq a < (q + 1)n$. En vue de la nature croissante de la suite doublement infinie $(kn)_{k \in \mathbf{Z}}$ (c'est-à-dire le fait que $kn < ln$ si $k < l$) cela sera clair dès que cette suite contient aussi bien des termes $\leq a$ que de termes $> a$: on pourra et devra alors prendre pour q le plus grand nombre tel que $qn \leq a$. Mais par exemple $-|a|n \leq -|a| \leq a$ et $(|a|+1)n > |a| \geq a$ montrent l'existence de tels termes, ce qui complète la preuve. □

3.4. *Anneaux de polynômes.*

On considéra maintenant l'anneau $K[X]$ des polynômes en X et à coefficients dans K. On connaît les polynômes d'abord dans le manipulation des expressions contenant une inconnue x (pour la classe expressions formées en utilisant uniquement des opérations additives et la multiplication), puis dans l'étude des fonction polynomiales, où x n'est pas une inconnue mais désigne l'argument de ces fonctions. Pour une considération algébrique des polynômes, le statut de x change en celui d'un élément comme les autres, qui ne cache pas une valeur inconnue ou variable ; on l'appelle *indéterminée*, et on l'écrit en majuscule pour marquer son statut. Si $x^2 = 3x + 1$ est une équation qui peut être vérifiée pour certains valeur concrètes de x, l'équation $X^2 = 3X + 1$ est tout simplement fausse dans $K[X]$.

Sans définir formellement $K[X]$ (ce n'est pas d'une grande importance pour ce cours), on peut le caractériser par les propriétés suivantes.

3.4.1. Caractérisation. *Pour un corps K et un symbole X, l'ensemble $K[X]$ vérifie*

(1) $K[X]$ contient le corps K (les constantes dans $K[X]$) ainsi qu'un autre élément désigné par X;
(2) $K[X]$ est un anneau commutatif, avec $0, 1 \in K$;
(3) Deux expressions formées en combinant des constantes et des copies de X avec les opérations d'un anneau ne donnent le même élément de $K[X]$ que si leur égalité est une conséquence des axiomes d'un anneau commutatif.

La deuxième condition implique l'existence d'une addition et multiplication de polynômes dont le résultat est toujours un polynôme. La troisième condition assure que X ne vérifie aucune propriété particulière ; notamment on pourra substituer, dans une identité de polynômes en X, n'importe quelle constante (dans K) pour X, et d'obtenir ainsi une égalité valable dans K (on précisera ceci plus tard).

Les axiomes d'un anneau commutatif (notamment la distributivité) permettent de transformer toute expression polynomiale en X (comme spécifié dans (3)) en une somme de produits. Dans chacun de ces produits on pourra regrouper (par associativité et commutativité de la multiplication) et multiplier ensemble toutes les constantes, pour obtenir une seule constante dans K (s'il n'y avait pas de constantes du tout dans le produit on prendra la constante $1 \in K$), et tous les autres facteurs du produit seront des copies de X, dont le produit sera noté X^i s'il y en avait i (là encore le produit X^0 d'aucune copie de X est identifié à la constante 1). Ces produits X^i sont appelés *monômes* en X. Ainsi on a transformé une expression polynomiale quelconque en une somme (finie) de termes, chacun étant le produit d'une constante et un monôme en X. Si nécessaire on pourra combiner (par distributivité) des termes ayant le même monôme (les termes similaires) pour obtenir une *combinaison K-linéaire* des monômes $X^0 = 1$, $X^1 = X$, $X^2, X^3, \ldots$ (la liste des monômes est infinie, mais par définition une combinaison linéaire ne peut avoir qu'un nombre fini de termes). Le langage d'espaces vectoriels est justifié : le fait que $K[X]$ est un anneau qui contient K assure l'existence d'une multiplication scalaire, et les propriétés nécessaires pour un espace vectoriel. Aucune combinaison linéaire non triviale des X^i donne $0 \in K[X]$, donc ces monômes forment une famille libre (infinie) dans $K[X]$, et en fait une base. On obtient ainsi une caractérisation de $K[X]$ plus concrète que la précédente : elle décrit une représentation explicite pour chaque polynôme.

3.4.2. Caractérisation. *$K[X]$ est anneau commutatif contenant le corps K, et vérifie :*
(1) $K[X]$ contient des monômes X^i pour $i \in \mathbf{N}$;
(2) pour la structure de K-espace vectoriel de $K[X]$, la famille $(X^i)_{i\in\mathbf{N}}$ est une base de $K[X]$;
(3) $X^iX^j = X^{i+j}$ pour tout $i, j \in \mathbf{N}$.

Condition (3) et le fait que tout polynôme est combinaison linéaire de monômes entraîne que X^0 est élément neutre de la multiplication, et donc identique à la constante $1 \in K$ de $K[X]$. Concrètement chaque $P \in K[X]$ s'écrit sous la forme $P = \sum_{i=0}^{d} p_i X^i$ avec $p_i \in K$ pour tout i. Dans cette écriture les p_i sont appelés les coefficients de P. Ils peuvent être nuls, ce qui nous permet d'utiliser toujours une ensemble contigu $1, X, X^2, \ldots, X^d$ de monômes ; la présence d'une borne d à la sommation assure que l'écriture est toujours une somme finie. On peut toujours augmenter d en rajoutant des termes à coefficient nul, ce qui s'avère souvent pratique ; ainsi cette écriture d'un polynôme n'est pas totalement unique, mais c'est aussi la seule liberté qu'on a dans cette écriture. On déduit facilement de 3.4.2 :

$$\left(\sum_{i=0}^{d} p_i X^i\right) + \left(\sum_{i=0}^{d} q_i X^i\right) = \sum_{i=0}^{d} (p_i + q_i) X^i \tag{16}$$

$$\left(\sum_{i=0}^{d_1} p_i X^i\right) \times \left(\sum_{j=0}^{d_2} q_j X^j\right) = \sum_{i=0}^{d_1}\sum_{j=0}^{d_2} p_i q_j X^{i+j} = \sum_{k=0}^{d_1+d_2} \left(\sum_{i+j=k} p_i q_j\right) X^k \tag{17}$$

3.4.3. 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* $\deg(0) = -\infty$.

Dans l'écriture d'un polynôme on peut arrêter la sommation au degré du polynôme (c'est-à-dire prendre $d = \deg(P)$) mais on n'y est pas obligé ; c'est la raison pour laquelle la définition de degré prend le plus grand indice dont le coefficient est *non nul*, l'indice avec cette propriété ne dépend pas de l'écriture choisie. Si $P = \sum_{i=0}^{N} p_i X^i \in K[X]$ est non nul et $d = \deg(P)$, on appellera le terme $p_d X^d$ le *terme dominant*, et p_d le *coefficient dominant* de P (ils sont non nuls par définition de $\deg(P)$). Un *polynôme unitaire* est un polynôme non nul dont le coefficient dominant est 1.

3.4.4. 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) $\deg(PQ) = \deg(P) + \deg(Q)$.

Preuve. La partie (1) découle de (16) où on peut prendre $d = \max(\deg(P), \deg(Q))$; comme $p_d \neq 0$ ou $q_d \neq 0$, on ne peut avoir $p_d + q_d = 0$ que si $p_d = -q_d \neq 0$, et donc $\deg(P) = \deg(Q)$. La partie (2) découle de (17), car la somme donnant le coefficient dominant du produit, $\sum_{i+j=\deg(P)+\deg(Q)} p_i q_j$, est réduite à $p_{\deg(P)} q_{\deg(Q)}$, le produit non nul (car $p_{\deg(P)} \neq 0$ et $q_{\deg(Q)} \neq 0$) des coefficients dominants. □

La valeur $\deg(0) = -\infty$ a été choisi pour rendre cet énoncé valable quand l'un ou l'autre des polynômes est nul, avec des conventions évidentes pour $-\infty$ (notamment que $-\infty + d = -\infty$ pour $d \in \mathbf{N} \cup \{-\infty\}$). Une conséquence du second point est que $K[X]$ est un anneau intègre : un produit de polynômes ne saurait être nul (de degré $-\infty$) que si l'un au moins des polynômes est nul. Il est aussi clair que pour un polynôme $P \neq 0$, tout multiple non nul de P (par un polynôme) est de degré au moins $\deg(P)$. On note l'ensemble $\{ PQ \mid Q \in K[X] \}$ des multiples de P par $P\,K[X]$, à l'analogie des multiplies $n\mathbf{Z}$ de n dans $\mathbf{Z}$. Aussi comme dans $\mathbf{Z}$ on dira que Q et Q' sont congruents modulo P si $Q - Q' \in P\,K[X]$ (c'est une relation d'équivalence), et on note $Q + P\,K[X]$ la classe de tous les polynômes congruents à Q modulo P. Ces notions sont utiles pour comprendre la division euclidienne dans $K[X]$:

3.4.5. Proposition. *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 $A = QB + R$, et le couple (Q, R) est unique.*

Preuve. On montre d'abord qu'une solution (Q, R) sera forcément unique. Pour une telle solution, R et A sont congruents modulo B, et donc les restes R dans deux solutions le seront aussi. Mais on a $\deg(R) < \deg(B)$, et aucun autre polynôme congruent à R peut être aussi de degré $< \deg(B)$: c'est la somme de R et d'un multiple non nul M de B, donc $\deg(M) \geq \deg(B) > \deg(R)$, et d'après la proposition 3.4.4(1), $\deg(R + M) = \deg(M) \geq \deg(B)$. Ainsi R sera unique, et comme $K[X]$ est intègre, Q sera unique aussi ($QB = Q'B$ implique $Q = Q'$). Pour l'existence de (Q, R), on prend pour R un élément de plus petit degré possible dans la classe de congruence $A + B\,K[X]$, qui admet donc par définition un Q tel que $A - QB = R$; il suffira de vérifier que $\deg(R) < \deg(B)$. Posons $r = \deg(R)$ et $b = \deg(B)$, et supposons $r \geq b$ pour en tirer une contradiction. Le terme principal de R est de la forme cX^r, et celui de B de la forme $c'X^b$. Le polynôme $R' = R - \frac{c}{c'}X^{r-b}B$ vérifie alors $\deg(R') \leq r$ d'après la proposition 3.4.4(1), et on a fait en sorte que le coefficient de X^r dans R' soit nul, donc en fait $\deg(R') < r$. Comme on a aussi $R' \in R + B\,K[X] = A + B\,K[X]$, ce R' contredit la minimalité supposée de $\deg(R)$. □

La preuve ne donne pas explicitement un algorithme pour trouver (Q, R), car prendre un élément de plus petit degré dans la classe $A + B\,K[X]$, qui est un ensemble infini, n'est pas une opération effective. Ceci dit, la preuve montre comment on peut procéder : en utilisant un polynôme *variable* R, qui sera toujours un représentant de la classe $A + B\,K[X]$, on initialise $R := A$, et tant que $\deg(R) \geq \deg(B)$ on replace R par un représentant de plus petit degré, comme indiqué dans la transition de R vers R' dans la preuve. Comme le degré baisse chaque fois d'au moins 1, on obtiendra $\deg(R) < \deg(B)$ après au plus $\deg(A) - \deg(B) + 1$ itérations, et la valeur finale de R sera le polynôme reste cherché ; le quotient Q est la somme des facteurs $\frac{c}{c'}X^{r-b}$ dans les multiples de B qui ont été soustraits dans les différentes étapes.

3.5. *Substitution dans $K[X]$, racines, idéaux de $K[X]$.*

Une des propriétés fondamentales des polynômes est la possibilité de remplacer X par une constante a quelconque, et d'évaluer l'expression obtenue pour trouver une valeur dans K (on dit qu'on évalue le polynôme en a). En fait, c'est de cette manière qu'on utilise les polynômes dans les équations et fonctions polynomiales. Plus précisément ce n'est pas tellement la possibilité d'évaluer en a qui est importante, mais le fait que ceci est compatible avec les opérations d'anneau : l'évaluation en a d'une somme ou d'un produit de polynômes P, Q donne la somme respectivement le produit de valeurs obtenues en évaluant P et Q individuellement en a. On appelle pour cette raison l'opération d'évaluation $K[X] \to K$ un *homomorphisme* d'anneaux (le terme homomorphisme est utilisé dans beaucoup d'autres contextes, pour des applications qui sont compatibles avec une certaine structure, ici celle d'un anneau commutatif).

La substitution de a pour X, que l'auteur de ce cours aime noter (avec une notation emprunté aux informaticiens) "$X := a$" prononcé "X devient a", produit évidemment parfois le même résultat pour différents polynômes, et définit ainsi une relation d'équivalence sur $K[X]$: "donnent la même valeur quand on met $X := a$". Chaque polynôme $P \in K[X]$ est équivalent pour cette relation à précisément un polynôme constant, à savoir la constante dont la valeur est justement le résultat de substituer $X := a$ dans P ; on notera ce résultat $P[X := a]$ (mais le plus souvent on le voit noté simplement $P(a)$, une notation non sans problèmes qui gomme la distinction entre un polynôme et sa fonction polynomiale).

D'après la proposition 3.4.5, on peut écrire $P = (X-a)Q + R$ avec $\deg(R) \leq 0$, autrement dit R est une constante ; en appliquant $X := a$ à cette égalité on obtient $P[X := a] = R$: la valeur du reste n'est autre que l'évaluation en a de P. (L'étude du calcul du reste R de cette division par $X - a$ montrera qu'il consiste essentiellement en une façon particulière de substituer (au fur et à mesure) $X := a$ dans P.)

3.5.1. Définition/Proposition. *Un élément $a \in K[X]$ est une racine du polynôme $P \in K[X]$ si $P[X := a] = 0$. Cela est le cas si et seulement si P est un multiple $(X-a)Q$ du polynôme $X - a$.* □

La possibilité de substituer des valeurs pour X n'est pas limité aux seuls éléments du corps de base K. On pourra aussi substituer des valeurs dans une anneau commutatif A, qui contient K (cette dernière condition sert à pouvoir interpréter les coefficients des polynômes). Par exemple pour les polynômes de $\mathbf{R}[X]$, il est possible de prendre $A = \mathbf{C}$ et de substituer l'unité imaginaire $\mathbf{i} \in \mathbf{C}$ pour X, et d'obtenir un nombre complexe comme résultat ; par exemple $(2X^3+7X-4)[X := \mathbf{i}] = 2\mathbf{i}^3+7\mathbf{i}-4 = 5\mathbf{i}-4$. Comme avant, le fait d'avoir la même évaluation définit une relation d'équivalence sur $K[X]$, mais cette fois-ci tout polynôme n'est pas équivalent à un polynôme constant : ce sera le cas seulement si l'évaluation donne un élément de $K \subseteq A$. Pour la substitution $X := \mathbf{i}$ tous les polynômes de degré au plus 1 dans $\mathbf{R}[X]$ seront non équivalents 2 à 2, car $(c_1X + c_0)[X := \mathbf{i}] = c_0 + c_1\mathbf{i}$. Et comme cela épuise (pour $c_0, c_1 \in \mathbf{R}$) tous les nombres complexes, tout polynôme de $\mathbf{R}[X]$ sera équivalent à un (et un seul) de ces polynômes de degré au plus 1. Le polynôme quadratique $X^2 + 1$ est annulé par la substitution $X := \mathbf{i}$, et donc équivalent à la constante 0, et tous ses multiples $(X^2 + 1)Q$ le sont également. Ici encore la division euclidienne montre que ces multiples forment *toute* la classe de polynômes équivalents à 0 : si $P = (X^2 + 1)Q + R$ avec $\deg(R) < 2$, alors R est l'unique polynôme de degré au plus 1 équivalent à P, et si $R = 0$ cela veut dire que P est multiple de $X^2 + 1$. La situation est très similaire pour l'évaluation en un autre nombre complexe non réel : pour la substitution $X := a + b\mathbf{i}$ avec $a, b \in \mathbf{R}$ et $b \neq 0$, les polynômes $c_1X + c_0$ de degré ≤ 1 donnent encore tous des valeurs complexes différentes (à savoir $c_0 + ac_1 + bc_1\mathbf{i}$), il y a un seul polynôme quadratique unitaire annulé par la substitution, qui est cette fois-ci $M = X^2 - 2aX + a^2 + b^2$, tout autre polynôme annulé étant un multiple QM de M, et tout polynôme est équivalent pour cette substitution à son reste (de degré au plus 1) après division par M.

Dans ces exemples d'une substitution définissant une équivalence dans $K[X]$, il y a chaque fois un polynôme unitaire (respectivement $X - a$, $X^2 + 1$, et $X^2 - 2aX + a^2 + b^2$) tel que ses multiples forment l'ensemble des polynômes équivalents à 0, et pour lequel tout polynôme est donc équivalent à son reste après division par ce polynôme. On va voir que cela ne dépend pas des détails de la substitution, ou même du fait que l'équivalence est définie par une substitution, mais est une propriété fondamentale de $K[X]$.

Supposons que $I \subseteq K[X]$ est le sous-ensemble de polynômes équivalents à 0, pour une certaine relation d'équivalence, où plus généralement deux polynômes sont équivalents si leur différence est dans I. Pour que cette équivalence soit compatible avec l'addition il faut que I soit un sous-groupe additif (fermé pour addition et soustraction), et pour être compatible avec la multiplication il faut en plus que tout multiple d'un élément de I reste dans I (car en multipliant 0 par n'importe quelle valeur on obtiendra toujours 0). On appellera I un *idéal* de $K[X]$, une notion qui existe pour tout anneau commutatif.

3.5.2. Définition. *Soit R un anneau commutatif. Une partie $I \subseteq R$ est un idéal de R vérifiant :*
(1) $0 \in I$, et I est fermé pour l'addition et la soustraction (c'est un sous-groupe du groupe additif de R);
(2) pour tout $a \in R$, l'ensemble I est fermé pour la multiplication par a : on a $ax \in I$ dès que $x \in I$.

Pour tout élément $a \in R$, l'ensemble $aR = \{\, ax \mid x \in R \,\}$ des multiples est un idéal, comme on vérifie immédiatement. On appelle ce type d'idéal un *idéal principal* de R, et a est un *générateur* de l'idéal aR. Dans $\mathbf{Z}$, tous les sous-groupes additifs $n\mathbf{Z}$ sont aussi des idéaux de $\mathbf{Z}$; ceci n'est pas étonnant, car la multiplication d'une valeur x par un élément $a \in \mathbf{Z}$ peut être réalisée par l'addition de a copies de x (ou de $-a$ copies de $-x$ si $a < 0$), donc pour $R = \mathbf{Z}$ la deuxième condition de la définition 3.5.2 est une conséquence de la première. Mais dans $K[X]$ elle ne l'est pas : les multiples d'un polynôme P sont bien plus que juste l'ensemble les polynômes obtenus par addition et soustraction de copies de P. D'ailleurs l'idéal principal $PK[X]$ est utilisé tellement souvent qu'on utilise parfois la notation allégée "(P)" pour le désigner, surtout dans des situations où il est clair qu'on parle d'un idéal de $K[X]$.

3.5.3. Proposition. *Pour tout idéal I de $K[X]$ qui n'est pas réduit à $\{0\}$ il existe un polynôme unitaire unique P tel que $I = (P) \overset{\text{déf}}{=} PK[X]$. En particulier tout idéal de $K[X]$ est un idéal principal.*

Preuve. Si $I \neq \{0\}$, soit P' un polynôme non nul de I de degré minimal, c son coefficient dominant, et $P = c^{-1}P$. Alors P est un élément unitaire de I, et de degré minimal parmi les éléments non nuls de I. Par la condition (2) de la définition 3.5.2 on a $(P) \subseteq I$, et d'après la proposition 3.4.4 (2), l'idéal ne contient que P comme polynôme unitaire de degré $\deg(P)$, donc pour pouvoir conclure il suffit de montrer l'inclusion $I \subseteq (P)$. Si $A \in I$ on écrit $A = PQ+R$ avec $\deg(R) < \deg(P)$, alors $R \in A+(P) \subseteq I$, et donc $R = 0$ par la minimalité dans le choix de P. On conclut $A = PQ \in (P)$ et donc $I \subseteq (P)$. □

Même si ce résultat parait un peu plus compliqué et abstrait, c'est l'analogue direct pour $K[X]$ du fait que les sous-groupes de $\mathbf{Z}$ sont tous de la forme $n\mathbf{Z}$ (en convenant que $\{0\} = 0\mathbf{Z}$). Dans les deux cas, c'est une conséquence de l'existence d'une division euclidienne, avec un reste qui est plus petit, dans un sens convenable, que le diviseur. Cette similitude nous permettra de développer la théorie de divisibilité et factorisation pour $\mathbf{Z}$ et de $K[X]$ au même temps. Les résultats sont familiers pour $\mathbf{Z}$, mais valent dans une forme très semblable pour $K[X]$ aussi, c'est tout l'avantage d'une approche un peu abstraite.

En fixant $P \in K[X]$, le fait d'avoir une racine $a \in K$ veut dire que P est multiple de $X - a$. Si c'est le cas pour un certain $a \in K$, donc $P = (X - a)P'$, alors toute *autre* racine a' de P doit aussi être racine de P' : il suffit d'appliquer $X := a'$ ce qui donne (d'après la compatibilité avec la multiplication) $0 = P[X := a'] = ((X - a)[X := a'])P'[X := a'] = (a' - a)P'[X := a']$, et d'observer que $a' - a \neq 0$. Réciproquement toute racine de P' est racine de P. Il n'est pas exclus que la première racine a soit également racine de P' ; dans ce cas on dit qu'elle est *racine multiple* de P. On peut répéter cette opération de mettre en facteur des polynômes $X - a_i$ pour les racines de P', et on trouvera ainsi donc toutes les racines de P dans K. Il reste un facteur qui n'a aucune racine dans K. Dans le cas $K = \mathbf{C}$, ce facteur ne peut être qu'une constante d'après le théorème d'Alembert–Gauss, et comme les autres facteurs sont tous des polynômes unitaires, la valeur de cette constante est le coefficient dominant de P.

On a vu qu'une substitution $X := \alpha$ dans les polynômes de $K[X]$ est définie quand α est dans un anneau commutatif A qui contient K, et elle définit dans ce cas un homomorphisme d'anneaux $K[X] \to A$. La notion de racine d'un polynôme est étendue à cette situation, du moins dans le cas où A est un corps, notamment si $A = \mathbf{C}$ et K est un sous-corps de $\mathbf{C}$, comme $\mathbf{R}$ ou $\mathbf{Q}$. Ainsi on dit que $\mathbf{i}$ et $-\mathbf{i}$ sont racines du polynôme $X^2+1 \in \mathbf{R}[X]$. Pour une telle racine $\alpha \notin K$ de $P \in K[X]$ on ne peut pas dire que P est un multiple dans $K[X]$ de $X-\alpha$, car ce dernier polynôme n'est pas un élément de $K[X]$. Mais l'ensemble de polynômes qui s'annulent quand on met $X := \alpha$ forme un idéal dans $K[X]$, qui d'après la proposition 3.5.3 est formé des multiples d'un polynôme unitaire U qui dépend de α (on l'appelle le polynôme minimal de α dans $K[X]$), a moins que l'idéal ne soit $\{0\}$ (quand aucun polynôme non nul possède α comme racine). Dans le cas particulier $K = \mathbf{R}$ et $\alpha \in \mathbf{C} \setminus \mathbf{R}$, on a vu ci-dessus que ce polynôme U existe toujours, et est quadratique et de discriminant strictement négatif : c'est $X^2 - 2aX + a^2 + b^2 \in \mathbf{R}[X]$ quand $\alpha = a + b\mathbf{i}$. (Ceci n'est pas vrai si $K = \mathbf{Q}$: le polynôme minimal de $\sqrt[3]{2}$ est $X^3 - 2$ qui n'est pas quadratique, et la substitution $X := \pi$ n'annule aucun polynôme non nul dans $\mathbf{Q}[X]$.) Un polynôme P ayant α pour racine s'écrira donc $P = UQ$, mais cette fois-ci les racines de P autres que α ne sont pas forcément des racines de Q, car elles peuvent aussi être racines de U. Dans le cas où $U \in \mathbf{R}[X]$ est quadratique de discriminant négatif, il possède deux racines distinctes, qui sont des conjugués complexes. Un polynôme réel sans racines réelles est donc le produit de son coefficient dominant et d'un nombre de tels facteurs quadratiques (pas forcément distincts) qui correspondent chacun à une paire de racines conjuguées.

3.5.4. Proposition. *Soit K un corps commutatif et $P \in K[X]$ un polynôme non nul.*

(1) On peut écrire, de façon unique à l'ordre des $a_i \in K$ près, $P = (X - a_1) \dots (X - a_r)Q$ avec $r \geq 0$, où $Q \in K[X]$ est sans racine dans K. Dans ce cas $\{a_1, \dots, a_r\}$ est l'ensemble de toutes les racines de P dans K, et une racine de multiplicité m dans P est présente m fois parmi $a_1, \dots, a_r$.

(2) Le nombre r des racines comptées avec multiplicité vérifie $r + \deg(Q) = \deg(P)$, et donc $r \leq \deg(P)$.

(3) Si $K = \mathbf{C}$, le polynôme Q du point (1) est une constante, le coefficient dominant de P, et $r = \deg(P)$.

(4) Si $K = \mathbf{R}$, le polynôme Q se décompose en un produit d'une constante (le coefficient dominant de P) et un nombre $s \geq 0$ de polynômes quadratiques unitaires à discriminant strictement négatif. □

On a vu dans le chapitre précédent qu'il est intéressant de substituer pour X non pas un nombre, mais un endomorphisme ϕ d'un K-espace vectoriel E. Dans ce cas un monôme X^i donnera une puissance ϕ^i de l'endomorphisme (et en particulier la constante $X^0 = 1$ donne $\phi^0 = \mathrm{id}_E$). Cette substitution est facile a définir, mais on n'est pas dans le cadre considéré auparavant : $\mathrm{End}(E)$ est bien un anneau qui contient K, si on convient d'identifier le scalaire λ avec l'homothétie $\lambda\,\mathrm{id}_E$, mais ce n'est pas un anneau commutatif (à moins que $\dim(E) \leq 1$). La formule (17) pour la multiplication dans $K[X]$ est trouvée en supposant les axiomes d'un anneau commutatif, donc il n'est pas évident qu'elle reste valable en remplaçant X pas un élément d'un anneau non-commutatif. Pourtant la substitution $X := \phi$ est compatible avec cette formule, car la seule instance de la commutativité que (17) suppose est celle dans $p_i X^i q_j X^j = p_i q_j X^{i+j}$, et on a $p_i \phi^i q_j \phi^j = p_i q_j \phi^{i+j}$ parce que $\phi^i \in \mathrm{End}(E)$ commute avec (l'homothétie de facteur) q_j. Il en résulte que $X := \phi$ définit un homomorphisme $K[X] \to \mathrm{End}(E)$. Les endomorphismes dans l'image de cet homomorphisme, qu'on appelle les polynômes en ϕ, forment un sous-anneau *commutatif* de $\mathrm{End}(E)$, car la composition de tels endomorphismes reflète la multiplication (commutative) des polynômes:

$$(P+Q)[X := \phi] = P[X := \phi] + Q[X := \phi], \tag{18}$$

$$(PQ)[X := \phi] = P[X := \phi] \circ Q[X := \phi] = Q[X := \phi] \circ P[X := \phi]. \tag{19}$$

3.5.5. Proposition. *Pour $\phi \in \mathrm{End}(E)$, l'application $\sum_{i=0}^{d} p_i X^i \mapsto \sum_{i=0}^{d} p_i \phi^i$ est un homomorphisme d'anneaux $K[X] \to \mathrm{End}(E)$. Les polynômes annulés par cet homomorphisme forment un idéal de $K[X]$.* □

Le résultat suivant, qu'on l'utilisera souvent sans le citer explicitement, nous permettra de considérer la restriction de ϕ à certains sous-espaces. On pourra affaiblir son hypothèse à la commutation de f avec ϕ.

3.5.6. Proposition. *Soit $f = P[X := \phi]$ un polynôme endomorphisme $\phi \in \mathrm{End}(E)$. Alors $\mathrm{Ker}(f)$ et $\mathrm{Im}(f)$ sont des sous-espaces ϕ-stables de E.*

Preuve. D'abord, vérifions si $v \in \mathrm{Ker}(f)$ qu'on ait aussi $\phi(v) \in \mathrm{Ker}(f)$; cela découle de $f(\phi(v)) = (PX)[X := \phi](v) = \phi(f(v)) = \phi(\vec{0}) = \vec{0}$. Ensuite vérifions si $v \in \mathrm{Im}(f)$ que $\phi(v) \in \mathrm{Im}(f)$: par hypothèse on peut écrire $v = f(w)$, et on aura alors $\phi(v) = \phi(f(w)) = (XP)[X := \phi](w) = f(\phi(v)) \in \mathrm{Im}(f)$. □

3.6. *Quelques éléments d'arithmétique dans* $\mathbf{Z}$ *et dans* $K[X]$.

On a vu dans la proposition 3.5.4 que les polynômes dans $\mathbf{C}[X]$ et dans $\mathbf{R}[X]$ admettent une décomposition en facteurs qui ressemble celle des entiers strictement positifs en facteurs premiers. Le but de cette section est de montrer qu'une telle décomposition existe pour tout corps commutatif K (mais sans que les facteurs soient limités à être de degré 1 ou 2), et que d'autres propriétés de $\mathbf{Z}$ ont un pendant dans les anneaux $K[X]$. On commence avec les notions de divisibilité.

Quand un polynôme A est un *multiple* QB de B, on dira également que B *divise* A, que A est *divisible* par B, ou que B est un *diviseur* ou un *facteur* de A (il faut admettre qu'il y a a beaucoup de façons de dire la même chose). Dans ce cas, si $B \neq 0$, on écrit $Q = A/B$ pour le quotient. Selon cette définition tout polynôme divise 0 qui lui ne divise aucun autre polynôme ; mais souvent on exclut explicitement le polynôme nul, par exemple pour la factorisation. Les autres polynômes constants sont à l'autre extrémité, car étant inversible ils divisent tout polynôme, et ils ne sont divisibles par aucun polynôme non constant. On n'exclut pas ces polynômes des considérations, mais des facteurs constants dans $K[X]$, comme des facteurs ± 1 dans $\mathbf{Z}$, ne sont pas intéressants ; les formulations doivent souvent en tenir compte. Parfois on évite la possibilité de facteurs constants en se limitant aux polynômes unitaires, comme on évite les facteurs -1 en $\mathbf{Z}$ en se limitant aux nombres strictement positifs.

3.6.1. Définition. *Un polynôme non nul $P \in K[X]$ est réductible s'il s'écrit comme le produit de deux polynômes non constants. Un polynôme non constant et non réductible est un polynôme irréductible.*

Ces notions correspondent à celles de nombres composés et premiers dans $\mathbf{Z}$, et on remarque que, pareillement à 1 qui n'est ni composé ni premier dans $\mathbf{Z}$, les polynômes constantes ne sont ni réductibles ni irréductibles dans $K[X]$. (Mais on n'exige pas aux polynômes irréductibles d'être unitaires, pendant qu'on exige aux nombres premiers d'être positifs.) Le résultat suivant est immédiat.

3.6.2. Proposition. *Tout polynôme non constant s'écrit comme un produit de polynômes irréductibles.*

Preuve. Par récurrence sur le degré : soit le polynôme est irréductible et donc le produit d'un seul facteur, soit il est réductible, et on l'écrit comme produit de deux facteurs non constants de degré plus bas, lesquels peuvent être écrits chacun comme produit de polynômes irréductibles par hypothèse de récurrence. □

Si cette décomposition est facile à obtenir du point de vue théorique, ce résultat n'est pas effectif, et il n'existe pas de méthode générale et effective de trouver une telle écriture. Pour certains corps (par exemple pour les corps finis, et aussi pour $\mathbf{Q}$) des méthodes effectives existent pour d'abord décider si un polynôme est réductible et ensuite pour trouver une décomposition. Mais de temmes méthodes ne sont pas générales, et dépendent beaucoup de la nature spécifique du corps.

Un polynôme de degré 1 est toujours irréductible, car le degré du produit de deux polynômes non constants est au moins 2. Pour un corps algébriquement clos tel que $\mathbf{C}$, ce sont les seuls polynômes irréductibles, car dans ce cas tout polynôme non constant possède une racine, et donc un facteur de degré 1. Réciproquement si les seuls polynômes irréductibles de $K[X]$ sont ceux de degré 1, alors K est algébriquement clos, car dans la décomposition de la proposition 3.6.2 chaque facteur, étant de degré 1, donne une racine du polynôme. Il ne faut pas pour autant confondre le fait d'être irréductible avec l'absence de racines : si tout polynôme réductible de degré 2 ou 3 possède forcément une racine, un produit de deux ou plus polynômes irréductibles tous de degré ≥ 2 est réductible sans avoir des racines.

Les polynômes irréductibles dans $\mathbf{R}[X]$ sont les polynômes de degré 1 et les polynômes quadratiques avec discriminant strictement négatif, comme on a pu le voir dans la proposition 3.5.4(4). Mais pour les corps $K = \mathbf{Q}$ et $K = \mathbf{Z}/p\mathbf{Z}$ avec p un nombre premier, il existe des polynômes irréductibles dans $K[X]$ de degré d, quel que soit $d > 0$ (à cet égard c'est le corps de nombres réels qui est atypique).

3.6.3. Définition. *Un polynôme est scindé s'il s'écrit comme un produit de polynômes de degré* 1.

Cette notion nous servira surtout pour faciliter le discours dans le dernier chapitre. On peut remarquer que l'utilisation du participe passé "scindé" est un peu optimiste, vu le fait que la proposition 3.6.2 n'est pas effective. Néanmoins la terminologie est traditionnelle : on imagine le polynôme déjà écrit sous forme décomposée, et on constate que ses facteurs sont de degré 1. Si K est algébriquement clos, tout polynôme non constant est évidemment scindé (ainsi que la constante 1, la valeur du produit vide).

Si I et J sont des idéaux, l'ensemble $I+J = \{ x+y \mid x \in I, y \in J \}$ est aussi un idéal ; c'est immédiat de la définition. En particulier pour tout couple d'entiers a, b, l'ensemble $a\mathbf{Z} + b\mathbf{Z}$ de tout les entiers de la forme $ak + bl$ (avec k, l entiers) est un idéal de $\mathbf{Z}$. Il contient donc (sauf si $a = b = 0$) un plus petit élément $d > 0$. On aura $d\mathbf{Z} = a\mathbf{Z} + b\mathbf{Z}$, et d est donc un diviseur commun de a et b ; c'est en fait le plus grand diviseur commun (on écrit $d = \mathrm{pgcd}(a, b)$), car si d' est un autre diviseur commun ($a, b \in d'\mathbf{Z}$) alors $d'\mathbf{Z} \supseteq a\mathbf{Z} + b\mathbf{Z}$ et en particulier d, étant élément de $a\mathbf{Z} + b\mathbf{Z}$, est un multiple de d'.

3.6.4. Définition. *Si* $A, B \in K[X]$ *sont non nuls, l'unique générateur unitaire* D *(qui existe d'après la proposition 3.5.3) de l'idéal* $(A) + (B) = \{ UA + VB \mid U, V \in K[X] \}$ *de* $K[X]$ *est le plus grand diviseur commun de* A *et de* B*, noté* $D = \mathrm{pgcd}(A, B)$*. Des polynômes* U, V *avec* $\mathrm{pgcd}(A, B) = UA + VB$ *sont appelés des coefficients de Bezout pour* A, B*. Si* $\mathrm{pgcd}(A, B) = 1$ *on appelle* A, B *premiers entre eux.*

Avec cette définition du pgcd, le "théorème de Bezout" (qui affirme l'existence des coefficients de Bezout) est une évidence. Ce qui n'est pas évident, mais néanmoins vrai, est qu'il est possible de trouver explicitement $\mathrm{pgcd}(A, B)$ ainsi que des coefficients de Bezout. Cela se fait par une variante de l'algorithme d'Euclide : on maintient un couple de polynômes (P, Q) tel que $(P) + (Q) = (A) + (B)$, initialisé $(P, Q) = (A, B)$, et pour chacun de P et Q on maintient des coefficients U, V qui l'exprime sous la forme $UA + VB$. Tant que $Q \neq 0$ on remplace (P, Q) par (Q, R), où R est le reste de la division de P par Q. Comme $R = P - SQ$ pour un certain (quotient) S, on trouve le coefficient U_R de A comme $U_R = U_P - SU_Q$ en terme des coefficients correspondents U_P, U_Q de P et Q, et pareillement pour le coefficient V_R de B. Une fois que $Q = 0$ (terminaison qui est garantie par le fait que $\deg(Q)$ baisse à chaque itération) on aura $(A) + (B) = (P) + (0) = (P)$ donc $P = \mathrm{pgcd}(A, B)$, et les coefficients U_P, V_P seront des coefficients de Bezout pour A, B. Tout cela se passe en principe exactement comme dans $\mathbf{Z}$.

Si P est un polynôme irréductible, qu'on supposera unitaire, alors les seuls diviseurs unitaires de P sont 1 et P. Donc pour tout Q qui n'est pas multiple de P, les polynômes P, Q sont premiers entre eux.

3.6.5. Lemme d'Euclide. *Si un polynôme irréductible P divise un produit AB de polynômes, il divise au moins un des facteurs A, B.*

On remarque d'abord qu'aucun polynôme réductible ne peut avoir cette propriété : si $Q = ST$ avec S, T non constants, alors Q divise ST sans diviser S ni T. L'énoncé dit donc que dès qu'un polynôme est irréductible, il possède automatiquement une propriété bien plus forte, quelle propriété est une conséquence de l'existence des coefficients de Bezout, donc ultimement de la division euclidienne.

Preuve. On montrera que si P ne divise pas A, il divise B. Soient U, V des coefficients de Bezout pour A, P, avec donc $\text{pgcd}(A, P) = 1 = UA + VP$. Alors comme P divise AB, il divise $UAB + VPB = B$. □

Pour P irréductible et A non multiple de P, une relation de Bezout $1 = UA+VP$ s'interprète comme $UA \equiv 1$ modulo P, ce qui montre que $K[X]/(P)$ est un corps. C'est ainsi qu'on montre aussi que $\mathbf{Z}/p\mathbf{Z}$ est un corps si p est un nombre premier. Le lemme de Gauss exprime le fait que $K[X]/(P)$ est intègre.

Le lemme d'Euclide se généralise par une récurrence facile aux produits de plusieurs facteurs (qu'on peut voir comme le premier facteur multiplié par le produit des autres) : si un polynôme irréductible P divise un tel produit, il divise au moins un de ses facteurs (c'est pareil pour un nombre premier dans $\mathbf{Z}$).

3.6.6. Théorème de factorisation unique dans $K[X]$**.** *Tout polynôme non nul dans $K[X]$ s'écrit comme le produit de son coefficient dominant et d'un produit de $l \geq 0$ polynômes irréductibles unitaires. Cette écriture est unique à l'ordre des facteurs irréductibles près.*

Preuve. L'existence d'une telle décomposition est déjà prouvée dans la proposition 3.6.2 ; pour assurer que les facteur irréductibles sont unitaires, il suffit de mettre en facteur pour chaque polynôme son coefficient dominant (le polynôme qui reste sera alors unitaire), et de multiplier tous ces coefficients ensemble, ce qui le coefficient dominant du polynôme de qui était décomposé. Le point essentiel est donc de montrer l'unicité de la factorisation, donc on suppose qu'on a deux factorisations d'un même polynôme Q. La preuve que les deux sont identiques, à l'ordre des facteurs irréductibles près, est fait par une récurrence sur le nombre de facteurs irréductibles P_i dans la première factorisation. Si ce nombre est 0 alors Q est constant et ses deux "factorisations" identiques. Sinon le premier facteur irréductible unitaire P_1 divise Q et donc d'après le lemme d'Euclide (généralisé), P_1 divise au moins un des facteurs irréductibles unitaires de la seconde factorisation, ce qui n'est possible que si ce facteur est égal à P_1 ; on isole ces facteurs P_1, et on applique l'hypothèse de récurrence à Q/P_1 pour conclure que ce qui reste des deux factorisations (à part le facteur P_1) est identique à l'ordre des facteurs près. □

Il est parfois utile de généraliser le lemme d'Euclide à l'énoncé suivant, attribué à Gauss. Cette attribution n'est pas très heureuse, car Gauss n'est pas le premier à formuler l'énoncé, il ne le formule pas comme lemme mais comme conséquence de la factorisation unique, et il y a déjà plein de résultats dans d'autres domaines connus comme "lemme de Gauss". On peut retrouver le lemme d'Euclide comme un cas particulier, car si P est irréductible, alors "P ne divise pas A" entraîne $\text{pgcd}(P, A) = 1$.

3.6.7. Lemme de Gauss. *Si $P \in K[X]$ divise un produit AB et* $\text{pgcd}(P, A) = 1$*, alors P divise B.*

Preuve. On pourra déduire ceci du théorème de factorisation unique : en écrivant $PQ = AB$ dans laquelle on remplace chaque polynôme par sa forme factorisée, on aura deux factorisations du même polynôme dans laquelle aucun facteur irréductible provenant de P à gauche ne peut figurer parmi les facteurs provenant de A à droite, donc ils correspondent tous à des facteurs distincts de B, et du coup P divise B. Mais il est plus simple d'utiliser, comme pour le lemme d'Euclide, une relation de Bezout $\text{pgcd}(P, A) = 1 = UP+VA$ et d'écrire $B = 1B = UPB+VAB$, où P divise chaque terme de la somme. □

Chapitre 4. Déterminants.

On revient maintenant à l'algèbre linéaire d'un K-espace vectoriel E de dimension n. Pour déterminer si un famille de k vecteurs, exprimés dans une base, est libre ou liée (on peut supposer $k \leq n$, sinon elle est forcément liée), on connaît la méthode de l'échelonnement (ou pivot de Gauss). Cette méthode est basée sur la possibilité, si un vecteur v a une coordonnée i non nulle, de rendre les coordonnées i des autres vecteurs nulles en leur rajoutant un multiple convenable de v (ce qui ne change pas le statut libre/liée de la famille). Mais cette méthode ne marche pas toujours bien quand par exemple les vecteurs de la famille dépendent d'un ou de plusieurs paramètres, et on demande pour quelles valeurs des paramètres la famille est liée, car la question si une coordonnée est nulle dépendra en général des paramètres.

Dans le cas particulier $k = n$ (ou la famille sera une base si elle est libre) il existe une *expression* dans les coordonnées des vecteurs, avec la propriété qu'elle prend la valeur 0 si et seulement si la famille est liée. Ainsi on trouve sans difficulté une équation qui exprime la dépendance des vecteurs, même dans le cas où leurs coordonnées dépendent des paramètres. Cet expression est le *déterminant* de la matrice des coordonnées, dont le calcul (dans des cas simples) est au programme de la première année. Le déterminant joue un rôle important pour les systèmes de n équations linéaires en autant d'inconnues (les systèmes "carrés") : un tel système possède une solution unique si et seulement si le déterminant de la matrice de ses coefficients des inconnues est non nul ; le système s'appelle alors un *système de Cramer*. (Dans le cas contraire, donc avec déterminant nul, le système possède soit aucune, soit multiples solutions.) La relation du déterminant avec la dépendance des vecteurs reste à ce point probablement un peu magique ; au mieux on vous a donné des explications *ad hoc* pour $n = 2, 3$. Dans ce chapitre on définira la déterminant rigoureusement, et on expliquera ses propriétés.

4.1. *Déterminants en dimension $n \leq 3$, formes linéaires.*

Il est instructif de considérer d'abord la situation quand la dimension n est petite. Les cas $n = 0, 1$ ne sont guère intéressants : une famille vide est toujours libre (même en dimension 0) donc il convient de définir le déterminant de la matrice 0×0 comme 1 (surtout pas 0) ; une famille d'un seul vecteur est libre dès que le vecteur est non nul, donc on prendra pour le déterminant d'une matrice 1×1 son unique coefficient. Pour $n = 2$ il s'agit de décider si un famille $[(a, b)_{\mathcal{B}}, (x, y)_{\mathcal{B}}]$ est libre, où les vecteurs sont exprimés dans une base $\mathcal{B}$. Supposons que $(a, b)_{\mathcal{B}}$ est fixé et non nul, alors la famille sera liée seulement si $(x, y) \in K^2$ est un multiple scalaire de (a, b). En fonction de (x, y), le déterminant doit donc être une fonction qui s'annule précisément sur ces multiples scalaires, et comme ils forment un sous-espace de dimension 1, on pourra trouver une fonction *linéaire* qui fait l'affaire ; dans ce cas il suffit de prendre une fonction non nulle qui s'annule en (a, b). Cela détermine la fonction, à multiplication par un scalaire non nul près, et l'application linéaire $K^2 \to K$ qui donnera le déterminant 2×2 est celle dont la matrice par rapport aux bases canoniques de K^2 et de K est $(-b \quad a)$, c'est-à-dire la fonction $(x, y) \mapsto -bx + ay$. Ce qui est remarquable dans ce choix d'application linéaire est que la fonction choisie dépend elle même de façon linéaire du vecteur (a, b), c'est-à-dire que l'application $K^2 \to \mathcal{L}(K^2, K)$ qui associe $(a, b) \mapsto (-b \quad a)$ est une application linéaire (or rappelle que l'ensemble $\mathcal{L}(E, F)$ des applications linéaires $E \to F$ est lui-même un K-espace vectoriel, quels que soient les K-espaces vectoriels E, F). Cela implique en particulier que l'application $(-b \quad a)$ sera nulle quand on prend $(a, b) = (0, 0)$, et c'est juste ce qu'on veut : dans ce cas la famille sera liée, quel que soit (x, y). Le déterminant d'une matrice 2×2, défini par

$$\begin{vmatrix} a & x \\ b & y \end{vmatrix} \quad \stackrel{\text{déf}}{=} \quad \det \begin{pmatrix} a & x \\ b & y \end{pmatrix} \quad = \quad ay - bx \tag{20}$$

a donc la propriété de s'annuler si et seulement si $(a, b)_{\mathcal{B}}$ et $(x, y)_{\mathcal{B}}$ sont linéairement dépendants. L'idée d'avoir une forme linéaire $(-b \quad a)$ qui dépend de façon linéaire d'un vecteur (a, b) donnera lieu à la notion d'une forme bilinéaire, qu'on définira ci-dessous. L'expression $ay - bx$ est bilinéaire en (a, b) et (x, y).

En général on appellera *forme linéaire* sur E un élément de $E^* \stackrel{\text{déf}}{=} \mathcal{L}(E, K)$, l'espace des applications linéaires sur K à valeur scalaire, espace qu'on appelle aussi l'*espace dual* de E. On a $\dim(E^*) = n = \dim(E)$, car par rapport à une base de E et la base canonique de K, les formes linéaires sont données par une matrice $1 \times n$. Des exemples typiques de formes linéaires sont les n fonctions coordonnées par

rapport à une base $\mathcal{B}$, à savoir les formes $(x_1,\dots,x_n)_\mathcal{B} \mapsto x_i$ pour $i = 1,\dots,n$. La représentation matricielle montre que toute forme linéaire est des façon unique une combinaison linéaire de ces fonctions coordonnées, qui forment donc une base de E^*, appelée la *base duale* de $\mathcal{B}$. Même si $\dim(E^*) = \dim(E)$, et s'il existe donc des isomorphismes entre E et E^*, il n'y a pas une manière *naturelle* (donc sans utiliser une donnée supplémentaire, comme le choix d'une base) d'associer un vecteur de E à une forme linéaire. Par contre on peut associer un sous-espace de E a une forme linéaire, à savoir son noyau ; ce sera toujours un sous-espace de dimension $n-1$, ce qu'on appelle un *hyperplan vectoriel* de E, sauf si la forme est nulle (le noyau de la forme nulle est évidemment tout l'espace E, et donc de dimension n). Chaque hyperplan vectoriel H de E est le noyau d'une forme linéaire (il suffit de choisir une base de H, et de la compléter par un vecteur à une base de E ; la dernière fonction coordonnée pour cette base a pour noyau H), et deux formes linéaires avec noyau H sont égales à un facteur scalaire non nul près.

Revenons à la question de l'indépendance linéaire, pour $n = 3$. Si $v_1 = (a,b,c)_\mathcal{B}$ et $v_2 = (p,q,r)_\mathcal{B}$ sont des vecteurs indépendants dans un K-espace vectoriel E de dimension 3, ils engendrent un (hyper)plan vectoriel $P = \mathrm{Vect}(v_1,v_2)$ dans E, et il existe donc une forme linéaire non nulle $f \in E^*$, unique à un scalaire près, telle que $f(v_1) = f(v_2) = 0$, et avec donc $\mathrm{Ker}(f) = P$. Un troisième vecteur $v_3 = (x,y,z)_\mathcal{B}$ formera alors une famille libre avec v_1, v_2 si et seulement si $f(v_3) \neq 0$. Ce que le déterminant 3×3 nous fournira est un choix d'une telle forme linéaire *en fonction de* v_1, v_2, à savoir la forme dont la matrice par rapport à $\mathcal{B}$ est $(br-cq \quad -ar+cp \quad aq-bp)$, donc $f : (x,y,z)_\mathcal{B} \mapsto brx - cqx - ary + cpy + aqz - bpz$. On vérifie facilement que $f((x,y,z)_\mathcal{B}) = 0$ quand $(x,y,z) = (a,b,c)$ et quand $(x,y,z) = (p,q,r)$. En plus, cette forme linéaire f dépend de façon linéaire de chacun des vecteurs $v_1 = (a,b,c)_\mathcal{B}$ et $v_2 = (p,q,r)_\mathcal{B}$ quand l'autre est fixé, ce qui mènera à la notion d'une forme multilinéaire. Et finalement, on peut vérifier que f est la forme nulle si et seulement si v_1 et v_2 sont déjà liés. Ainsi le déterminant de matrice 3×3

$$\begin{vmatrix} a & p & x \\ b & q & y \\ c & r & z \end{vmatrix} \overset{\text{déf}}{=} \det\begin{pmatrix} a & p & x \\ b & q & y \\ c & r & z \end{pmatrix} = aqz - ary - bpz + brx + cpy - cqx \tag{21}$$

détermine une forme 3-linéaire en les vecteurs v_1, v_2, v_3 dont les coordonnées dans $\mathcal{B}$ forment les colonnes de la matrice, et cette forme s'annule précisément quand ces vecteurs sont dépendants.

4.2. *Formes multilinéaires alternées.*

En vue de ce qu'on vient de discuter, on pourrait définir l'espace vectoriel $\mathcal{L}_k(E)$ des formes k-linéaires sur E récursivement en prenant $\mathcal{L}_0(E) = K$ comme cas de base, et en définissant $\mathcal{L}_{k+1}(E)$ comme $\mathcal{L}(E,\mathcal{L}_k(E))$ pour $k \in \mathbf{N}$; autrement dit une forme k-linéaire sur E est une constante si $k = 0$, et sinon une application qui produit une forme $k-1$-linéaire sur E, de manière linéaire en fonction d'un vecteur de E. C'est essentiellement comme cela que les formes k-linéaires sont définies, mais l'idée d'une application qui à $v_1 \in E$ associe une application qui à $v_2 \in E$ associe… une application qui à $v_k \in E$ associe un scalaire est un peu difficile pour la discussion et pour la notation. On définira donc une forme k-linéaire comme une seule fonction qui à un k-uplet de vecteurs $[v_1,\dots,v_k]$ associe directement un scalaire. Mais les propriétés de la notion récursive seront préservées, notamment les fait que $\mathcal{L}_k(E)$ est un K-espace vectoriel, et que sa dimension est n^k (qu'on déduit de la règle générale que $\dim(\mathcal{L}(E,F)) = \dim(E)\times\dim(F)$, qui résulte de la représentation matricielle des applications linéaires).

4.2.1. Définition. *Pour un K-espace vectoriel E et $k \in \mathbf{N}$, une forme k-linéaire sur E est une application $f : E^k \to K$ avec la propriété que pour tout indice $1 \le i \le k$ et tout choix de $k-1$ vecteurs $v_1,\dots,v_{i-1},v_{i+1},\dots,v_k$ fixés, l'application $x \mapsto f(v_1,\dots,v_{i-1},x,v_{i+1},\dots,v_k)$ est une forme linéaire sur E. Les formes k-linéaires sur E forment un K espace vectoriel noté $\mathcal{L}_k(E)$, de dimension $\dim(E)^k$.*

Une conséquence immédiate de la linéarité est qu'une forme k-linéaire s'annule dès que l'un de ses k arguments est le vecteur nul. La plus simple façon de produire une forme k-linéaire non nulle est de prendre k formes linéaires non nulles $\ell_1,\dots,\ell_k \in E^*$ et de former leur produit où elles opèrent chacune sur l'un des K vecteurs, c'est-à-dire la fonction $f : (v_1,\dots,v_k) \mapsto \ell_1(v_1)\dots\ell_k(v_k)$. Une forme k-linéaire n'est en général pas de cette forme, mais peut être obtenue comme une combinaison linéaire de telles

formes simples. En fait, si on fixe une base $\mathcal{B}$ de E et on fait parcourir, indépendamment les uns des autres, à chaque ℓ_i la base duale des n fonctions coordonnées pour $\mathcal{B}$, on obtiendra une base de l'espace $\mathcal{L}_k(E)$ (quelle base a effectivement n^k éléments). Une autre façon de dire ceci est qu'une forme k-linéaire est déterminée par les valeurs qu'elle prend pour les k-uplets de vecteurs particuliers, où chaque vecteur est l'un des vecteurs de la base $\mathcal{B}$, et que ces valeurs sont indépendantes (pour chaque jeu de n^k valeurs dans K il existe une forme k-linéaire qui prend précisément ces valeurs). Cette propriété fait écho à celle des applications linéaires (à la base de leur représentation matricielle) qui dit qu'elles sont déterminées par leurs valeurs pour les vecteurs de $\mathcal{B}$. Pour comprendre cette propriété des formes k-linéaires, il suffit d'observer que la forme k-linéaire $f : (v_1, \ldots, v_k) \mapsto \ell_1(v_1) \ldots \ell_k(v_k)$, où chaque ℓ_j est une fonction coordonnée pour $\mathcal{B} = [b_1, \ldots, b_n]$ disons la i_j-ème fonction coordonnée ($1 \leq i_j \leq n$), s'annule pour chacun de ces k-uplets de vecteurs particuliers à l'exception de $[b_{i_1}, \ldots, b_{i_k}]$, pour lequel on a $f(b_{i_1}, \ldots, b_{i_k}) = 1$.

En fonction des vecteurs $v_1 = (a, b)_{\mathcal{B}}$ et $v_2(x, y)_{\mathcal{B}}$ d'une espace de dimension 2, les expressions ay et bx définissent des formes bilinéaires (c'est-à-dire 2-linéaires), et le déterminant de (20) en est une combinaison linéaire. (Le déterminant n'utilise pas les deux autres formes bilinéaires simples ax et by, qui sont par contre utilisées pour former $ax + by$, forme bilinéaire importante dans le cas $K = \mathbf{R}$, connue comme le *produit scalaire* qui munit $\mathbf{R}^2$ d'une structure d'*espace euclidien*. Des espaces vectoriels munis de ce type de structure forment le sujet d'un autre cours.) Dans le cas d'un espace de dimension 3, chacun des 6 termes dans la définition (21) provient d'une forme 3-linéaire simple, qui est formée comme le produit de 3 fonctions coordonnées (ces 6 formes figurent parmi les $3^3 = 27$ telles formes possibles).

Pour un espace de dimension n, les formes n-linéaires forment le cadre dans lequel on va définir le déterminant, mais il nous faudra un peu de direction pour trouver la bonne combinaison linéaire dans cet immense espace vectoriel de dimension n^n. Le principe qui va nous guider est qu'on veut avoir une forme multilinéaire qui s'annule dès qu'il existe une relation de dépendance linéaire entre les vecteurs qui forment ses arguments. Cela sera assuré dans le cas d'une forme multilinéaire *alternée*.

4.2.2. Définition. *Une forme k-linéaire sur un espace vectoriel E est appelé alternée si elle prend la valeur 0 sur tout k-uplet d'arguments contentant au moins une paire de vecteurs égaux.*

La condition pour une forme k-linéaire d'être alternée réduit considérablement leur possibilités. On voit tout de suite que, parmi les valeurs pour les k-uplets particuliers choisis parmi les vecteurs de $\mathcal{B}$, celles pour un choix faisant apparaître deux fois un même vecteur de base seront toutes nulles. Cela nous apprend déjà que pour $k > n$ la seule forme k-linéaire alternée est la forme nulle. Mais on peut dire plus.

4.2.3. Proposition. *Soit f une forme k-linéaire alternée sur E, et $i, j \in \{1, \ldots, k\}$ des indices distincts.*
(1) Pour tout $\lambda \in K$ on a $f(v_1, \ldots, v_{i-1}, v_i + \lambda v_j, v_{i+1}, \ldots, v_k) = f(v_1, \ldots, v_k)$.
(2) Si la famille de vecteurs $[v_1, \ldots, v_k]$ est liée, on a $f(v_1, \ldots, v_k) = 0$.
(3) Si on suppose $i < j$, on a $f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_{j-1}, v_i, v_{i+1}, \ldots, v_k) = -f(v_1, \ldots, v_k)$.

Preuve. Pour (1) on utilise la linéarité de f par rapport à v_i : $f(v_1, \ldots, v_{i-1}, v_i + \lambda v_j, v_{i+1}, \ldots, v_k) = f(v_1, \ldots, v_k) + \lambda f(v_1, \ldots, v_{i-1}, v_j, v_{i+1}, \ldots, v_k)$, où on observe que le second terme est nul car f est alternée (le vecteur v_j est présent deux fois parmi les arguments). La partie (2) en découle, car par hypothèse l'un des vecteurs v_i est égal à une combinaison linéaire des *autres* vecteurs v_j, et en appliquant (1) successivement pour les différents v_j pour soustraire leur contribution dans cette combinaison linéaire de v_i, on réduit ce i-ème argument de f au vecteur nul, et ceci étant fait la valeur de f sera certainement nulle. Pour (3) les arguments autres que v_i et v_j ne changent pas, donc on simplifiera l'argument en considérant $g : E^2 \to K$ avec $g(x, y) = f(v_1, \ldots, v_{i-1}, x, v_{i+1}, \ldots, v_{j-1}, y, v_{i+1}, \ldots, v_k)$, qui est une forme bilinéaire alternée. Alors on a $0 = g(x + y, x + y) = g(x, x) + g(x, y) + g(y, x) + g(y, y) = g(x, y) + g(y, x)$ pour tout $x, y \in E$, d'où $g(y, x) = -g(x, y)$ comme l'affirme l'énoncé (pour $x = v_i$ et $y = v_j$). □

4.2.4. Corollaire. *Si $\mathcal{B} = [b_1, \ldots, b_n]$ est une base de E, une forme k-linéaire alternée f est déterminée par les $\binom{n}{k}$ valeurs $f(b_{i_1}, \ldots, b_{i_k}) \in K$ avec indices strictement croissants : $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.*

Preuve. On a vu qu'une forme k-linéaire est déterminée par de telles valeurs où $(i_1, \ldots, i_k)$ parcourt $\{1, \ldots, n\}^k$, et que pour une forme alternée ces valeurs sont nulles dès qu'il y a deux indices égaux. Il

reste les valeurs prises par f pour les suites d'indices tous distincts (les *arrangements* de k indices choisis parmi $\{1,\dots,n\}$). Les valeurs obtenues pour l'ensemble de telles suites qui concernent toutes le même sous-ensemble (*combinaison*) S d'indices sont toutes reliées par les relations de la proposition 4.2.3(3) qui intervertissent une paire d'indices. Il suffit donc de connaître l'une de ces valeurs, pour laquelle on peut prendre $f(b_{i_1},\dots,b_{i_k})$ ou $S=\{i_1,\dots,i_k\}$ avec $i_1<\cdots<i_k$. □

En particulier une forme n-linéaire alternée f sur E est déterminée par la seule valeur $f(b_1,\dots,b_n)$, et l'espace des formes n-linéaires alternées est de dimension 1 au plus. C'est la propriété clé qui mène à la définition du déterminant, mais il reste un tout petit souci à enlever : un arrangement d'indices qui n'est pas croissant peut en général être lié de différents manières à l'arrangement croissant de la même combinaison d'indices, chaque manière introduisant un nombre de changements de signe égal au nombre d'échanges de deux indices utilisées pour les ranger en ordre croissant. Si jamais il y avait deux façons différentes d'arriver à l'arrangement croissant pour lesquelles la *parité* des nombres d'échanges était *différente*, cela donnerait une relation $f(b_{i_1},\dots,b_{i_k})=-f(b_{i_1},\dots,b_{i_k})$ qui exclurait toute forme k-linéaire alternée non nulle (car le même argument s'appliquerait, quelle que soit la combinaison de k indices). C'est un fait fondamental de la théorie des permutations qu'une telle situation ne se produit jamais. On vérifie ceci facilement pour $k\leq 3$, mais donnons en argument pour $k\in\mathbf{N}$ quelconque.

Il n'est pas nécessaire ici de définir les permutations en général. On appellera permutation de n tout arrangement $\sigma=(\sigma_1,\dots,\sigma_n)$ tel que $\{\sigma_1,\dots,\sigma_n\}=\{1,\dots,n\}$ (on trouve donc tous les nombres $1,\dots,n$ chacun une fois dans l'arrangement, mais pas dans un ordre particulier). On notera $\mathbf{S}_n$ l'ensemble des permutations de n (elles sont $n!=n\times(n-1)\times\cdots\times 2\times 1$ en nombre), et on appelle $\sigma'=(\sigma'_1,\dots,\sigma'_n)$ une permutation obtenue de σ par une transposition simple s'il existe i avec $1\leq i<n$ tel que $\sigma'_i=\sigma_{i+1}$, $\sigma'_{i+1}=\sigma_i$, et $\sigma'_j=\sigma_j$ pour tout $j\in\{1,\dots,n\}\setminus\{i,i+1\}$. La permutation particulière $(1,2,\dots,n)$ est appelée la permutation identique de n.

4.2.5. Proposition/Définition. *Pour tout $n\in\mathbf{N}$ il existe une application $f_n:\mathbf{S}_n\to\{1,-1\}$ telle que $f_n(\sigma)=-f_n(\sigma')$ pour tout $\sigma\in\mathbf{S}_n$ et toute permutation σ' obtenue de σ par une transposition simple, et $f_n(\sigma)=1$ quand σ est la permutation identique de n. On appelle $f_n(\sigma)$ la signature de σ, notée* $\mathrm{sg}(\sigma)$.

Preuve. On définit f_n par récurrence sur n : pour $n\leq 1$ c'est la fonction constante à valeur 1 (sur la permutation identique, qui est la seule dans ces cas) ; pour $n>1$ et $\sigma\in\mathbf{S}_n$ soit i l'indice tel que $\sigma_i=n$, alors on pose $f_n(\sigma)=(-1)^{n-i}f_{n-1}(\tau)$ où $\tau=(\sigma_1,\dots,\sigma_{i-1},\sigma_{i+1},\dots,\sigma_n)\in\mathbf{S}_{n-1}$. Pour σ la permutation identique de n on aura $i=n$ et τ est la permutation identique de $n-1$, donc $f_n(\sigma)=1$ par l'hypothèse de récurrence. Pour vérifier $f_n(\sigma)=-f_n(\sigma')$ quand σ' est obtenue de σ par une transposition simple, on distingue deux cas : $\sigma'_i=n$ (la transposition ne déplace pas le terme $\sigma_i=n$), et $\sigma'_i\neq n$ auquel cas on aura $\sigma'_{i\pm1}=n$ pour un indice $i\pm1$ voisin de i, c'est-à-dire $i\pm1\in\{i-1,i+1\}$. Dans le premier cas on a $f_n(\sigma')=(-1)^{n-i}f_{n-1}(\tau')$ où τ' est obtenue de τ par une transposition simple, donc $f_n(\sigma')=-f_n(\sigma)$ par l'hypothèse de récurrence ; dans le second cas on a $f_n(\sigma')=(-1)^{n-(i\pm1)}f_{n-1}(\tau)=-f_n(\sigma)$. □

On remarque que dans la définition $n-i$ est le nombre de termes de σ qui viennent *après* le terme $\sigma_i=n$ (et qui sont forcément plus petit que celui-ci). On voit alors facilement que $\mathrm{sg}(\sigma)=(-1)^N$, où N est le nombre d'*inversions* de σ, c'est-à-dire de couples d'indices (i,j) avec $i<j$ et $\sigma_i>\sigma_j$.

4.2.6. Corollaire. *Pour toute base $\mathcal{B}=[b_1,\dots,b_n]$ de E il existe une forme n linéaire alternée unique f sur E telle que $f(b_1,\dots,b_n)=1$. Si $[\ell_1,\dots,\ell_n]$ est la base de E^* duale de $\mathcal{B}$, on a pour $v_1,\dots,v_n\in E$:*

$$f(v_1,\dots,v_n)=\sum_{\sigma\in\mathbf{S}_n}\mathrm{sg}(\sigma)\ell_{\sigma_1}(v_1)\dots\ell_{\sigma_n}(v_n).$$

La formule définit la forme k-linéaire indiquée telle que $f(b_{\sigma_1},\dots,b_{\sigma_n})=\mathrm{sg}(\sigma)$ pour tout $\sigma\in\mathbf{S}_n$, et $f(b_{i_1},\dots,b_{i_n})=0$ pour tout n-uplet d'indices $(i_1,\dots,i_n)\in\{1,\dots,n\}^n$ qui n'est pas permutation de n, c'est-à-dire avec au moins un indice qui apparaît deux fois. D'après la proposition 4.2.5, cette forme est en fait alternée (pour voir le changement de signe pour une transposition de (i,j) avec $j-i>1$, on la réalise par une suite d'un nombre impair $2(j-i)-1$ de transpositions simples, à trouver comme exercice). Par une expression similaire (mais encore plus compliquée à écrire) on montre que les $\binom{n}{k}$ valeurs dont parle le corollaire 4.2.4 peuvent toutes être choisies indépendamment, et la dimension de l'espace des formes k-linéaires alternées sur E est donc donnée par ce coefficient binomial $\binom{n}{k}$ (qui vaut 1 si $k=n$).

4.3. *Définitions de déterminant.*

Jusqu'ici on a travaillé dans un K-espace vectoriel E de dimension n pour motiver l'existence d'une forme n-linéaire alternée. Mais pour définir le déterminant de façon unique, le mieux est de prendre comme point de départ une matrice : d'une part on ne saura pas faire un choix d'une forme n-linéaire alternée particulière (elle forment un espace de dimension 1) sans utiliser une base de E, et d'autre part en considérant une matrice on n'est pas obligé de supposer qu'elle représente un n-uplet de vecteurs, et notamment pas que ses coefficients soient pris dans un corps. En fait on donnera une définition du déterminant dans trois cadres distincts : celui des matrices carrées à coefficients dans un anneau commutatif, celui des n-uplets de vecteurs de E, et celui des endomorphismes de E.

Déterminant d'une matrice à coefficients dans un anneau commutatif.

La définition du déterminant d'une matrice carrée ne fait intervenir que des additions, soustractions et des multiplications; il est donc possible de la formuler pour les matrices à coefficients dans un anneau, pas nécessairement dans un corps. Et ceci est très utile, car notre raison principale de considérer le déterminant et pouvoir définir le polynôme caractéristique, qui est le déterminant d'une matrice à coefficients dans $K[X]$. Cependant, pour avoir les propriétés les plus basiques des déterminants, il faut supposer que l'anneau soit *commutatif* (car la multilinéarité nécessite des scalaires qui commutent).

4.3.1. Définition. *Pour un anneau commutatif R et $n\mathbf{N}$, le déterminant d'une matrice $A = (A_{i,j})_{i,j=1}^n$ à coefficients dans R est :*

$$\det(A) = \sum_{\sigma\in\mathbf{S}_n} \left(\operatorname{sg}(\sigma) \prod_{j=1}^{n} A_{\sigma_j,j} \right) \in R. \tag{22}$$

Un nombre d'identités fondamentales, valables pour tout anneau commutatif R, suivent directement de cette définition. Nous formulons seulement les cas de base ; d'autres identités peuvent être déduites en les combinant (notamment (2) permet de conclure pour les lignes tout ce qui est dit pour les colonnes).

4.3.2. Théorème. *Le déterminant d'une matrice $n \times n$ sur R possède les propriétés suivantes.*

(1) Si $f : R \to S$ est un homomorphisme d'anneaux commutatifs, l'application de f au déterminant a le même effet que d'appliquer f à tous les coefficients: $f\big(\det\big((A_{i,j})_{i,j=1}^n\big)\big) = \det\big((f(A_{i,j}))_{i,j=1}^n\big)$.

(2) La transposition de la matrice ne change pas le déterminant: $\det(A) = \det({}^tA)$.

(3) Le déterminant de la matrice identité est 1.

(4) Comme fonction d'une seule colonne (donc en fixant les autres) le déterminant définit une application R-linéaire: si j et les colonnes de A autres que la colonne j sont fixés, et si $f : R^n \to R$ est défini par $f(C) = \det(A \leftarrow_j C)$ où $A \leftarrow_j C$ désigne la matrice obtenue de A en remplaçant la colonne j par C, on a $f(\sum r_i C_i) = \sum_i r_i f(C_i)$ pour toute combinaison R-linéaire $\sum r_i C_i$ avec $r_i \in R$ et $C_i \in R^n$.

(5) Le déterminant est alterné en les colonnes: si deux colonnes de A sont identiques, alors $\det(A) = 0$.

(6) Si A possède un bloc de zéros en bas à gauche qui touche à la diagonale principale, le déterminant se factorise comme le produit des deux déterminants des blocs carrés sur la diagonale principale:

$$\det\begin{pmatrix} A & C \\ \mathbf{0} & B \end{pmatrix} = \det(A)\det(B) \qquad \text{si } A \text{ et } B \text{ sont des blocs carrés.}$$

(7) Si A est triangulaire, alors le déterminant est égal au produit des coefficients diagonaux:

$$\det\begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \dots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \dots & a_{2,n} \\ 0 & 0 & a_{3,3} & \dots & a_{3,n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{n,n} \end{pmatrix} = a_{1,1}\, a_{2,2}\, a_{3,3} \dots a_{n,n}$$

(8) Le déterminant est multiplicatif par rapport au produit matriciel de matrices carrées:

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

(9) Si $n > 0$, le déterminant peut être "développé" en termes de déterminants de matrices $(n-1)\times(n-1)$: pour tout $j \in \{1, \dots, n\}$ fixé, on a $\det(A) = \sum_{i=1}^n (-1)^{i-j} A_{i,j} \det(A_{\widehat{i,j}})$, où $A_{\widehat{i,j}}$ est la matrice $(n-1) \times (n-1)$ obtenue à partir de A en supprimant la ligne et la colonne du coefficient $A_{i,j}$.

Toutes ces identités sont obtenues à partir de la définition par des manipulations algébriques plus ou moins compliquées. On ne donnera ici seulement une indication du type de manipulation et du raisonnement qui mènent à ces identités, parfois accompagnée d'une explication de leur signification.

L'identité (1) est une conséquence directe de la notion de homomorphisme, en vue du fait que le déterminant est donné par une expression explicite (aussi grande qu'elle soit). Elle est d'un importance fondamentale, ne serait-ce que parce qu'elle dit que la valeur du déterminant ne change pas quand on interprète des coefficients dans une anneau commutatif plus grand (avec f l'inclusion des anneaux en question). Une utilisation moins évidente est avec f un morphisme non injectif: cela permet par exemple de calculer l'évaluation en $a \in K$ du déterminant d'une matrice de polynômes en évaluant tous les polynômes de la matrice en a, et de calculer le déterminant de la matrice de scalaires ainsi obtenue.

L'identité (2) est une conséquence du fait que chaque terme $\mathrm{sg}(\sigma)\prod_{i=1}^{n} A_{i,\sigma_i}$ de l'expression pour le déterminant de ${}^tA$ est en correspondance avec un terme du déterminant de A pour une autre permutation, appelée la permutation inverse σ^{-1} de σ, qui vérifie $\mathrm{sg}(\sigma^{-1}) = \mathrm{sg}(\sigma)$. Si on a appelle matrice de la permutation σ la matrice ayant pour colonne j le vecteur $\mathbf{e}_{\sigma_j}$ de la base canonique de K^n (pour $j = 1, \ldots, n$), la permutation σ^{-1} est par définition celle dont la matrice est la transposée de celle de σ. Il s'agit clairement d'une correspondance bijective (et même involutive : elle est sa propre réciproque). Pour voir que $\mathrm{sg}(\sigma^{-1}) = \mathrm{sg}(\sigma)$, on peut soit argumenter que les inversions de σ sont en bijection avec celles de σ^{-1}, soit argumenter que le fait d'effectuer une transposition simple sur σ résulte toujours en une transposition effectuée sur σ^{-1}, qui correspond à un nombre *impair* de transpositions simples.

L'identité (3) est évidente. Avec les identités (4) et (5) elle permet de calculer la valeur d'un déterminant quelconque sans faire référence à la définition (22).

L'identité (4) affirme le caractère n-linéaire en les colonnes du déterminant (et grâce à (2) il est aussi n-linéaire en les lignes), mais on a évité ce terme qui a été introduit dans le cadre des espaces vectoriels. Elle est fondamentale pour le calcul pratique des déterminants (l'application directe de la définition étant en général trop compliquée), permettant le développement par une colonne, ou (avec l'identité (5)) la simplification par opérations sur les colonnes. Cette identité est une conséquence de la forme de l'expression définissant le déterminant : chaque terme $\mathrm{sg}(\sigma)\prod_{j=1}^{n} A_{\sigma_j,j}$ de (22) est une fonction R-linéaire de la colonne j de A (le reste de A étant fixe), car c'est un multiple de la coordonnée numéro σ_j de cette colonne, dont la valeur pour la matrice A est $A_{\sigma_j,j}$. La somme pour $\sigma \in \mathbf{S}_n$ de ces termes est donc aussi une fonction R-linéaire de la colonne j de A.

L'identité (5) affirme le caractère alterné en les colonnes du déterminant; c'est la raison d'être de la sommation alternée sur toutes les permutations dans la définition du déterminant. Si les colonnes i et j de A sont identiques, on considère les paires formées d'une permutation σ et de la permutation τ obtenu en effectuant la transposition (i, j) sur σ (comme la même transposition appliqué à τ redonne σ, il s'agit d'une partition de l'ensemble des permutation en paires). Les produits dans les termes associés à σ et τ dans (22) diffèrent seulement aux indices i et j, mais on a $A_{\sigma_i,i}A_{\sigma_j,j} = A_{\tau_j,i}A_{\tau_i,j} = A_{\tau_j,j}A_{\tau_i,i}$ (la dernière équation car $A_{k,i} = A_{k,j}$ pour tout k), donc ce produit de deux facteurs est le même pour les deux termes, et par conséquence le produit entier aussi. Mais $\mathrm{sg}(\sigma) = -\mathrm{sg}(\tau)$ les deux termes s'annulent ; en prenant la somme sur toutes les paires on obtient $\det(A) = 0$.

Quand l'identité (6) s'applique, elle permet une simplification considérable du calcul du déterminant, car le nombre combiné de termes dans les expressions pour $\det(A)$ et $\det(B)$ est en général nettement plus petit que le nombre de termes dans le déterminant global. Cette simplification est une conséquence du fait que, pour qu'une permutation σ donne une contribution non nulle au déterminant, il faut qu'elle permute les indices $(k+1, \ldots, n)$ entre eux (où k est la taille de A), car si $\sigma_j > k$ pour un indice $j \leq k$ on aura $A_{\sigma_j,j} = 0$. Par conséquent les indices $(1, \ldots, k)$ doivent aussi être permutés entre eux, et la permutation σ consiste effectivement en deux permutations indépendantes $\sigma^{(1)} \in \mathbf{S}_k$, $\sigma^{(2)} \in \mathbf{S}_{n-k}$ de ces deux parties. Aussi une inversion d'un tel σ ne peut concerner qu'une des deux parties, donc $\mathrm{sg}(\sigma)$ est le produit $\mathrm{sg}(\sigma^{(1)})\,\mathrm{sg}(\sigma^{(2)})$ des signatures des deux permutations. Du coup la somme dans (22) se factorise comme le produit de deux sommes, correspondants respectivement à $\det(A)$ et $\det(B)$.

L'identité (7) découle de (6) par récurrence sur n, ou se démontre directement car toute permutation non identique aura $j < \sigma_j$ pour au moins un indice j, ce qui annule sa contribution à cause de $A_{\sigma_j,j} = 0$.

L'identité (8) est une propriété fondamentale, qui dit que le déterminant est multiplicatif (ou de façon équivalente, qu'il est un morphisme du monoïde multiplicatif de $\mathcal{M}_n(R)$ vers celui de R). Elle est un peu plus difficile à démontrer que les autres identités, car l'expansion de (22) pour la matrice $A \cdot B$ est extrêmement compliquée. On verra que pour le cas où R est un corps, elle découle facilement de ce qu'on a dit sur les formes n-linéaires alternées. Pour la montrer pour le cas général d'un anneau commutatif, on pourra utiliser que chaque colonne du produit $A \cdot B$ est le produit de A et de la colonne correspondante de B, et pour A fixé elle est donc une fonction R-linéaire de cette colonne ; d'après (4) les expressions $\det(A \cdot B)$ et $\det(A)\det(B)$ sont donc des fonctions R-multilinéaires des n colonnes de B. Par conséquent, l'égalité des ces expressions sera assurée si elles prennent les mêmes valeurs dans les cas particuliers où chaque colonne de B est un des générateurs canoniques $\mathbf{e}_i$ de R^n (elle contient un seul coefficient non nul, qui est 1). Pour ces cas, si deux colonnes de B sont le même générateur, alors les deux colonnes correspondantes de $A \cdot B$ seront aussi égales (et égale à une colonne de A), donc $\det(B) = 0 = \det(A \cdot B)$. Il reste le cas où B est une matrice de permutation pour un certain $\sigma \in \mathbf{S}_n$. L'identité pour ce cas se démontre par récurrence sur le nombre d'inversions de σ : elle est évidente si σ est la permutation identique (car B sera alors la matrice identité) et sinon on pourra réduire le nombre d'inversions de B en effectuant une transposition simple sur σ, qui résulte en l'échange de deux colonnes voisines dans B et dans $A \cdot B$, et donc en le changement du signe du déterminant pour ces deux matrices.

Finalement l'identité (9) est une conséquence de (4), qui permet de développer le déterminant comme combinaison R-linéaire de déterminants $n \times n$, à savoir $\det(A) = \sum_{i=1}^{n} A_{i,j} \det(A \leftarrow_j \mathbf{e}_i)$, où $\mathbf{e}_i$ est la colonne donnant le i-ème générateur canonique de R^n. Le fait que $\det(A \leftarrow_j \mathbf{e}_i) = (-1)^{i-j} \det(A_{\widehat{i,j}})$ peut être montré en utilisant le caractère alterné du déterminant par lignes et par colonnes, ou directement de la définition, ainsi. Seules les permutations σ avec $\sigma_j = i$ donnent une contribution à $\det(A \leftarrow_j \mathbf{e}_i)$, et elles sont en bijection avec les permutations de $n-1$, par l'opération de supprimer le terme $\sigma_j = i$, et de soustraire 1 de la valeur des termes restants qui sont $> i$ (rendant le résultat une permutation). On a $\#\{\, j' > j \mid \sigma_{j'} < i \,\} - \#\{\, j' < j \mid \sigma_{j'} > i \,\} = i - j$ par un simple argument de comptage, donc la parité du nombre d'inversions que cette opération enlève de la permutation est celle de $i - j$.

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

4.3.3. Définition. *Si E est un K-espace vectoriel de dimension n et $\mathcal{E} = [e_1, \ldots, e_n]$ une base de E. Le déterminant $\det_B(v_1, \ldots, v_n)$ par rapport à $\mathcal{E}$ d'un n-uplet de vecteurs $v_1, \ldots, v_n \in E$ est le déterminant de la matrice dont les colonnes donnent les coordonnées des vecteurs v_j dans la base $\mathcal{E}$:*

$$\det_{\mathcal{E}}(v_1, \ldots, v_n) = \det(M) \quad \text{où } M \in \mathrm{Mat}_n(K) \text{ vérifie } v_j = (M_{1,j}, \ldots, M_{n,j})_{\mathcal{E}} \text{ pour } 1 \leq j \leq n. \tag{23}$$

Une autre façon de décrire M est que, avec $[\ell_1, \ldots, \ell_n]$ la base duale de $\mathcal{E}$ (les fonctions coordonnées pour $\mathcal{E}$), on a $M = \big(\ell_i(v_j)\big)_{i,j=1,\ldots,n}$. Alors l'expression (22) appliquée à $\det(M)$ donne la formule du corollaire 4.2.6. Les propriétés fondamentales de ce déterminant de n vecteurs sont les suivantes.

4.3.4. Théorème.
(1) $\det_{\mathcal{E}} : E \times \cdots \times E \to K$ est une forme n-linéaire alternée, et $\det_{\mathcal{E}}(e_1, \ldots, e_n) = 1$.
(2) Pour tout forme n-linéaire alternée $f : E \times \cdots \times E \to K$ il existe $\lambda \in K$ tel que $f = \lambda \det_{\mathcal{E}}$.
(3) Si $\mathcal{B} = [b_1, \ldots, b_n]$ est une autre base de E, le facteur λ du point précédent pour $f = \det_{\mathcal{B}}$ est donné par $\lambda = \det_{\mathcal{E}}(b_1, \ldots, b_n)^{-1}$ (et en particulier $\det_{\mathcal{E}}(b_1, \ldots, b_n) \neq 0$).
(4) Dans ce cas on a $\det_{\mathcal{E}}(v_1, \ldots, v_n) = \det_{\mathcal{E}}(b_1, \ldots, b_n) \det_{\mathcal{B}}(v_1, \ldots, v_n)$ pour tout $v_1, \ldots, v_n \in E$.
(5) Pour $v_1, \ldots, v_n \in E$ on a $\det_{\mathcal{E}}(v_1, \ldots, v_n) \neq 0$ si et seulement si $[v_1, \ldots, v_n]$ forme une base de E.

Preuve. Le point (1) est évident, il découle aussi bien des propriétés (3)–(5) du théorème 4.3.2 que du fait qu'on vient de retrouver la formule du corollaire 4.2.6. Le point (2) exprime le fait que les formes n-linéaires alternées forment un espace vectoriel de dimension 1, et que $\det_{\mathcal{E}}$ n'est pas la forme nulle, ce qui est clair par ce qui précède. Pour (3) on sait que $\det_{\mathcal{B}}(b_1, \ldots, b_n) = 1$, ce qui donne $\lambda \det_{\mathcal{E}}(b_1, \ldots, b_n) = 1$. Le point (4) n'est qu'une réécriture de $\det_{\mathcal{B}} = \lambda \det_{\mathcal{E}}$ comme $\det_{\mathcal{E}} = \lambda^{-1} \det_{\mathcal{B}}$. La partie "si" du point (5) est mentionnée dans (3) ; réciproquement si $[v_1, \ldots, v_n]$ n'est pas une base, la famille est liée, et toute forme n-linéaire alternée, en particulier $\det_{\mathcal{E}}$, s'annule en $[v_1, \ldots, v_n]$ d'après la proposition 4.2.3(2). □

On remarque que si A est la matrice dont les colonnes expriment les vecteurs $v_1, \ldots, v_n$ dans la base $\mathcal{B}$, et P est la matrice de passage de $\mathcal{E}$ vers $\mathcal{B}$, alors $P \cdot A$ est la matrice dont les colonnes expriment les vecteurs $v_1, \ldots, v_n$ dans la base $\mathcal{E}$. Le point (4) se traduit alors par $\det(P \cdot A) = \det(P)\det(A)$.

Déterminant d'un endomorphisme.

Pour un endomorphisme de E, on pourrait définir son déterminant comme le déterminant de sa matrice par rapport à une base $\mathcal{B}$ de E. Mais une telle définition suggère une dépendance de la base $\mathcal{B}$, pendant que, contrairement au déterminant $\det_{\mathcal{B}}$ de n vecteurs, le déterminant d'un endomorphisme est en fait indépendant de la base. Pour cette raison on préfère donc de donner une définition qui rend cette indépendance manifeste, en évitant toute mention d'une base.

4.3.5. Proposition/Définition. *Pour tout endomorphisme ϕ d'un K-espace vectoriel E de dimension n, il existe un $\lambda \in K$ unique tel que, pour toute forme n-linéaire alternée f sur E et tout $v_1, \ldots, v_n \in E$ on ait $f(\phi(v_1), \ldots, \phi(v_n)) = \lambda f(v_1, \ldots, v_n)$. Ce scalaire λ est par définition le déterminant* $\det(\phi)$ *de ϕ.*

Preuve. Il découle de la linéarité de ϕ que si $f : E \times \cdots \times E \to K$ est n-linéaire alternée, alors l'application $(v_1, \ldots, v_n) \mapsto f(\phi(v_1), \ldots, \phi(v_n))$ l'est aussi. Pour $f = 0$ la condition donnée est vérifiée indépendamment de λ, donc on supposera $f \neq 0$ (ce qui est possible, par exemple en prenant $f = \det_{\mathcal{B}}$ pour une base quelconque $\mathcal{B}$ de E). Alors l'existence d'un unique scalaire λ qui vérifie la condition pour f est une conséquence du fait que les formes n-linéaires alternées forment un espace vectoriel de dimension 1. Or un tel λ vérifie aussi la condition pour tout multiple scalaire μf de f, et encore par l'argument de dimension 1, cela montre que λ vérifie la condition pour toute forme n-linéaire alternée sur E. □

Les propriétés principales de cette notion de déterminant découlent facilement de la définition. Nous les résumons pour référence.

4.3.6. Théorème. *Le déterminant d'un endomorphisme ϕ d'un espace vectoriel E vérifie les propriétés :*
(1) $f(\phi(v_1), \ldots, \phi(v_n)) = \det(\phi) f(v_1, \ldots, v_n)$ pour toute forme n-linéaire alternée f et $v_1, \ldots, v_n \in E$.
(2) Si $\mathcal{B} = [b_1, \ldots, b_n]$ est une base de E, alors $\det_{\mathcal{B}}(\phi(v_1), \ldots, \phi(v_n)) = \det(\phi)\det_{\mathcal{B}}(v_1, \ldots, v_n)$.
(3) Pour toute base $\mathcal{B} = [b_1, \ldots, b_n]$ de E on a $\det(\phi) = \det_{\mathcal{B}}(\phi(b_1), \ldots, \phi(b_n)) = \det(\mathrm{Mat}_{\mathcal{B}}(\phi))$.
(4) Si ψ est un autre endomorphisme de E, on a $\det(\phi \circ \psi) = \det(\phi)\det(\psi)$.
(5) On a $\det(\phi) \neq 0$ si et seulement si ϕ est un isomorphisme.

Preuve. Le point (1) est juste la définition, (2) en est le cas particulier pour $f = \det_{\mathcal{B}}$, dont (3) découle en prenant $[v_!, \ldots, v_n] = \mathcal{B}$, remarquant pour la dernière égalité reflète directement les définitions de $\det_{\mathcal{B}}$ et de $\mathrm{Mat}_{\mathcal{B}}(\phi)$. Pour (4) il suffit de remplacer chaque v_i par $\psi(v_i)$ dans (1). Pour (5) on sait que ϕ est un isomorphisme si et seulement si l'image par ϕ d'une base de E est de nouveau une base, et le point est donc une conséquence de théorème 4.3.4(5), compte tenu de (3) du théorème actuel. □

On voit en particulier que dans le point (3) que effectivement, le déterminant d'un endomorphisme peut être calculé comme le déterminant de sa matrice par rapport à une base quelconque. Et le point (4) nous donne la multiplicativité du déterminant, comme énoncée dans le théorème 4.3.2(8), pour le cas général des matrices carrées à coefficients dans K (pendant que la règle $\det(P \cdot A) = \det(P)\det(A)$ vue ci-dessus se limitait encore au cas où P est une matrice de passage, donc inversible).

4.4. *Déterminants et matrices inverses: la règle de Cramer.*

Fixons l'indice j d'une colonne, pour les matrices $n \times n$ à coefficients dans un anneau commutatif R (on peut penser au cas particulier d'un corps, mais le cas général nous sera utile). Le théorème 4.3.2(4) dit que pour une telle matrice A donnée, l'application $f : R^n \to R$ donnée par $f(C) = \det(A \leftarrow_j C)$ est R-linéaire, donc elle est donnée par une matrice $1 \times n$ à coefficients dans R qu'on appellera $M^{(j)}$. En fait (9) du théorème en donne les coefficients : le coefficient $M^{(j)}_{1,i}$ est $(-1)^{i-j}\det(A_{\widehat{i,j}})$ pour $i = 1, \ldots, n$. On a évidemment $f(C_j(A)) = \det(A)$ si $C_j(A)$ est la colonne j de A, et par le caractère alterné du déterminant f s'annule pour les autres colonnes de A, c'est-à-dire $f(C_k(A)) = 0$ si $k \neq j$. Sous forme matricielle, cela veut dire que $M^{(j)} \cdot A = \det(A)\,{}^t\mathbf{e}_j$, où comme avant $\mathbf{e}_j \in R^n$ est le j-ème vecteur de la

base canonique, et ${}^t\mathbf{e}_j$ désigne qu'il est vu ici comme une matrice $1 \times n$. Il est clair qu'on peut également décrire ${}^t\mathbf{e}_j$ comme la ligne j de la matrice identité $n \times n$, matrice qu'on notera id_n.

L'identité $M^{(j)} \cdot A = \det(A)\,{}^t\mathbf{e}_j$ a une application au systèmes d'équations en n inconnues ayant A comme matrice de coefficients. Ces systèmes s'écrivent sous forme matriciel, si on appelle $x_1, \ldots, x_n$ les inconnues, comme

$$\begin{pmatrix} A_{1,1} & A_{1,2} & \ldots & A_{1,n} \\ A_{2,1} & A_{2,2} & \ldots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \ldots & A_{n,n} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}. \tag{24}$$

On utilisera les abréviations $\mathbf{x}$ et $\mathbf{b}$ pour les deux vecteurs colonnes dans cette équation, de sorte qu'elle s'écrive $A \cdot \mathbf{x} = \mathbf{b}$. Si on suppose qu'on a une solution $\mathbf{x}$ de ce système, et on multiplie l'équation à gauche par $M^{(j)}$ on trouve pour le premier membre $M^{(j)} \cdot A \cdot \mathbf{x} = \det(A)\,{}^t\mathbf{e}_j \cdot \mathbf{x} = \det(A)x_j$, et pour le second membre $M^{(j)} \cdot \mathbf{b} = f(\mathbf{b}) = \det(A \leftarrow_j \mathbf{b})$. Ce qu'on a fait est de former une combinaison R-linéaire des équations dans laquelle toutes les inconnues autre que x_j ont disparu. Ce constat, même si ce n'est pas une méthode très efficace pour résoudre le système, est un important outil théorique.

4.4.1. Proposition. *Pour que $\mathbf{x} \in R^n$ soit solution de $A \cdot \mathbf{x} = \mathbf{b}$ avec $A \in \mathrm{Mat}_n(R)$ et $\mathbf{b} \in R^n$, il est nécessaire que $\det(A)x_j = \det(A \leftarrow_j \mathbf{b})$ pour $j = 1, \ldots, n$, où $\leftarrow_j$ indique remplacement de la colonne j.* □

Rassemblons ensuite les n matrices $M^{(j)}$ pour $j = 1, \ldots, n$, qui sont de taille $1 \times n$, verticalement en une matrice $M \in \mathrm{Mat}_n(R)$:

$$M = \big((-1)^{i-j} \det(A_{\widehat{i,j}})\big)_{j,i=1,\ldots,n}. \tag{25}$$

Attention, par la force des choses, dans cette formule c'est j qui est l'indice des lignes $M^{(j)}$ de la matrice M, et i qui est l'indice des colonnes, ce qui est inhabituel. En prenant le produit matriciel avec A, les n résultats $\det(A)\,{}^t\mathbf{e}_j$ se rassemblent également en une matrice $n \times n$, qui est le multiple scalaire par $\det(A)$ de la matrice identité id_n :

$$M \cdot A = \det(A)\,\mathrm{id}_n\,. \tag{26}$$

On voit que la matrice M possède une propriété bien particulière par rapport à la matrice A, pendant que ses coefficients sont (à signe près) des déterminants de matrices extraites de M (quel type de déterminant est connu plus généralement sous le nom *mineur* de M). Pour cette raison on appelle M la *comatrice* de A, ou plutôt c'est la transposée de M qui est appelée ainsi en terminologie française ; à tort, car ${}^tM$ ne bénéficie d'aucune propriété particulière par rapport à A qui ne passe pas par une transposition (par contre, dans la terminologie anglaise c'est bien M elle-même qui est appelée *adjugate matrix* pour A).

4.4.2. Définition/Proposition. *Pour une matrice carrée A de taille $n \times n$ à coefficients dans un anneau commutatif R, on appelle comatrice-transposée de A la matrice (du même type)*

$$\mathrm{comatr}(A) = \big((-1)^{j-i} \det(A_{\widehat{j,i}})\big)_{i,j=1,\ldots,n}, \tag{27}$$

où la matrice $A_{\widehat{j,i}}$ est celle obtenue à partir de A en supprimant la ligne j et la colonne i. Elle vérifie les égalités

$$\mathrm{comatr}(A) \cdot A = \det(A)\,\mathrm{id}_n = A \cdot \mathrm{comatr}(A). \tag{28}$$

La première égalité est celle qui nous a motivés de définir $\mathrm{comatr}(A)$. Or il découle de la formule définissant $\mathrm{comatr}(A)$ que $\mathrm{comatr}({}^tA) = {}^t\mathrm{comatr}(A)$ et donc ${}^t\mathrm{comatr}(A) \cdot {}^tA = \det({}^tA)\,\mathrm{id}_n$; en utilisant dans cette équation ${}^tP\,{}^tQ = {}^t(QP)$ ainsi que $\det({}^tA) = \det(A)$, on obtient la seconde égalité (transposée).

4.4.3. Théorème. *Dans l'anneau $\mathrm{Mat}_n(R)$ des matrices $n \times n$ à coefficients dans un anneau commutatif R, une matrice A possède une matrice inverse (dans $\mathrm{Mat}_n(R)$) si et seulement si $\det(A)$ possède un inverse dans R. Si c'est la cas, la matrice inverse A^{-1} est donnée par $A^{-1} = \det(A)^{-1}\,\mathrm{comatr}(A)$.*

Preuve. Si B est une matrice inverse pour A, la multiplicativité du déterminant appliquée à $A \cdot B = \mathrm{id}_n$ donne $\det(A)\det(B) = 1$ dans R, donc pour que A soit inversible dans $\mathrm{Mat}_n(R)$ il est nécessaire que $\det(A)$ soit inversible dans R. Mais si c'est le cas, proposition 4.4.2 montre que la matrice $\det(A)^{-1}\,\mathrm{comatr}(A)$ est bien une matrice inverse de A. Or, si un inverse existe, il est toujours unique. □

Ce théorème, appliqué à la matrice d'un endomorphisme $\phi \in \mathrm{End}(E)$ exprimé par rapport à une base quelconque, confirme que (la matrice et donc) ϕ est inversible (c'est-à-dire un isomorphisme) si et seulement si $\det(\phi) \neq 0$. Mais le théorème dit bien plus : il affirme par exemple aussi qu'une matrice à coefficients entiers possède une matrice inverse du même type si et seulement si son déterminant est 1 ou -1, car ce sont les seuls éléments avec un inverse (multiplicatif) dans $\mathbf{Z}$. Et une matrice A à coefficients entiers admet une autre matrice B qui lui est inverse modulo n (ce qui veut dire que AB et BA, sans forcément être égaux à id_n, ont des coefficients qui sont congruents modulo n aux coefficients correspondants de id_n, c'est-à-dire à 1 sur la diagonale et à 0 ailleurs) si et seulement si $\det(A)$ et n sont premiers entre eux (car c'est la condition pour que la classe de $\det(A)$ soit inversible dans $\mathbf{Z}/n\mathbf{Z}$).

4.4.4. Théorème [règle de Cramer]. *Un système d'équations de la forme (24) avec coefficients et inconnues dans un corps commutatif K possède une solution unique si et seulement si* $\det(A) \neq 0$. *Si c'est le cas, les solutions pour les inconnues sont données par* $x_j = \frac{\det(A \leftarrow_j \mathbf{b})}{\det(A)}$ *pour* $j = 1, \ldots, n$.

Preuve. Si $\det(A) \neq 0 \in K$ alors A est inversible, et $\mathbf{x} = A^{-1} \cdot \mathbf{b}$ est une solution. L'unicité de cette solution et l'expression pour les inconnues individuelles sont des conséquences de la proposition 4.4.1. Si par contre $\det(A) = 0$, alors l'équation $A \cdot \mathbf{x} = \vec{0}$ a des solutions non nulles, qui peuvent être ajoutées à toute solution éventuelle de $A \cdot \mathbf{x} = \mathbf{b}$, montrant qu'une telle solution ne peut jamais être unique. □

4.5. *Le polynôme caractéristique.*

Avec le déterminant comme outil pour tester si un endomorphisme est un isomorphisme par une *expression* en les coefficients de sa matrice (par rapport à une base fixée), nous pouvons caractériser les valeurs propres de l'endomorphisme comme les solutions d'une équation. Pour cela on constate d'abord :

4.5.1. Proposition. *Un scalaire $\lambda \in K$ est une valeur propre d'un endomorphisme $\phi \in \mathrm{End}(E)$ (avec E un K-espace vectoriel de dimension finie) si et seulement si* $\det(\lambda \,\mathrm{id}_E - \phi) = 0$.

Preuve. Un vecteur propre pour ϕ est par définition un vecteur non nul dans le noyau de l'application $\lambda \,\mathrm{id}_E - \phi$, et un tel vecteur existe si et seulement si cette application *n'est pas* injective, ce qui pour un endomorphisme de E est la même chose que de ne pas être un isomorphisme (grâce au théorème du rang). Or, d'après le théorème 4.3.6(5), c'est le cas précisément quand $\det(\lambda \,\mathrm{id}_E - \phi) = 0$. □

Si $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ pour une certaine base $\mathcal{B}$ de E, alors $\det(\lambda \,\mathrm{id}_E - \phi) = \det(\lambda \,\mathrm{id}_n - A)$ pour tout $\lambda \in K$, d'après théorème 4.3.6(3). Pour trouver une équation pour les valeurs propres, il suffit donc de trouver une expression pour $\det(\lambda \,\mathrm{id}_n - A)$ en termes de λ et des coefficients de A.

4.5.2. Définition. *Si $A \in \mathrm{Mat}_n(K)$, alors le polynôme caractéristique χ_A de A est*

$$\chi_A = \det(X \,\mathrm{id}_n - A) \in K[X],$$

où $X \,\mathrm{id}_n - A$ est la matrice $M \in \mathrm{Mat}_n(K[X])$ telle que $M_{i,i} = X - A_{i,i}$ et $M_{i,j} = -A_{i,j}$ si $i \neq j$.

Par exemple si $A = \left(\begin{smallmatrix} 3 & 5 \\ -1 & 7 \end{smallmatrix}\right)$ on a $\chi_A = \left|\begin{smallmatrix} X-3 & -5 \\ 1 & X-7 \end{smallmatrix}\right| = X^2 - 10X + 26$. En fait, le polynôme caractéristique peut être calculé sans connaître explicitement les coefficients de A ; par exemple pour une matrice 3×3 générique

$$A = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},$$

le polynôme caractéristique est donné par

$$\chi_A = \begin{vmatrix} X-a & -b & -c \\ -d & X-e & -f \\ -g & -h & X-i \end{vmatrix} = X^3 - (a+e+i)X^2 + (ae - bd + ai - cg + ei - hf)X - \det(A).$$

Cela nous mène au constat suivant pour polynôme caractéristique des matrices $n \times n$ en général.

4.5.3. Proposition. *Pour tout $A \in \mathrm{Mat}_n(K)$, le polynôme caractéristique χ_A est un polynôme unitaire de degré n. En fonction de A, chaque coefficient de χ_A est donné par une expression en les coefficients de A, dont en particulier $-\mathrm{tr}(A)$ pour le coefficient de X^{n-1}, où $\mathrm{tr}(A) = \sum_{i=1}^{n} A_{i,i}$ est la trace de A, et $\det(-A) = (-1)^n \det(A)$ pour le coefficient de X^0 (le coefficient constant) dans χ_A.*

Preuve. Le fait que χ_A est de degré n et que ses coefficients sont donnés par des expressions en les coefficients de A est clair à partir des définitions. La description concrète du coefficient de X^i se déduit assez facilement pour $i = n, n-1, 0$ en regardant comment on trouve des termes de ces degrés dans l'expansion du déterminant (pour $i = 0$ on peut aussi considérer la substitution $X := 0$). Pour être plus précis, on obtient en général une contribution au terme de X^i en multipliant ensemble une combinaison de i facteurs X parmi les n qui sont présents sur le diagonal, et qui sont encore multipliés par un produit de $n-i$ coefficients de $-A$; pour une combinaison S fixée on obtient comme contribution le déterminant $(n-i) \times (n-i)$ (le mineur) extrait de $-A$ sur les lignes et les colonnes correspondantes au *complément* de S. Pour $i = n$ et $i = 0$ il n'y a qu'une telle combinaison (la combinaison pleine respectivement vide), et pour $i = n-1$ il y a $\binom{n}{n-1} = n$ combinaisons, qui contribuent chacun un coefficient diagonal de $-A$. □

4.5.4. Théorème. *Les valeurs propres de $\phi \in \mathrm{End}(E)$ sont les racines du polynôme caractéristique χ_A, où $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ pour une base quelconque $\mathcal{B}$ de E.*

Preuve. Pour $\lambda \in K$, la matrice de $\lambda \mathrm{id}_E - \phi$ par rapport à la base $\mathcal{B}$ est $\lambda \mathrm{id}_n - A$ (car la matrice de l'homothétie $\lambda \mathrm{id}_E$ est toujours $\lambda \mathrm{id}_n$, quelle que soit la base). Son déterminant s'annule si et seulement si λ est une valeur propre de ϕ, et il vaut $\det(\lambda \mathrm{id}_n - A) = \chi_A[X := \lambda]$ d'après le théorème 4.3.2(1). □

On voit que si ϕ possède n valeurs propres distinctes $\lambda_1, \ldots, \lambda_n$, alors avec $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ pour une base quelconque $\mathcal{B}$ de E, le polynôme χ_A est unitaire de degré n avec $\lambda_1, \ldots, \lambda_n$ pour racines, ce qui n'est possible que si $\chi_A = (X - \lambda_1) \ldots (X - \lambda_n)$. Donc dans ce cas le polynôme caractéristique ne dépend pas de la base utilisée pour exprimer la matrice de ϕ. Mais il n'est pas difficile de montrer directement que cela est le cas en général, ce qui permettra de parler simplement du polynôme caractéristique de ϕ.

4.5.5. Proposition/Définition. *Pour $\phi \in \mathrm{End}(E)$, le polynôme χ_A pour $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ ne dépend pas de la base $\mathcal{B}$, et est appelé le polynôme caractéristique χ_ϕ de ϕ.*

Preuve. Si $\mathcal{B}'$ est une autre base et $A' = \mathrm{Mat}_{\mathcal{B}'}(\phi)$, on aura $A' = P^{-1} \cdot A \cdot P$, où P est la matrice de passage de $\mathcal{B}$ à $\mathcal{B}'$. Comme $P^{-1} \cdot \mathrm{id}_n \cdot P = \mathrm{id}_n$, on aura $X \mathrm{id}_n - A' = P^{-1} \cdot (X \mathrm{id}_n - A) \cdot P$ dans $\mathrm{Mat}_n(K[X])$, et donc $\chi_{A'} = \det(P^{-1} \cdot (X \mathrm{id}_n - A) \cdot P) = \det(P)\, \chi_A\, \det(P^{-1}) = \chi_A$ par multiplicativité du déterminant. □

Si ϕ est diagonalisable, cette proposition appliquée pour une base de diagonalisation montre que χ_ϕ est alors scindé, et la multiplicité de chaque racine λ de χ_ϕ est égale à la dimension de son espace propre. Réciproquement, si χ_ϕ est scindé et la dimension de chaque espace propre atteint la multiplicité de la racine correspondante dans χ_ϕ, alors la somme (toujours directe) des espace propres est de dimension $\deg(\chi_\phi) = \dim(E)$, donc elle est E tout entier, et ϕ est diagonalisable. On a donc

4.5.6. Critère. *Un endomorphisme ϕ est diagonalisable si et seulement si χ_ϕ est scindé, et la dimension de l'espace propre pour chaque valeur propre λ égale à la multiplicité de λ comme racine de χ_ϕ.*

Le calcul du polynôme caractéristique est simple en principe, mais peut être fastidieux quand la dimension n n'est pas petite (disons pour $n > 3$), car les méthodes habituelles pour simplifier une matrice dont on veut calculer le déterminant, en utilisant son caractère n-linéaire alterné, peuvent être frustrées par la présence de X (on ne peut pas inverser un polynôme non constant). Quelques principes simples peuvent néanmoins souvent simplifier le calcul du polynôme caractéristique.

4.5.7. Proposition. *Si pour une base $\mathcal{B}$ de E la matrice $M = \mathrm{Mat}_{\mathcal{B}}(\phi)$ est de la forme $\left(\begin{smallmatrix} A & C \\ \mathbf{0} & B \end{smallmatrix}\right)$ décrite dans théorème 4.3.2(6) avec $A \in \mathrm{Mat}_i(K)$ et $B \in \mathrm{Mat}_{n-i}(K)$, on a une décomposition $\chi_\phi = \chi_A\, \chi_B$.*

Preuve. Il suffit d'appliquer le théorème 4.3.2(6), compte tenu de la forme de $X \mathrm{id}_n - M$. □

4.5.8. Corollaire. *Pour un sous-espace ϕ-stable V de E, le polynôme caractéristique $\chi_{\phi|_V}$ divise χ_ϕ.*

Preuve. En choisissant une base de E qui commence avec une base de V, la proposition précédente s'applique, et A est la matrice de la restriction $\phi|_V$, par rapport à la base de V, donc $\chi_{\phi|_V} = \chi_A$. □

4.5.9. Proposition. *Si $\phi \in \mathrm{End}(E)$ possède une matrice triangulaire T par rapport à une certaine base (on dit dans ce cas que ϕ est trigonalisable), alors on a une factorisation $\chi_\phi = \prod_{i=1}^{n}(X - a_i)$ du polynôme caractéristique, où $a_1, \ldots, a_n \in K$ sont les coefficients diagonaux de T.* □

La réciproque est aussi vraie (si χ_ϕ est scindé dans $K[X]$ alors ϕ est trigonalisable), mais on laisse cela pour le dernier chapitre (théorème 5.2.6), car il n'y a pas de rapport direct avec les déterminants.

Chapitre 5. Réduction d'endomorphismes.

Dans ce dernier chapitre nous allons approfondir l'étude des endomorphismes d'un espace vectoriel de dimension finie n. Rappelons (proposition 2.2.2 et la remarque qui suit) qu'un tel endomorphisme est certainement diagonalisable s'il admet n valeurs propres distinctes (ce qui est le maximum possible). On sait aussi (théorème 4.5.4) qu'on peut trouver les valeurs propres de ϕ comme les racines du polynôme caractéristique χ_ϕ. Par conséquent, si χ_ϕ est scindé *et a racines simples*, alors ϕ sera toujours diagonalisable (c'est aussi évident dans le critère 4.5.6). Si ce cas est le cas générique pour $K = \mathbf{C}$ (c'est-à-dire qu'il se produit presque toujours, en fonction des coefficients d'une matrice de ϕ), il n'est pas le seul cas possible, et les autres cas, même s'ils sont plus rares, sont d'autant plus intéressants. On va étudier notamment ce qu'on pourra dire dans le cas où χ_ϕ possède une ou plusieurs racines multiples. On verra qu'il est extrêmement rare que la dimension d'un espace propre pour d'une racine mulitple de χ_ϕ atteigne sa multiplicité, et dans le cas contraire ϕ n'est pas diagonalisable, encore d'après le critère 4.5.6.

5.1. *Le polynôme minimal.*

Le fait que le polynôme caractéristique de $\phi \in \mathrm{End}(E)$ s'annule si on y substitue une valeur propre λ est basé sur la détection par le déterminant (en s'annulant) de la manque d'injectivité de $\lambda\,\mathrm{id}_E - \phi$, qui correspond précisément à l'existence de vecteurs propres. Mais le déterminant ne peut pas détecter plus un fois qu'il s'annule ; il serait notamment une erreur de croire que pour une racine multiple λ de χ_ϕ la dimension de $\mathrm{Ker}(\lambda\,\mathrm{id}_E - \phi)$ soit liée à la multiplicité de la racine (même si elle ne peut pas la dépasser). L'exemple du endomorphisme de K^2 de matrice $\left(\begin{smallmatrix}\lambda & x\\ 0 & \lambda\end{smallmatrix}\right)$, avec $\lambda, x \in K$, illustre cela : son polynôme caractéristique est $(X-\lambda)^2$ dans tous les cas, avec donc λ comme racine double, mais la condition $\dim \mathrm{Ker}(\lambda\,\mathrm{id}_E - \phi) = 2$ n'est vérifiée que dans le cas très particulier où $x = 0$, où ϕ est l'homothétie $\lambda\,\mathrm{id}_E$.

On obtientdra plus d'information par l'étude du polynôme minimal, un polynôme qui comme le polynôme caractéristique est défini en termes de ϕ et qui aura comme racines les valeurs propres, mais dont la définition est basée sur l'annulation de l'espace E tout entier, après la substitution $X := \phi$. Ainsi dans l'exemple le polynôme minimal ne sera $X - \lambda$ que dans le cas $x = 0$, pour indiquer que dans ce cas l'espace propre $\mathrm{Ker}(\lambda\,\mathrm{id}_E - \phi)$ est déjà E tout entier, mais pour $x \neq 0$ le polynôme minimal sera $(X-\lambda)^2$, indiquant que $\mathrm{Ker}\big((\lambda\,\mathrm{id}_E - \phi)^2\big) = E$. Le polynôme minimal détecte donc la différence entre la situation diagonalisable $x = 0$ et le reste, mais il ne mesure pas non plus la dimension de l'espace propre ; en effet, plus la dimension de l'espace propre pour λ est grande, moins le polynôme minimal sera en général de degré élevé, car $\lambda\,\mathrm{id}_E - \phi$ annule déjà une grande partie de l'espace.

5.1.1. Proposition/Définition. *Pour un endomorphisme ϕ d'un K-espace vectoriel de dimension finie, il existe des polynômes non nuls de $K[X]$ annulés par la substitution $X := \phi$. Le polynôme minimal μ_ϕ de ϕ est le générateur unitaire de l'idéal de $K[X]$ formé des polynômes annulés par cette substitution.*

Preuve. Comme le K-espace vectoriel $\mathrm{End}(E)$ est de dimension finie (de dimension n^2 si $n = \dim(E)$, pour être précis), la substitution $X := \phi$, qui est une application linéaire $K[X] \to \mathrm{End}(E)$, ne peut pas être injectif, car $K[X]$ est un K-espace vectoriel de dimension infinie. Cela montre que l'idéal en question n'est pas $\{0\}$, et les propositions 3.5.5 et 3.5.3 justifient alors la définition du polynôme minimal. □

Concrètement, on peut trouver le polynôme minimal μ_ϕ à partir de la matrice $A = \mathrm{Mat}_{\mathcal{B}}(\phi)$ par rapport une base $\mathcal{B}$ de E, ainsi. On considère dans le K-espace vectoriel $\mathrm{Mat}_n(K)$ de dimension n^2, la suite de matrices $A^0 = \mathrm{id}_n$, A, A^2, A^3, et on cherche la première relation de dépendance linéaire entre ces matrices, qui consistera à exprimer l'une de ces matrices comme combinaison linéaire des matrices précédentes : $A^d = c_0A^0 + \cdots + c_{n-1}A^{n-1}$. On aura alors $\mu_\phi = X^d - c_{n-1}X^{n-1} - \cdots - c_0X^0$. La recherche d'une combinaison linéaire des $A^0, \ldots, A^{i-1}$ qui vaille A^i est celle d'une solution pour un système *linéaire*, et en tant que telle la réponse ne changera pas si on élargit le corps K à un corps plus grand, ce qui permet de constater une chose valable aussi pour le polynôme caractéristique (pour des raisons plus simples) :

5.1.2. Proposition. *Le polynôme minimal μ_ϕ est déterminé par une matrice quelconque de ϕ, et ne change pas si cette matrice est interprétée comme élément de $\mathrm{Mat}_n(K')$ pour un corps K' contenant K.* □

Cette construction de μ_ϕ ressemble à celle, dans la preuve de la proposition 2.3.1, du polynôme unitaire $P \in K[X]$ de degré minimal tel que $P \cdot_\phi v = \vec{0}$, pour $v \in E$ choisi. En comparaison avec cette preuve, où l'opération vérifie $P \cdot_\phi v = P[X := \phi](v)$, on évite pour le polynôme minimal μ_ϕ le choix de v, en exigeant que $\mu_\phi[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$, c'est-à-dire qu'il annule *tous* les vecteurs de E. Comme pour le polynôme P de cette preuve, les racines de μ_ϕ sont des valeurs propres de ϕ, ce qu'on montrera maintenant avec une légère variante de l'argument utilisé ; en fait on a ici également la réciproque.

5.1.3. Théorème. *Les valeurs propres de ϕ sont précisément les racines du polynôme minimal μ_ϕ.*

Preuve. Si $v \in E$ est vecteur propre de ϕ avec valeur propre λ, on a $\phi^k(v) = \lambda^k v$ pour tout $k \in \mathbf{N}$ et donc $P[X := \phi](v) = P[X := \lambda]v$ pour tout $P \in K[X]$. Par conséquent $\vec{0} = \mathbf{0}(v) = \mu_\phi[X := \phi](v) = \mu_\phi[X := \lambda]v$, et donc $\mu_\phi[X := \lambda] = 0$ car $v \neq \vec{0}$; ceci montre que λ est racine de μ_ϕ. Réciproquement, si λ est racine de μ_ϕ, celui-ci est divisible par $X - \lambda$: on peut écrire $\mu_\phi = (X - \lambda)Q$ avec $Q \in K[X]$, et comme $\deg(Q) < \deg(\mu_\phi)$ on a $Q[X := \phi] \neq \mathbf{0} \in \mathrm{End}(E)$. Il existe donc un vecteur non nul v dans $\mathrm{Im}(Q[X := \phi])$, qu'on peut donc écrire $v = Q \cdot_\phi w$ pour un certain $w \in E$, et on aura alors $(\phi - \lambda\,\mathrm{id}_E)(v) = (X - \lambda) \cdot_\phi (Q \cdot_\phi w) = \mu_\phi \cdot_\phi w = \mathbf{0}(w) = \vec{0}$. Ainsi λ est une valeur propre de ϕ. □

Cette propriété rapproche le polynôme minimal beaucoup au polynôme caractéristique. On verra ci-dessous qu'ils sont souvent le même polynôme, mais comme on a déjà montré, ce n'est pas toujours le cas. Une différence importante est que le polynôme minimal de $\phi \in \mathrm{End}(E)$ n'est pas forcément de degré $n = \dim(E)$, comme son polynôme caractéristique. Un exemple extrême est une homothétie $\lambda\,\mathrm{id}_E$, dont le polynôme minimal est toujours de degré 1, à savoir $X - \lambda$, quel que soit n.† Une autre différence est que les coefficients de μ_ϕ ne s'expriment pas directement en termes des coefficients d'une matrice de ϕ (contrairement à ceux du polynôme caractéristique, cf. la proposition 4.5.3) ; ce n'est pas possible, sachant que μ_ϕ est toujours unitaire, mais de degré variable. Les coefficients de μ_ϕ sont bien déterminés à partir d'une matrice de ϕ, comme indiqué ci-dessus, mais le système linéaire à résoudre n'est pas de nature "carrée", et ne peut donc pas être résolu par la règle de Cramer (ce qui donnerait des formules).

On peut étendre l'exemple des homothéties aux endomorphismes diagonalisables : leur polynôme minimal est déterminé par les valeurs propres distinctes, sans tenir compte de la dimension de leurs espaces propres, comme le montre le résultat suivant.

5.1.4. Proposition. *Si ϕ est diagonalisable, alors le polynôme μ_ϕ est scindé et à racines simples.*

Preuve. Si ϕ est diagonalisable, avec comme valeurs propres distinctes $\lambda_1, \ldots, \lambda_l \in K$, alors μ_ϕ est donné par le produit $P = (X - \lambda_1)\ldots(X - \lambda_l)$: d'une part on voit que $P[X := \phi]$ annule chaque vecteur propre de ϕ, et donc E tout entier qui possède une base de tels vecteurs ; d'autre part μ_ϕ doit avoir chacun des λ_i comme racine (d'après le théorème 5.1.3), et ne peut donc pas être de degré $< l$. □

En particulier, si χ_ϕ est scindé et à racines simples, on a vu que ϕ est diagonalisable, donc μ_ϕ est aussi scindé et à racines simples, et comme l'ensemble de racines est l'ensemble des valeurs propres de ϕ pour μ_ϕ aussi bien que pour χ_ϕ, on a nécessairement $\chi_\phi = \mu_\phi$ dans ce cas.

† Pour être tout à fait précis, il faut exclure $n = 0$: on peut appeler l'unique $\phi \in \mathrm{End}(E)$ une homothétie (pour n'importe quel facteur λ), mais μ_ϕ est de degré 0 ; en fait $\mu_\phi = 1$ ici, car $1[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$.

Mais contrairement à la situation pour χ_ϕ, réciproque de la proposition 5.1.4 est également vraie. Pour le voir, il est utile d'étudier d'abord plus généralement la situation où on connaît une décomposition du polynôme minimal.

5.1.5. Lemme. *Si $\mu_\phi = QP$ est une décomposition, où P et Q sont unitaires et $P \neq 1$, alors $V = \mathrm{Im}(P[X := \phi])$ est un sous-espace strict de E, et Q est le polynôme minimal de la restriction $\phi|_V$.*

Preuve. On a $Q[X := \phi|_V] = \mathbf{0} \in \mathrm{End}(V)$, car pour $v = P \cdot_\phi w \in \mathrm{Im}(P[X := \phi]) = V$ on calcule $Q[X := \phi|_V](v) = Q \cdot_\phi (P \cdot_\phi w) = \mu_\phi \cdot_\phi w = \vec{0}$. Comme $\deg(Q) = \deg(\mu_\phi) - \deg(P) < \deg(\mu_\phi)$ on a $Q[X := \phi] \neq \mathbf{0} \in \mathrm{End}(E)$, donc V est strictement plus petit que E. Finalement si Q' était un polynôme non nul avec $Q'[X := \phi|_V] = \mathbf{0}$ et $\deg(Q') < \deg(Q)$, on aurait $(Q'P)[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$ par le même calcul, pendant que $\deg(Q'P) < \deg(QP) = \deg(\mu_\phi)$, ce qui n'est pas possible ; donc $Q = \mu_{\phi|_V}$. □

On observe tout de suite une conséquence qui rapproche encore plus le polynôme minimal au polynôme caractéristique, et dont on verra plus loin que la réciproque est vraie aussi.

5.1.6. Corollaire. *Si le polynôme caractéristique χ_ϕ de $\phi \in \mathrm{End}(E)$ est scindé, μ_ϕ est aussi scindé.*

Preuve. Supposons que χ_ϕ est scindé. On écrit $\mu_\phi = QP$ où P scindé et unitaire, et où Q est sans racine dans K, comme dans la proposition 3.5.4(1). Le lemme 5.1.5 décrit un sous-espace ϕ-stable V tel que Q soit le polynôme minimal de la restriction $\phi|_V$; il s'agit de montrer que $Q = 1$ et donc $V = \{\vec{0}\}$. Comme Q est sans racine, $\phi|_V$ n'a pas de valeurs propres, et $\chi_{\phi|_V}$ est aussi sans racine dans K, et il divise χ_ϕ d'après le corollaire 4.5.8. Or χ_ϕ est scindé, donc ceci n'est possible que si $\chi_{\phi|_V} = 1$, c'est-à-dire $\dim(V) = 0$. □

En itérant le lemme 5.1.5, on voit que si on a une décomposition en plusieurs facteurs $\mu_\phi = Q_1 \cdots Q_l$, et si on pose $P_i = Q_{i+1} \cdots Q_l$ pour le produit des $l - i$ derniers facteurs ($i = 0, 1, \ldots, l$), alors on a une chaîne strictement croissante $\{\vec{0}\} = V_0 \subset \cdots \subset V_{l-1} \subset V_l = E$ de sous-espaces ϕ-stables $V_i = \mathrm{Im}(P_i)$. L'espace $\mathrm{Ker}(P_i)$ a la bonne dimension pour être un supplémentaire de V_i, mais que ce n'est pas toujours le cas (il est même possible que V_i n'ait pas de supplémentaire ϕ-stable du tout). Un cas où $\mathrm{Ker}(P_i)$ est effectivement un supplémentaire de V_i arrive dans la démonstration ce résultat, annoncé ci-dessus.

5.1.7. Théorème. *Le polynôme μ_ϕ est scindé à racines simples si et seulement si ϕ est diagonalisable.*

Preuve. La proposition 5.1.4 affirme la partie "si". Réciproquement supposons μ_ϕ scindé à racines simples $\lambda_1, \ldots, \lambda_l$, donc $\mu_\phi = (X - \lambda_1) \ldots (X - \lambda_l)$, où d'après le théorème 5.1.3, $\lambda_1, \ldots, \lambda_l$ sont les valeurs propres distinctes de ϕ. Montrons par récurrence sur l que $E = \bigoplus_{i=1}^{l} \mathrm{Ker}(\phi - \lambda_i \,\mathrm{id}_E)$, la somme directe des espaces propres, le cas $l = 0$ étant évident (car $\mu_\phi = 1$ implique $\dim(E) = 0$). En appliquant le lemme 5.1.5 avec $P = X - \lambda_l$ on trouve que $\mu_{\phi|_V} = (X - \lambda_1) \ldots (X - \lambda_{l-1})$ pour $V = \mathrm{Im}(\phi - \lambda_l \,\mathrm{id}_E)$, et donc par hypothèse de récurrence $V = \bigoplus_{i=1}^{l-1} \mathrm{Ker}(\phi - \lambda_i \,\mathrm{id}_E)$. On sait que $\dim(V) + \dim \mathrm{Ker}(\phi - \lambda_l \,\mathrm{id}_E) = \dim(E)$ par le théorème du rang, donc il suffit de montrer que V et l'espace propre $\mathrm{Ker}(\phi - \lambda_l \,\mathrm{id}_E)$ sont en somme directe. Mais λ_l n'étant pas parmi les racines $\lambda_1, \ldots, \lambda_{l-1}$ de $\mu_{\phi|_V}$, il n'est pas valeur propre de $\phi|_V$, donc $V \cap \mathrm{Ker}(\phi - \lambda_l \,\mathrm{id}_E) = \{\vec{0}\}$ comme voulu. □

On aurait aussi pu déduire de la remarque après la proposition 2.2.2 que les différents espaces propres sont toujours en somme directe, et terminer ainsi la preuve. Le point essentiel utilisé reste en tout cas que les λ_i sont distincts.

5.1.8. Corollaire. *Si $\phi \in \mathrm{End}(E)$ vérifie $P[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$ pour un certain $P \in K[X]$ qui est scindé et à racines simples, alors ϕ est diagonalisable.*

Preuve. La condition $P[X := \phi] = \mathbf{0}$ veut dire que μ_ϕ divise P, ce qui entraîne que μ_ϕ, comme P, est scindé et à racines simples (grâce au théorème 3.6.6 : les facteurs irréductibles de μ_ϕ le sont aussi de P). □

Ce corollaire est parfois plus simple à appliquer que le théorème 5.1.7, car il n'y a pas d'obligation de vérifier que le polynôme annulé par l'évaluation $X := \phi$ soit minimal. Par exemple tout endomorphisme vérifiant $\phi^2 = \phi$, qu'on appelle un *projecteur*, est diagonalisable, car $X^2 - X = X(X - 1)$ est scindé à racines $0, 1$ qui sont simples (quel que soit le corps K). Pour $K = \mathbf{C}$, un endomorphisme vérifiant $\phi^k = \mathrm{id}_E$ pour $k \in \mathbf{N}$ est aussi toujours diagonalisable, car $X^k - 1$ possède k racines distinctes dans $\mathbf{C}$.

5.2. *Sous-espaces caractéristiques, trigonalisation.*

Considérons maintenant la situation où μ_ϕ possède une racine multiple λ, de multiplicité $m > 1$. Cela veut dire qu'on peut décomposer $\mu_\phi = Q(X-\lambda)^m$, avec un quotient $Q \in K[X]$ qui n'est pas divisible par $X - \lambda$. Dans ce cas proposition 5.1.4 dit que ϕ ne peut pas être diagonalisable. On obtient, en itérant le lemme 5.1.5, une chaîne de sous-espaces vectoriels ϕ-stables $E = V_0 \supset V_1 \supset \cdots \supset V_m$ où $V_i = \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^i\big)$, qui est strictement décroissante, et dans laquelle on a donc $V_i = (\phi - \lambda\,\mathrm{id}_E)(V_{i-1})$ pour $i = 1, \ldots, m$. Mais au-delà de la m-ème itération, l'application de $\phi - \lambda\,\mathrm{id}_E$ ne réduira plus l'espace, autrement dit $(\phi - \lambda\,\mathrm{id}_E)(V_m) = V_m$: la restriction $\phi|_{V_m}$ n'a plus λ comme valeur propre, car λ n'est pas racine de son polynôme minimal Q, et $(\phi - \lambda\,\mathrm{id})|_{V_m}$ est donc un isomorphisme $V_m \to V_m$. Ceci montre :

5.2.1. Proposition. *La multiplicité m d'une valeur propre λ de ϕ comme racine de μ_ϕ est égale au plus grand nombre pour lequel $\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$ est strictement contenu dans $\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^{m-1}\big)$, et pour lequel (ce qui est équivalent) $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$ contient strictement $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^{m-1}\big)$.*

Preuve. Il reste juste de montrer l'équivalence des deux conditions concernant m. On a les relations $\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^k\big) \subseteq \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^{k-1}\big)$ et $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^k\big) \supseteq \mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^{k-1}\big)$ quel que soit $k > 0$. Or d'après le théorème du rang $\dim \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^k\big) + \mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^k\big) = \dim(E)$ pour tout k, donc une comparaison des dimensions des sous-espace concernés donne la même différence dans les deux cas. □

Contrairement à V_m, les sous-espaces V_i pour $i < m$ contiennent des vecteurs propres pour λ : ce sont les vecteurs non nuls dans le noyau de l'application $(\phi - \lambda\,\mathrm{id})|_{V_i} : V_i \to V_{i+1}$. Un tel V_i n'est donc pas en somme directe avec l'espace propre $\mathrm{Ker}(\phi - \lambda\,\mathrm{id}_E)$, et en particulier pour $0 < i < m$ le sous-espace $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^i\big)$ (qui contient cet espace propre) *n'est pas un sous-espace supplémentaire* de $V_i = \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^i\big)$. Pour obtenir un espace qui soit facteur dans une décomposition de E en somme directe de sous-espaces ϕ-stables, il sera nécessaire de prendre pour l'exposant de $(\phi - \lambda\,\mathrm{id}_E)$ (au moins) la multiplicité m mentionnée dans la proposition, ce qui nous conduit à la définition suivante (qui est, malgré l'emploi du terme "caractéristique", sans lien direct avec le polynôme caractéristique).

5.2.2. Définition. *Pour $\phi \in \mathrm{End}(E)$ et λ une valeur propre de ϕ, le sous-espace caractéristique de ϕ pour λ est $E_\lambda = \mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$, où m est la multiplicité de λ comme racine de μ_ϕ.*

Cette multiplicité m est le plus petit exposant k tel que $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^k\big)$ ne s'élargisse plus. On peut donc la remplacer dans la définition par un nombre plus grand, sans que cela ne change l'espace défini ; notamment on voit souvent que la multiplicité m' de λ dans le polynôme caractéristique χ_ϕ est utilisée à la place de m, dont on verra qu'elle vérifie $m' \geq m$. Voilà une décomposition attendue de E en somme directe de sous-espaces ϕ-stables avec E_λ comme facteur.

5.2.3. Lemme. *Pour toute valeur propre λ de $\phi \in \mathrm{End}(E)$, on a une décomposition en somme directe $E = \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big) \oplus E_\lambda$ où m est comme dans la définition du sous-espace caractéristique E_λ. La restriction de ϕ au facteur $\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$ n'a plus λ comme valeur propre.*

Preuve. Le facteur $\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$ est le sous-espace appelé V_m ci-dessus ; on a vu qu'il ne contient pas de vecteur propre pour λ, et que la restriction $(\phi - \lambda\,\mathrm{id})|_{V_m}$ est un isomorphisme. Par le théorème du rang $\dim(E_\lambda) + \dim(\mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big) = \dim(E)$, et il suffit de montrer que la somme est directe : un vecteur de $V_m \cap E_\lambda$, annulé par l'isomorphisme $((\phi - \lambda\,\mathrm{id})|_{V_m})^m$ par définition de E_λ, est nécessairement nul. □

On obtient ainsi une généralisation du théorème 5.1.7, qui décrit une décomposition en somme directe d'espaces propres pour le cas où μ_ϕ est scindé à racines simples, au cas où les racines peuvent êtres multiples, et dans laquelle les espaces caractéristiques prennent la place des espaces propres.

5.2.4. Théorème. *Si μ_ϕ est scindé, alors $E = \bigoplus_{i=1}^k E_{\lambda_i}$ où $\lambda_1, \ldots, \lambda_k$ sont ses racines distinctes dans K.*

Preuve. Récurrence sur k, avec $k = 0$ donnant $\dim(E) = 0$. Pour $k > 0$ le lemme 5.2.3 pour $\lambda = \lambda_k$ donne $E = V \oplus E_{\lambda_k}$ où $V = \mathrm{Im}\big((\phi - \lambda\,\mathrm{id}_E)^m\big)$, pour quel sous-espace on a $\mu_{\phi|_V} = \mu_\phi/(X-\lambda)^m$, qui est scindé et aux racines distinctes $\lambda_1, \ldots, \lambda_{k-1}$. Donc $V = \bigoplus_{i=1}^{k-1} E_{\lambda_i}$ par récurrence, et le théorème suit. □

On remarque que d'après le corollaire 5.1.6 et les théorèmes 4.5.4 et 5.1.3, l'énoncé du théorème 5.2.4 reste vrai si l'on remplace μ_ϕ par χ_ϕ, et c'est sous cette forme qu'est souvent énoncé ce résultat.

Trigonalisation.

Si E_λ est un sous-espace caractéristique pour ϕ, il est clair que le polynôme minimal de la restriction de ϕ à E_λ est $(X-\lambda)^m$. Le raisonnement du début de cette section appliqué à E_λ donnera donc une chaîne de sous-espaces qui descend jusqu'au sous-espace nul : $E_\lambda = V_0 \supset V_1 \supset \cdots \supset V_m = \{\vec{0}\}$, où $V_i = (\phi - \lambda\,\mathrm{id}_E)(V_{i-1})$ pour $i = 1, \ldots, m$. En choisissant d'abord une base de V_{m-1} et l'étendant successivement à une base de V_{m-2}, …, à une base de V_1, et finalement à une base $\mathcal{B}$ de $V_0 = E_\lambda$, l'image par $\phi - \lambda\,\mathrm{id}_E$ de chaque vecteur de $\mathcal{B}$ sera une combinaison linéaire des vecteurs de base précédents, autrement dit $\mathrm{Mat}_{\mathcal{B}}((\phi - \lambda\,\mathrm{id})|_{E_\lambda})$ est *triangulaire* supérieur avec des coefficients nuls sur le diagonal. (Ce n'est pas la seule façon de choisir une telle base ; il serait aussi naturel de choisir d'abord une base de l'espace propre $\mathrm{Ker}(\phi - \lambda\,\mathrm{id}_E)$, puis l'étendre à une base de $\mathrm{Ker}\big((\phi - \lambda\,\mathrm{id}_E)^2\big)$, etc.) Pour $\mathrm{Mat}_{\mathcal{B}}(\phi|_{E_\lambda})$ on aura évidemment des coefficients λ sur le diagonal, donc en utilisant la proposition 4.5.9 et le lemme 5.2.3, on obtient le résultat suivant.

5.2.5. Proposition. *Le polynôme caractéristique de la restriction $\phi|_{E_\lambda}$ est $(X-\lambda)^{\dim(E_\lambda)}$, et dans cette expression l'exposant est égal à la multiplicité de λ comme racine du polynôme caractéristique χ_ϕ.* □

Demander, comme le fait le critère 4.5.6, que la dimension de l'*espace propre* pour λ soit égale à la multiplicité de λ comme racine de χ_ϕ, demande donc que cet espace propre soit égal à E_λ, ou encore que la restriction $\phi|_{E_\lambda}$ soit l'homothétie de facteur λ. Si l'on impose que E_λ soit un sous-espace donné V de dimension au moins 2, il n'y a donc *qu'une seule possibilité* pour $\phi|_V$ qui permet à ϕ d'être diagonalisable, parmi plein d'autres possibilités : pour une base quelconque de V, toute matrice triangulaire à coefficients diagonaux tous λ en donne une. Ceci justifie notre affirmation au début du chapitre qu'il est extrêmement rare que la dimension d'un espace propre pour une racine multiple de χ_ϕ atteigne sa multiplicité,

5.2.6. Théorème. *Un endomorphisme ϕ est trigonalisable si et seulement si χ_ϕ est scindé dans $K[X]$.*

Preuve. La partie "seulement si" est contenue dans l'énoncé de la proposition 4.5.9. Réciproquement si χ_ϕ est scindé, μ_ϕ l'est aussi d'après le corollaire 5.1.6, donc on a la décomposition $E = \bigoplus_{i=1}^k E_{\lambda_i}$ du théorème 5.2.4. En choisissant une base de trigonalisation dans chaque facteur E_{λ_i} comme indiqué ci-dessus, et en les combinant pour $i = 1, \ldots, k$ en une base de E, on obtient une base par rapport à laquelle la matrice de ϕ est triangulaire supérieure. □

5.2.7. Corollaire. *Le polynôme μ_ϕ est scindé dans $K[X]$ si et seulement si c'est le cas de χ_ϕ.*

Preuve. On a déjà vu la partie "si" dans le corollaire 5.1.6. Réciproquement si μ_ϕ est scindé, la partie "si" dans la preuve du théorème 5.2.6 s'applique (elle n'utilise que le fait que μ_ϕ est scindé) pour montrer que ϕ est trigonalisable, et la partie "seulement si" du même théorème montre que χ_ϕ est scindé. □

En résumé, dans le cas où μ_ϕ, ou de façon équivalente χ_ϕ, est scindé dans $K[X]$, l'espace E se décompose en en somme directe des espaces caractéristiques E_{λ_i}, et μ_ϕ et χ_ϕ se décomposent chacun de façon correspondante en facteurs de la forme $(X-\lambda_i)^{d_i}$. Dans le cas de μ_ϕ, l'exposant d_i de ce facteur est le plus petit nombre tel que $E_\lambda = \mathrm{Ker}\big((\phi - \lambda_i\,\mathrm{id}_E)^{d_i}\big)$ (c'est l'exposant m qui figure dans la définition de E_λ), et dans le cas de χ_ϕ il est donné par $d_i = \dim(E_{\lambda_i})$.

5.3. *Noyaux de polynômes d'un endomorphisme.*

Notre résultat principal est le théorème 5.2.4 ; le terme "décomposition des endomorphismes" fait probablement référence à cette décomposition. On n'a pas l'ambition dans ce cours ni d'étudier en détail ce qu'on peut dire quand χ_ϕ et μ_ϕ ne sont pas scindés, ni de faire une analyse fine de la structure des sous-espaces caractéristiques E_λ, bien que ce soient des sujets intéressants ; on peut les réserver pour un cours plus approfondi. Nous nous contenterons donc de donner quelques compléments qui sont faciles à obtenir, et qui éclairciront notre compréhension de la situation.

5.3.1. Lemme. *Si pour $\phi \in \mathrm{End}(E)$, $P \in K[X]$ et $v \in E$ on a $P \cdot_\phi v = \vec{0}$, on a également $P \cdot_\phi w = \vec{0}$ pour tout $w \in E$ qui s'écrit sous la forme $w = Q \cdot_\phi v$ pour $Q \in K[X]$.*

Preuve. On a l'embarras du choix entre les manières élémentaires pour voir ceci. L'ensemble des vecteurs annulés par P, $\mathrm{Ker}(P[X := \phi])$, est un sous-espace ϕ-stable (proposition 3.5.6), donc avec v il contient aussi tout vecteur $\phi^k(v)$ et donc $Q \cdot_\phi v$. Le fait que $P \cdot_\phi v = \vec{0}$ entraîne $P \cdot_\phi \phi(v) = \vec{0}$ peut aussi être obtenu directement en appliquant ϕ aux deux membres de l'équation, si on fait l'expansion de $P \cdot_\phi v$. Finalement on peut aussi simplement récrire $P \cdot_\phi w = (PQ) \cdot_\phi w = Q \cdot_\phi (P \cdot_\phi v) = Q \cdot_\phi \vec{0} = \vec{0}$. □

5.3.2. Corollaire. *Pour $\phi \in \mathrm{End}(E)$, et $v \in E$, $I = \{ P \in K[X] \mid P \cdot_\phi v = \vec{0} \}$ est un idéal de $K[X]$.*

Preuve. Il est évident que I vérifie la condition 3.5.2(1), et pour (2) c'est ce qui dit le lemme 5.3.1. □

D'après la proposition 3.5.3, le polynôme unitaire du plus petit degré vérifiant $P \cdot_\phi v$ divise donc tout autre tel polynôme (le fait que $I \neq \{0\}$ découle de la dimension finie de E, ou de $\mu_\phi \in I$). On appellera ce polynôme (qui figurait déjà dans la preuve de la proposition 2.3.1) le *polynôme minimal de v pour ϕ*.

Le polynôme minimal de v pour ϕ est plus simple à calculer que μ_ϕ, car il ne nécessite que le calcul des images itérées du seul vecteur v, et non pas celles d'une base entière (c'est-à-dire les puissances d'une matrice) ; aussi les questions de dépendance linéaire à résoudre sont dans E au lieu de $\mathrm{End}(E)$. En plus, le polynôme trouvé annule automatiquement tout l'espace engendré par les images calculés (lemme 5.3.1) qui risque fort d'être E tout entier, auquel cas on aura trouvé μ_ϕ sans le chercher. Dans le cas contraire on aura en tout cas trouvé un diviseur de μ_ϕ (corollaire 5.3.2), et le lemme 5.1.5 facilitera la recherche du facteur éventuellement manquant. Ainsi le calcul du polynôme minimal n'est pas vraiment plus difficile que celui du polynôme caractéristique, même si la recette est un peu moins rectiligne.

5.3.3. Théorème. *Pour $\phi \in \mathrm{End}(E)$ on a $\deg(\mu_\phi) \leq \dim(E)$.*

Preuve. Par récurrence sur le nombre de facteurs irréductibles de μ_ϕ (si ce nombre est nul, l'inégalité devient $0 \leq 0$). Si P est un tel facteur, qu'on peut supposer unitaire, on sait d'après le lemme 5.1.5 que le quotient $Q = \mu_\phi / P$ est le polynôme minimal d'un sous-espace ϕ-stable $V = \mathrm{Im}(P[X := \phi])$, et par le théorème du rang on a $\dim(V) + \dim \mathrm{Ker}(P[X := \phi]) = \dim(E)$. Pour passer de l'hypothèse de récurrence $\deg(Q) \leq \dim(V)$ à la conclusion souhaitée, il suffit donc d'établir $\deg(P) \leq \dim \mathrm{Ker}(P[X := \phi])$. Mais pour un vecteur non nul v de $\mathrm{Ker}(P[X := \phi])$ (on sait qu'il en existe car $\dim(V) < \dim(E)$), le polynôme minimal de v pour ϕ est P, car il divise P qui est irréductible. Cela veut dire que les vecteurs $\phi^k(v)$ pour $k = 0, 1, \ldots, \deg(P)-1$ forment une famille libre à $\deg(P)$ vecteurs, qui est contenue dans $\mathrm{Ker}(P[X := \phi])$ d'après le lemme 5.3.1. On obtient donc $\deg(P) \leq \dim \mathrm{Ker}(P[X := \phi])$, ce qui complète la preuve. □

On change maintenant le point de vue, en fixant un polynôme $P \in K[X]$ plutôt que $v \in E$. On étudiera donc le sous-espace $\mathrm{Ker}(P[X := \phi]) = \{ v \in E \mid P \cdot_\phi v = \vec{0} \}$, qu'on notera simplement $\mathrm{Ker}_\phi(P)$.

5.3.4. Proposition. *Si P divise P' dans $K[X]$, alors $\mathrm{Ker}_\phi(P) \subseteq \mathrm{Ker}_\phi(P')$.*

Preuve. Si $P' = PQ$ alors $v \in \mathrm{Ker}_\phi(P)$ entraîne $Q \cdot_\phi v \in \mathrm{Ker}_\phi(P)$ (lemme 5.3.1) et donc $v \in \mathrm{Ker}_\phi(P')$. □

Souvent $\mathrm{Ker}_\phi(P)$ est nul, et il n'est intéressant que si P divise μ_ϕ, à cause de la proposition suivante.

5.3.5. Proposition. *Pour $\phi \in \mathrm{End}(E)$ et $P \in K[X]$ on a $\mathrm{Ker}_\phi(P) = \mathrm{Ker}_\phi(D)$ où $D = \mathrm{pgcd}(P, \mu_\phi)$.*

Preuve. L'inclusion '$\supseteq$' est donnée par la proposition précédente. Pour l'inclusion $\mathrm{Ker}_\phi(P) \subseteq \mathrm{Ker}_\phi(D)$, il suffit de montrer que si on écrit $P = QD$, alors $Q[X := \phi]$ est injectif. Par construction Q et μ_ϕ sont premiers entre eux, donc si $Q \cdot_\phi v = \vec{0}$, le polynôme minimal de v pour ϕ est forcément 1, car il divise à la fois Q et μ_ϕ (on a toujours $\mu_\phi \cdot_\phi v = \vec{0}$). Par conséquent $v = \vec{0}$, établissant l'injectivité de $Q[X := \phi]$. □

Le résultat principal sur ces noyaux $\mathrm{Ker}_\phi(P)$ est le théorème de décomposition des noyaux, qui décrit de façon très générale une situation dans laquelle ils forment une somme directe. On a déjà vu des situations particulières où c'est le cas, dans les théorèmes 5.1.7 (dans lequel les noyaux sont les espaces propres de ϕ) et 5.2.4 (idem pour les espaces caractéristiques). Dans ces décompositions les polynômes des facteurs sont premiers entre eux, ce qui ne doit pas étonner car le noyau d'un facteur commun serait contenu dans les deux noyaux, ce qui est incompatible avec une somme directe. Le lemme suivant montre que le seul fait d'être premiers entre eux suffit pour avoir une bonne décomposition en somme directe.

5.3.6. Lemme. *Si $P, Q \in K[X]$ sont premiers entre eux, alors $\mathrm{Ker}_\phi(PQ) = \mathrm{Ker}_\phi(P) \oplus \mathrm{Ker}_\phi(Q)$ pour tout $\phi \in \mathrm{End}(E)$, et les projections de $\mathrm{Ker}_\phi(PQ)$ sur $\mathrm{Ker}_\phi(P)$ et sur $\mathrm{Ker}_\phi(Q)$ selon la somme directe peuvent être écrites comme des restrictions à $\mathrm{Ker}_\phi(PQ)$ de certains polynômes en ϕ.*

Preuve. Comme l'énoncé ne parle que de la restriction de ϕ à $\mathrm{Ker}_\phi(PQ)$ (qui contient $\mathrm{Ker}_\phi(P)$ et $\mathrm{Ker}_\phi(Q)$ d'après la proposition 5.3.4), on pourra sans perte de généralité remplacer E par $\mathrm{Ker}_\phi(PQ)$ et ϕ par sa restriction, et donc supposer $(PQ)[X := \phi] = \mathbf{0}$. Soient $U, V \in K[X]$ des coefficients de Bezout tels que $\mathrm{pgcd}(P, Q) = 1 = UP + VQ$, et posons $\pi_1 = (VQ)[X := \phi]$ et $\pi_2 = (UP)[X := \phi]$. On a donc $\pi_1 + \pi_2 = \mathrm{id}_E$, ainsi que $\mathrm{Im}(\pi_1) \subseteq \mathrm{Ker}_\phi(P)$ et $\mathrm{Im}(\pi_2) \subseteq \mathrm{Ker}_\phi(Q)$, car PQ divise PVQ et QUP pendant que $(PQ)[X := \phi] = \mathbf{0}$. Alors d'une part pour $v \in E$ on a $v = \pi_1(v) + \pi_2(v) \in \mathrm{Ker}_\phi(P) + \mathrm{Ker}_\phi(Q)$, ce qui montre que $\mathrm{Ker}_\phi(P) + \mathrm{Ker}_\phi(Q) = E$, et d'autre part pour $v_1 \in \mathrm{Ker}_\phi(P) \subseteq \mathrm{Ker}(\pi_2)$ et $v_2 \in \mathrm{Ker}_\phi(Q) \subseteq \mathrm{Ker}(\pi_1)$ on a $\pi_1(v_1 + v_2) = \pi_1(v_1) = (\pi_1 + \pi_2)(v_1) = v_1$ et $\pi_2(v_1 + v_2) = \pi_2(v_2) = (\pi_1 + \pi_2)(v_2) = v_2$, donc la somme $\mathrm{Ker}_\phi(P) + \mathrm{Ker}_\phi(Q)$ est directe, avec π_1, π_2 comme les projections sur ses facteurs. □

La démonstration donnée est la traditionnelle, qui fait ressortir au maximum le rôle de l'arithmétique des polynômes, notamment d'une relation de Bezout. On remarque cependant que la restriction faite à $\mathrm{Ker}_\phi(PQ)$ est essentielle : si on avait défini π_1, π_2 sur un espace où $(PQ)[X := \phi] \neq \mathbf{0}$, on ne saurait pas dire des choses très précises sur leurs images et leurs noyaux. (En revanche on peut remarquer que la démonstration donnée ne dépend pas de l'hypothèse que E ou $\mathrm{Ker}_\phi(PQ)$ soient de dimension finie.)

Une démonstration dans un esprit un peu plus "géométrique", similaire à celles dans les sections précédentes, pourrait continuer ainsi après la réduction au cas $(PQ)[X := \phi] = \mathbf{0}$. Cette équation implique $\mathrm{Im}(P[X := \phi]) \subseteq \mathrm{Ker}_\phi(Q)$ et on a $\mathrm{Ker}_\phi(P) \cap \mathrm{Ker}_\phi(Q) = \{\vec{0}\}$ car le polynôme minimal d'un vecteur v dans l'intersection est un diviseur commun de P et Q, donc il est 1 et $v = \vec{0}$ (comme dans la preuve de la proposition 5.3.5). La somme $\mathrm{Ker}_\phi(P) + \mathrm{Ker}_\phi(Q)$ est donc directe, et sa dimension est égale à $\dim \mathrm{Ker}_\phi(P) + \dim \mathrm{Ker}_\phi(Q) \geq \dim \mathrm{Ker}_\phi(P) + \dim \mathrm{Im}(P[X := \phi]) = \dim(E)$ (encore une fois le théorème du rang). Par conséquent la somme est égale à E, et on a égalité $\mathrm{Im}(P[X := \phi]) = \mathrm{Ker}_\phi(Q)$.

Il ne nous reste qu'à montrer que les projections sur chacun des facteurs sont des polynômes en ϕ ; les deux étant semblables, on considéra celle sur $\mathrm{Ker}_\phi(Q)$. L'endomorphisme $P[X := \phi]$ a pour noyau $\mathrm{Ker}_\phi(P)$ et (comme on vient de voir) pour image $\mathrm{Ker}_\phi(Q)$, mais sa restriction à cette image n'est pas l'identité, comme il le faudra pour la projection sur $\mathrm{Ker}_\phi(Q)$. Pour trouver cette projection il convient donc de corriger $P[X := \phi]$, en la faisant suivre par l'application inverse de sa restriction à $\mathrm{Ker}_\phi(Q)$. Pour cela on se sert du coefficient de Bezout U mentionné avant : comme $UP \equiv 1 \pmod{Q}$, la restriction à $\mathrm{Ker}_\phi(Q)$ de l'endomorphisme $U[X := \phi]$ donne l'inverse de celle de $P[X := \phi]$ (car dans $\mathrm{Ker}_\phi(Q)$ les polynômes "agissent modulo Q"), et la projection $E \to \mathrm{Ker}_\phi(Q)$ est donc donnée par $(UP)[X := \phi]$ (comme dans la première démonstration donnée).

5.3.7. 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*

$$\mathrm{Ker}_\phi(P_1 \cdots P_l) = \mathrm{Ker}_\phi(P_1) \oplus \cdots \oplus \mathrm{Ker}_\phi(P_l),$$

et les projections de la somme sur chacun des facteurs sont des restrictions de certains polynômes en ϕ.

Preuve. Par récurrence sur l ; pour $l \leq 1$ c'est évident. Pour $l \geq 2$ on applique le lemme 5.3.6 avec $P = P_1 \cdots P_{l-1}$ et $Q = P_l$ (ce qui est possible car $\mathrm{pgcd}(P_1 \cdots P_{l-1}, P_l) = 1$, qu'on déduit par exemple du lemme de Gauss (3.6.7) et le fait que P_i et P_l sont premiers entre eux pour tout $i < l$)). On obtient

$$\mathrm{Ker}_\phi(P_1 \cdots P_l) = \mathrm{Ker}_\phi(P_1 \cdots P_{l-1}) \oplus \mathrm{Ker}_\phi(P_l) = \mathrm{Ker}_\phi(P_1) \oplus \cdots \oplus \mathrm{Ker}_\phi(P_{l-1}) \oplus \mathrm{Ker}_\phi(P_l),$$

et la composition des projections réalise celle $\mathrm{Ker}_\phi(P_1 \cdots P_l) \to \mathrm{Ker}_\phi(P_j)$ par un polynôme en ϕ. □

Ce théorème s'applique notamment avec $\mu_\phi = P_1 \cdots P_l$ une décomposition du polynôme minimal en ses facteurs le plus petits qui sont premiers entre eux, autrement dit où chaque P_i est *une puissance* d'un polynôme irréductible, avec exposant égal à la multiplicité de ce polynôme irréductible dans μ_ϕ (pour que les autres P_j n'en contiennent pas). On a alors $\mathrm{Ker}_\phi(P_1 \cdots P_l) = E$, et on retrouve ainsi les théorèmes 5.1.7 et 5.2.4 comme des cas particuliers. Mais même si μ_ϕ n'est pas scindé, ce théorème nous fournit une décomposition de l'espace en somme directe, avec à coté des sous-espaces caractéristiques aussi des facteurs qui sont chacun le noyau d'une puissance d'un facteur irréductible de degré > 1 de μ_ϕ.

5.4. ***Le théorème de Cayley–Hamilton.***

Dans cette dernière section on mentionne un résultat célèbre qui rapproche encore plus le polynôme caractéristique au polynôme minimal, car il dit que $\chi_\phi[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$. Des résultats déjà vus nous font soupçonner ceci : le théorème 5.1.3, le corollaire 5.1.6, la proposition 5.2.5, et le théorème 5.3.3.

5.4.1. Théorème de Cayley–Hamilton. *Pour tout $\phi \in \mathrm{End}(E)$, le polynôme caractéristique χ_ϕ vérifie $\chi_\phi[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$. De façon équivalente, le polynôme minimal μ_ϕ divise χ_ϕ.*

Preuve pour le cas où χ_ϕ est scindé. Grâce au corollaire 5.1.6, il s'agit juste de vérifier que pour toute valeur propre λ sa multiplicité comme racine de χ_ϕ est au moins égale à celle comme racine de μ_ϕ. Mais notre analyse de la situation rend cela évident : on a $\dim(E_\lambda) \geq m$ dans la définition 5.2.2, compte tenu de la proposition 5.2.1 et le théorème du rang ; c'est aussi ce qui dit le théorème 5.3.3 appliqué à E_λ. □

Cette démonstration partielle suffit quand K est un corps algébriquement clos comme $\mathbf{C}$, car dans ce cas χ_ϕ est toujours scindé. La démonstration suffit aussi quand on peut voir K comme un sous-corps d'un corps K' pour lequel χ_ϕ est scindé dans $K'[X]$ (par exemple pour $K = \mathbf{R}$ ou $K = \mathbf{Q}$ en prenant $K' = \mathbf{C}$), pour la raison suivante. Si on représente ϕ par une matrice $A = \mathrm{Mat}_{\mathcal{B}}(\phi) \in \mathrm{Mat}_n(K)$ pour une base $\mathcal{B}$ quelconque, les polynômes caractéristique et minimal associés à A ne changent pas si on interprète A comme un élément de $\mathrm{Mat}_n(K')$, et donc comme la matrice d'un endomorphisme ϕ' d'un K'-espace vectoriel : c'est évident pour $\chi_\phi = \chi_A$, et pour μ_ϕ c'est ce que dit la proposition 5.1.2. On pourra alors appliquer le cas démontré pour ϕ', pourvu que le sens de "μ_ϕ divise χ_ϕ" ne change pas si on remplace K par K', ce qui est vrai : on peut interpréter cette condition soit sous la forme "$\chi_\phi[X := A] = \mathbf{0} \in \mathrm{Mat}_n(K)$", soit sous la forme "le reste de la division euclidienne de χ_ϕ par μ_ϕ dans $K[X]$ est 0", et dans ces deux formes il est clair que le fait de remplacer K par K' ne change rien. Finalement on peut *toujours* trouver (ou plutôt construire) un tel corps K' (mais pour une démonstration de cela on renvoie à un cours d'algèbre plus avancé), donc en fait le cas démontré couvre le cas général.

Néanmoins, il est étrange de devoir faire référence à un corps plus grand pour prouver une propriété qui ne concerne que des K-espaces. On propose donc une démonstration alternative qui reste entièrement dans ce cadre. Quand on ne peut pas être sûr de l'existence d'un espace propre, l'idée est d'utiliser à sa place un espace V engendré par les ϕ-images itérées d'un vecteur v non nul quelconque, comme dans la preuve du théorème 5.3.3. La base naturelle à utiliser dans V est $[v, \phi(v), \ldots, \phi^{d-1}(v)]$ où $d = \dim(V)$, et la matrice de $\phi|_V$ par rapport à cette base est particulièrement simple : chaque vecteur de base est envoyé vers le suivant, sauf le dernier dont l'image est déterminé par le polynôme minimal de v pour ϕ.

5.4.2. Définition. *La matrice compagnon d'un polynôme unitaire $P = X^n + c_{n-1}X^{n-1} + \cdots + c_0X^0$ est*

$$\begin{pmatrix} 0 & 0 & 0 & \ldots & -c_0 \\ 1 & 0 & 0 & \ldots & -c_1 \\ 0 & 1 & 0 & \ldots & -c_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & -c_{d-1} \end{pmatrix} \tag{29}$$

Dans la situation mentionnée ci-dessus, la matrice de la restriction $\phi|_V$ de ϕ à l'espace engendré par les image itérées de v, par rapport à la base $[v, \phi(v), \ldots, \phi^{d-1}(v)]$, est la matrice compagnon du polynôme minimal P de v pour ϕ. D'après la lemme 5.3.1, P est aussi le polynôme minimal de $\phi|_V$ tout entier.

5.4.3. Lemme. *Si $\mathrm{Mat}_{\mathcal{B}}(\phi)$ est la matrice compagnon d'un polynôme unitaire $P \in K[X]$, alors $\chi_\phi = P$.*

Preuve. C'est un simple calcul. Par définition du polynôme caractéristique, il faut montrer

$$\begin{vmatrix} X & 0 & 0 & \ldots & c_0 \\ -1 & X & 0 & \ldots & c_1 \\ 0 & -1 & X & \ldots & c_2 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & c_{n-1} + X \end{vmatrix} = X^n + \sum_{i=0}^{n-1} c_i X^i = P.$$

On peut commencer à modifier la matrice, rendant ses coefficients diagonaux à l'exception du dernier tous nuls, en ajoutant la dernière ligne X fois de la précédente, puis cette ligne X fois de celle qui la précède, et ainsi de suite jusqu'à la première ligne. On obtient comme dernier coefficient de la première ligne $c_0 + X(c_1 + X(\cdots(c_{n-2} + X(c_{n-1} + X))\cdots)) = P$, et on vérifie facilement que

$$\begin{vmatrix} 0 & 0 & 0 & \dots & P \\ -1 & 0 & 0 & \dots & * \\ 0 & -1 & 0 & \dots & * \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -1 & c_{n-1}+X \end{vmatrix} = P.$$

Une autre méthode consiste à développer le déterminant de la matrice originale M directement par la dernière colonne. On remarque que pour $0 \leq i < n$ la matrice $M_{(i)}$ obtenue de M en supprimant la ligne contenant c_i et la dernière colonne est de la forme en blocs $\left(\begin{smallmatrix} L & \mathbf{0} \\ \mathbf{0} & U \end{smallmatrix}\right)$ où $L \in \mathrm{Mat}_i(K)$ est triangulaire inférieure avec des coefficients diagonaux X, et $U \in \mathrm{Mat}_{n-1-i}(K)$ est triangulaire supérieure avec des coefficients diagonaux -1. Donc $\det(M_i) = (-1)^{n-1-i}X^i$, ce qui donne $\det(M) = X^n + \sum_{i=0}^{n-1} c_i X^i = P$ comme voulu. Finalement on pourra aussi faire une démonstration par récurrence sur n, les cas $n \leq 1$ étant immédiats, et les cas $n > 1$ suivent par récurrence après un développement par la première ligne. □

Preuve du théorème de Cayley–Hamilton. Par récurrence sur $n = \dim(E)$, le cas $n = 0$ étant évident (car alors $\mathrm{End}(E) = \{\mathbf{0}_E\}$). Si $n > 0$ on choisit un vecteur non nul v, et soit $P \in K[X]$ le polynôme minimal de v pour ϕ. Alors $\mathcal{B}_v = [v, \phi(v), \dots, \phi^{d-1}(v)]$, où $d = \deg(P) > 0$, est une famille libre. C'est une base du sous-espace $V = \mathrm{Vect}(v, \dots, \phi^{d-1}(v))$ qui est ϕ-stable, et la matrice $\mathrm{Mat}_{\mathcal{B}'}(\phi|_V)$ est la matrice compagnon de P. En étendant $\mathcal{B}_v$ à une base $\mathcal{B}$ de E, la matrice $\mathrm{Mat}_{\mathcal{B}}(\phi)$ sera de la forme en blocs $M = \left(\begin{smallmatrix} A & * \\ \mathbf{0} & B \end{smallmatrix}\right)$ où '$*$' désigne un bloc dont le contenu nous n'intéressera pas (et qui sera différent pour les autres utilisations de '$*$' dans la suite). On voit par une calcul direct que M^n est de la forme $\left(\begin{smallmatrix} A^n & * \\ \mathbf{0} & B^n \end{smallmatrix}\right)$ pour $n \in \mathbf{N}$, et donc $Q[X := M] = \left(\begin{smallmatrix} Q[X:=A] & * \\ \mathbf{0} & Q[X:=B] \end{smallmatrix}\right)$ pour tout $Q \in K[X]$. D'après la proposition 4.5.7 on a $\chi_\phi = \chi_A \chi_B$. Comme $A = \mathrm{Mat}_{\mathcal{B}'}(\phi|_V)$ est la matrice compagnon de P, on a $\chi_A = P$ (lemme 5.4.3) et $P[X := A] = \mathbf{0} \in \mathrm{Mat}_d(K)$. Or $B \in \mathrm{Mat}_{n-d}(K)$ où $n - d < n$, donc l'hypothèse de récurrence donne $\chi_B[X := B] = \mathbf{0} \in \mathrm{Mat}_{n-d}(K)$. On conclut par un calcul sous la forme en blocs : $\chi_\phi[X := M] = P[X := M] \cdot \chi_B[X := M] = \left(\begin{smallmatrix} \mathbf{0} & * \\ \mathbf{0} & * \end{smallmatrix}\right) \cdot \left(\begin{smallmatrix} * & * \\ \mathbf{0} & \mathbf{0} \end{smallmatrix}\right) = \left(\begin{smallmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{smallmatrix}\right)$. □

Cette démonstration du théorème est tout à fait satisfaisante pour la forme de l'énoncé qu'on a donné, mais nous voulons néanmoins mentionner que ce théorème, qui est particulièrement riche en formulations et démonstrations diverses, peut aussi être donné sous une forme qui demande encore un autre type de preuve. Le polynôme caractéristique χ_ϕ étant défini par l'intermédiaire d'une matrice A de ϕ par rapport à une base, l'égalité $\chi_\phi[X := \phi] = \mathbf{0} \in \mathrm{End}(E)$ revient concrètement à l'égalité $\chi_A[X := A] = \mathbf{0} \in \mathrm{Mat}_n(K)$, valable pour toute matrice $A \in \mathrm{Mat}_n(K)$. Dans cette forme on peut observer que les coefficients du polynôme χ_A sont donnés par des expressions en les coefficients de A (proposition 4.5.3), et il est de même pour les coefficients des différents puissances A^k qu'on substitue pour les X^k dans ce polynôme, donc au bout du compte $\chi_A[X := A] = \mathbf{0}$ exprime un système de n^2 identités, extrêmement compliquées quand n est grand, en les coefficients de A. Ces identités ne faisant intervenir que des opérations additives et des multiplications, elles devraient être des conséquences des seuls axiomes d'un anneau commutatif, et donc rester valable pour les matrices carrées à coefficients dans un tel anneau. Une preuve dans ce contexte doit s'attaquer directement à la forme matricielle de l'identité, sans pouvoir tirer profit de considérations structurelles comme l'ont fait nos démonstrations (notamment les choix de bases adaptés à des sous-espaces particulières). Des telles preuves existent, et prennent en général comme point de départ l'identité (28) appliquée à la matrice $XI_n - A \in \mathrm{Mat}_n(K[X])$.

Finalement il y a un complément qu'on peut apporter au théorème, dans sa forme pour les endomorphismes ; non seulement μ_ϕ divise χ_ϕ, mais ils ont les *mêmes* facteurs irréductibles (la seule différence étant que la multiplicité d'un tel facteur peut être plus grand dans χ_ϕ). On l'a vu pour les facteurs de degré 1, correspondant aux racines. D'après le théorème de décomposition des noyaux on peut décomposer E en somme directe de sous-espaces V_i annulés chacun par une puissance d'un facteur irréductible différent P_i de μ_ϕ. Or on pourra montrer (pareillement à notre dernière preuve) que $\chi_{\phi|_{V_i}}$ est aussi une puissance de P_i, ce qui ne laisse pas de place pour d'autres facteurs irréductibles dans χ_ϕ.

Résumé des objectifs du cours.

Pour le premier chapitre toutes les notions doivent être bien comprises : combinaison linéaire, sous-espace, famille libre ou génératrice, base, coordonnées, dimension, application linéaire, matrice par rapport à des bases, isomorphisme, endomorphisme, rang, matrice de passage, changement de base. Dans les deux dernières sections il faut retenir surtout la proposition 1.4.2 et le théorème 1.4.3 du rang, qui avec le théorème 1.2.1 de la base incomplète forment le socle théorique de ce chapitre.

Le chapitre 2 sert surtout pour introduire et illustrer les problèmes de valeurs propres. Toutes les définitions ainsi que la proposition 2.2.2 sont fondamentales et donc à retenir. La proposition 2.3.1 sera supplantée par des résultats plus précis, donc ni sa formulation ni sa démonstration sont nécessaires à retenir en tant que tel. Avec cela, les deux dernières sections de ce chapitre servent surtout de motivation, mais les exemples de la section 2.4 méritent d'être bien étudiés, car ils sont proche du type d'applications auquel on peut s'attendre dans les contrôles.

Le chapitre 3 est long et nécessaire pour servir de fondement pour la suite, mais une fois ses fondements assimilés, relativement peu de choses seront directement pertinentes pour le cours : la construction de $\mathbf{Z}/n\mathbf{Z}$ sert surtout pour sa relation avec la division euclidienne dans $\mathbf{Z}$ (mais il est utile de retenir l'exemple des corps $\mathbf{Z}/p\mathbf{Z}$ avec p premier, même s'il est rare qu'on ose les faire intervenir dans les contrôles), pour les polynômes le plus importante à retenir, mis à part les propriétés basiques du degré, est encore la division euclidienne, puis les notions de substitution et de racine. Sont importants à retenir dans ce chapitre : les propositions 3.3.1, 3.4.4, 3.4.5, 3.5.1, 3.5.3, 3.5.5, et 3.5.6, et le théorème 3.6.6.

Dans le chapitre 4, les deux premières sections sont préparatoires, les propriétés des formes multilinéaires alternées sont surtout à retenir dans leur application au déterminant d'un système de vecteurs, le seul exemple qu'on en verra dans la pratique. Pour le déterminant il est important de savoir qu'il est défini de façon naturelle pour les matrices et les endomorphismes, mais que le déterminant d'un système de n vecteurs a besoin d'une base pour la fixer (normaliser). Les énoncés contenus dans le théorème 4.3.2 servent continuellement et doivent être connus. Les propriétés des deux autres formes du déterminant (théorèmes 4.3.4 et 4.3.6) sont importantes également. La section 4.4 n'est pas beaucoup utilisée dans ce cours, mais on a intérêt à avoir vu ses résultats, surtout pour leur utilisation dans d'autres cours d'algèbre. La dernière section sur le polynôme caractéristique est fondamentale dans ce cours.

Le deux premières sections du chapitre 5 mènent aux résultats les plus complets de ce cours, mais beaucoup de résultats sont préparatoires pour d'autres qui les supplantent. Pour le polynôme minimal on retiendra sa définition ainsi que les théorèmes 5.1.3 et 5.1.7, avec son corollaire 5.1.8. Pour les sous-espaces caractéristiques on retiendra surtout leur définition, le théorème 5.2.4, et la proposition 5.2.5. Les énoncés des résultats 5.2.6, 5.2.7, 5.3.7 et 5.4.1 sont jolis et faciles à mémoriser, donc de ce point de vue ce serait bien de les retenir ; ceci dit, ils sont à considérer comme des compléments facultatifs des résultats cités avant, car à eux seuls ils donnent trop peu de détails pour bien comprendre l'enjeu de la situation, et donc leur importance. Par ailleurs les deux dernières sections sont exclusivement pour information, bien qu'il puisse être utile de retenir ce qu'il y est dit concernant le calcul du polynôme minimal.

Table de matières.